ISSN 0258-6770 (PRINT) ISSN 1564-698X (ONLINE)

THE WORLD BANK ECONOMIC REVIEW

Volume 27 • 2013 • Number 2

www.wber.oxfordjournals.org

OXFORD UNIVERSITY PRESS

THE WORLD BANK ECONOMIC REVIEW

The *World Bank Economic Review* is a professional journal used for the dissemination of research in development economics broadly relevant to the development profession and to the World Bank in pursuing its development mandate. It is directed to an international readership among economists and social scientists in government, business, international agencies, universities, and development research institutions. The Review seeks to provide the most current and best research in the field of quantitative development policy analysis, emphasizing policy relevance and operational aspects of economics, rather than primarily theoretical and methodological issues. Consistency with World Bank policy plays no role in the selection of articles.

The Review is managed by one or two independent editors selected for their academic excellence in the field of development economics and policy. The editors are assisted by an editorial board composed in equal parts of scholars internal and external to the World Bank. World Bank staff and outside researchers are equally invited to submit their research papers to the Review.

For more information, please visit the Web sites of the Economic Review at Oxford University Press at www.wber.oxfordjournals.org and at the World Bank at www.worldbank.org/research/journals.

Instructions for authors wishing to submit articles are available online at www.wber.oxfordjournals.org. Please direct all editorial correspondence to the Editor at wber@worldbank.org.

SUBSCRIPTIONS: A subscription to *The World Bank Economic Review* (ISSN 0258-6770) comprises 3 issues. Prices include postage; for subscribers outside the Americas, issues are sent air freight.

Annual Subscription Rate (Volume 27, 3 Issues, 2013): *Institutions*—Print edition and site-wide online access: £184/$277/€277, Print edition only: £169/$255/€254, Site-wide online access only: £154/$232/€231; *Corporate*—Print edition and site-wide online access: £277/$415/€415, Print edition only: £254/$381/€380, Site-wide online access only: £230/$346/€347; *Personal*—Print edition and individual online access: £45/$67/€67. US$ rate applies to US & Canada, Euros€ applies to Europe, UK£ applies to UK and Rest of World. There may be other subscription rates available; for a complete listing, please visit www.wber.oxfordjournals.org/subscriptions. *Readers with mailing addresses in non-OECD countries and in socialist economies in transition are eligible to receive complimentary subscriptions on request by writing to the UK address below.*

Full prepayment in the correct currency is required for all orders. Orders are regarded as firm, and payments are not refundable. Subscriptions are accepted and entered on a complete volume basis. Claims cannot be considered more than four months after publication or date of order, whichever is later. All subscriptions in Canada are subject to GST. Subscriptions in the EU may be subject to European VAT. If registered, please supply details to avoid unnecessary charges. For subscriptions that include online versions, a proportion of the subscription price may be subject to UK VAT. Personal rates are applicable only when a subscription is for individual use and are not available if delivery is made to a corporate address.

BACK ISSUES: The current year and two previous years' issues are available from Oxford University Press. Previous volumes can be obtained from the Periodicals Service Company, 11 Main Street, Germantown, NY 12526, USA. E-mail: psc@periodicals.com. Tel: (518) 537-4700. Fax: (518) 537-5899.

CONTACT INFORMATION: Journals Customer Service Department, Oxford University Press, Great Clarendon Street, OxfordOX2 6DP, UK. E-mail: jnls.cust.serv@oup.com. Tel: +44 (0)1865 353907. Fax: +44 (0)1865 353485. *In the Americas, please contact*: Journals Customer Service Department, Oxford University Press, 2001 Evans Road, Cary, NC 27513, USA. E-mail: jnlorders@oup.com. Tel: (800) 852-7323 (toll-free in USA/Canada) or (919) 677-0977. Fax: (919) 677-1714. *In Japan, please contact*: Journals Customer Service Department, Oxford University Press, Tokyo, 4-5-10-8F Shiba, Minato-ku, Tokyo, 108-8386, Japan. E-mail: custserv.jp@oup.com. Tel: +81 3 5444 5858. Fax: +81 3 3454 2929.

POSTAL INFORMATION: *The World Bank Economic Review* (ISSN 0258-6770) is published three times a year, in February, June, and October, by Oxford University Press for the International Bank for Reconstruction and Development/THE WORLD BANK. Send address changes to *The World Bank Economic Review*, Journals Customer Service Department, Oxford University Press, 2001 Evans Road, Cary, NC 27513-2009. Periodicals postage paid at Cary, NC and at additional mailing offices. Communications regarding original articles and editorial management should be addressed to The Editor, *The World Bank Economic Review*, The World Bank, 3, Chemin Louis Dunant, CP66 1211 Geneva 20, Switzerland.

ENVIRONMENTAL AND ETHICAL POLICIES: Oxford Journals, a division of Oxford University Press, is committed to working with the global community to bring the highest quality research to the widest possible audience. Oxford Journals will protect the environment by implementing environmentally friendly policies and practices wherever possible. Please see http://www.oxfordjournals.org/ethicalpolicies.html for further information on environmental and ethical policies.

DIGITAL OBJECT IDENTIFIERS: For information on dois and to resolve them, please visit www.doi.org.

PERMISSIONS: For information on how to request permissions to reproduce articles or information from this journal, please visit www.oxfordjournals.org/jnls/permissions.

ADVERTISING: Advertising, inserts, and artwork enquiries should be addressed to Advertising and Special Sales, Oxford Journals, Oxford University Press, Great Clarendon Street, Oxford, OX2 6DP, UK. Tel: +44 (0)1865 354767; Fax: +44(0)1865 353774; E-mail: jnlsadvertising@oup.com.

PAPER USED: *The World Bank Economic Review* is printed on acid-free paper that meets the minimum requirements of ANSI Standard Z39.48-1984 (Permanence of Paper).

INDEXING AND ABSTRACTING: *The World Bank Economic Review* is indexed and/or abstracted by *CAB Abstracts, Current Contents/Social and Behavioral Sciences, Journal of Economic Literature/EconLit, PAIS International, RePEc (Research in Economic Papers)*, and *Social Services Citation Index*.

Typeset by Techset Composition Limited, Chennai, India; Printed by The Sheridan Press.

THE WORLD BANK ECONOMIC REVIEW

Volume 27 · 2013 · Number 2

Does Urbanization Affect Rural Poverty? Evidence from Indian Districts

Massimiliano Calì and Carlo Menon

Although a high rate of urbanization and a high incidence of rural poverty are two distinct features of many developing countries, there is little knowledge of the effects of the former on the latter. Using a large sample of Indian districts from the 1983–1999 period, we find that urbanization has a substantial and systematic poverty-reducing effect in the surrounding rural areas. The results obtained through an instrumental variable estimation suggest that this effect is causal in nature and is largely attributable to the positive spillovers of urbanization on the rural economy rather than to the movement of the rural poor to urban areas. This rural poverty-reducing effect of urbanization is primarily explained by increased demand for local agricultural products and, to a lesser extent, by urban-rural remittances, the rural land/population ratio, and rural nonfarm employment. JEL codes: O12, O18, O2, 13

The transformation from an agricultural and mainly rural economy to an industrial and predominantly urban economy is a typical feature of the process of economic development (Lewis 1954; Kuznets 1955). During this process, as urban areas grow, so does the productivity of their workers as a result of denser and larger markets for goods and production factors (Fujita and others 1999; Duranton and Puga 2004). However, whether the welfare gains from this process have any implications for welfare in surrounding rural areas and

Massimiliano Calì (corresponding author) is research associate at the Overseas Development Institute, UK. His e-mail address is m.cali.ra@odi.org.uk. Carlo Menon is an economist within the OECD, Science, Technology and Industry Directorate and the Bank of Italy.

The authors thank Andrew Shepherd, who inspired the main question of the research, and Henry Overman for their detailed comments on an earlier draft. Thanks also go to Kristian Behrens, Gilles Duranton, Giovanni Favero, Steve Gibbons, Vegard Iversen, Veena Jha, Guy Michaels, Elisabeth Sadoulet, Kunal Sen, Stephen Sheppard, Tony Venables, Hugh Wenban-Smith, three anonymous referees, and the participants at the LSE SERC junior seminar, the Urban Economics Association conference in New York and the ISI Annual Conference on Economic Growth and Development in Delhi, who provided useful comments. We are also grateful to Petia Topalova and Pingali Parthasarathy Rao for sharing some of the data and to the Tata Institute of Social Science in Mumbai for their kind hospitality. Financial support by the Chronic Poverty Research Centre and the India Observatory at LSE is gratefully acknowledged. The views expressed here do not necessarily reflect those of the OECD or the Bank of Italy. The other usual disclaimers also apply. A supplementary appendix to this article is available at http://wber.oxfordjournals.org/.

THE WORLD BANK ECONOMIC REVIEW, VOL. 27, NO. 2, pp. 171–201 doi:10.1093/wber/lhs019
Advance Access Publication August 3, 2012

the extent of such gains are not clear. These questions have been important in the analysis of the structural transformation in developed countries during the industrial revolution (Bairoch 1988; Williamson 1990; Allen 2009). However, only a few studies have investigated these issues in the context of today's developing countries (Dercon and Hoddinot (2005) is an exception), and little quantification is available for the effects of urbanization on rural poverty. In a period of increasing urbanization in most developing countries, the answers to these questions may have important implications for development policies.

This paper represents one of the first efforts to fill this gap by identifying and measuring the impact of urbanization on rural poverty in a large developing country. The relevance of the analysis is underscored by the fact that most of the world's poor reside in rural areas, where the incidence of poverty is higher than in urban areas across all developing regions. In 1993, rural areas accounted for 62 percent of the world population and 81 percent of the world's poor at the $1/day poverty line. In 2002, after a period of intensive urbanization, the same figures stood at 58 percent and 76 percent, respectively (Ravallion and others 2007).[1] The recent process of urbanization (which mostly involves the developing world) has been accompanied by an unequal distribution of the global reduction in poverty rates. Between 1993 and 2002, although the number of $1/day poor in rural areas declined by 100 million, the number of urban poor increased by 50 million. Ravallion and others (2007) explain this "urbanization of poverty" through two related effects. First, a large number of rural poor migrated to urban areas, thus ceasing to be rural poor, and they were either lifted out of poverty in the process or became urban poor. We define this effect as a *location* effect because it results from allocating the same people into different categories (i.e., rural vs. urban). Second, the process of urbanization also affects the welfare of those who remain in rural areas through urban-rural linkage effects. We call this effect an *economic linkage* effect.

No direct evidence is available on the relative importance of these two types of effects, but distinguishing between the *location* and the *economic linkage* effects is important. The former involves no structural links between urbanization and rural poverty and entails variation in rural poverty simply due to the change in residency from rural areas to cities of some of the rural poor. However, the *economic linkage* effects capture the impact of urban population growth on the rural rate of poverty. This relationship is structural in nature and indicates how good or bad urbanization is for rural poverty. Understanding this relationship is particularly important in developing countries because most of their population will continue to be rural for at least another decade—and for another three decades in the least developed countries.[2] This figure, along with

1. In fact, the actual poverty line used by Ravallion and others (2007) is $1.08/day; to simplify, we refer to it as the $1/day poverty line.

2. On the basis of calculations from UN (2008) data, developing countries are expected to become more urban than rural in 2018, as are least developed countries in 2045.

the recognition that poverty has a higher incidence in rural than in urban areas, suggests that the implications of urbanization will be most important in the near future for global poverty reduction among this rural nonmigrant population. The focus on developing countries is essential, given that almost all of the future population growth in urban areas (94 percent from 2005 to 2030) is predicted to occur in developing countries (UN 2008).

We measure the impact of urbanization on rural poverty in surrounding rural areas, distinguishing between the *location* and the *economic linkage* effects, using a large sample of Indian districts for the 1983–1999 period. During this period, the country urbanized at a relatively slow rate: the urban population was 23.3 percent of the total in 1981 and 27.8 percent of the total in 2001 (Government of India 2001). However, given the sheer size of the Indian population, this moderate increase translates into a massive rise in the absolute number of urban dwellers (126 million). This number represents an increase of almost 80 percent in the urban population over this period. These figures mask large variability in urbanization patterns at the sub-national level; in particular, districts have urbanized at very different rates (figure S3.1a in the supplemental appendix, available at http://wber.oxfordjournals.org/). For instance, a district such as Idukki in Kerala increased its urban population by 13,000 (+29 percent) between 1981 and 2001, whereas the urban population increased by 1.6 million (+416 percent) in Rangareddi (Andhra Pradesh) and by 2.4 million (+130 percent) in Pune (Maharashtra) over the same period. In the subsequent analysis, we attempt to exploit this variability to identify the impact of urbanization on rural poverty.

India also provides an interesting case in terms of the policy environment and economic performance given the structural changes in economic policy, the rate of growth, and the poverty levels experienced by the country during this period. Despite disagreements regarding the extent to which economic growth increased the welfare of India's poor, poverty in India declined steadily in the 1990s, particularly in rural areas (Kijima and Lanjouw 2003). The geography of the decrease in the share of the poor, however, is extremely varied (figure S3.1b in the supplemental appendix, available at http://wber.oxfordjournals.org/). Although over 30 percent of the rural population was lifted out of poverty in many districts between 1983 and 1999, for approximately a quarter of the districts, the incidence of poverty has remained roughly constant or has worsened over the same period.

India is the country with the largest number of both rural and urban poor. India's number of $1/day rural poor in 2002 was over 316 million, representing 36 percent of the world's rural poor. The country is expected to add a further 280 million urban dwellers by 2030.[3] Thus, estimating the impact of urbanization on rural poverty in India can help to identify the potential effects of this expected growth of the urban population on the world's largest stock of

3. This estimate is based on the authors' calculations using UN (2008) data.

rural poor. Our results suggest that urban growth has a significant poverty-reducing effect in rural areas within the same district, explaining between 13 percent and 25 percent of the overall decrease in rural poverty over the period.

The rest of the paper is structured as follows. The next section identifies the possible channels through which urban population growth affects rural poverty. Section two describes the data used. Section three details the empirical methodology employed, including the variables used and the strategy to distinguish *location* and *economic linkage* effects. Section four presents the results, and section five concludes.

Urban-rural linkages

We can draw from existing theoretical and empirical insights to identify the main mechanisms underlying the *economic linkage* effects of urbanization on economic conditions in nearby rural areas. There are at least six channels through which urban population growth can affect rural poverty in surrounding areas: consumption linkages, rural nonfarm employment, remittances, the rural land/labor ratio, rural land prices, and consumer prices.

Consumption Linkages

An expanding urban area will generate an increase in the demand for rural goods. This channel is likely to operate via an *income* effect as well as a *substitution* effect. The income effect is related to increased demand for agricultural goods due to higher (nominal) incomes in urban areas relative to rural areas. This phenomenon was documented as early as the Industrial Revolution; Allen (2009) describes how trade and proto-industrialization in British cities increased the urban consumption of goods produced in the countryside. This higher income is explained by urbanized economies: urban areas have denser markets for products and factors, which increase labor productivity and wage levels over the levels of rural areas (Duranton and Puga, 2004). The substitution effect is related to the increased share of higher value-added products in total agricultural demand that is typical of more sophisticated urban consumers. Empirical evidence from India and Vietnam confirms this composition effect (Parthasarathy Rao and others 2004; Thanh and others 2008). As noted below, we expect these effects to grow stronger closer to urban areas as a result of the weakly integrated agricultural markets within India (Jha and others 2005).

Rural Nonfarm Employment

Expanding urban areas can also favor the diversification of economic activity away from farming, which typically has a positive effect on income (e.g., Berdegue and others 2001; Lanjouw and Shariff 2002; Jacoby and Minten 2009). This effect is particularly important in the rural areas surrounding the cities. Three concomitant effects can explain this increased diversification. First, proximity to cities may allow some of the peripheral urban workforce to

commute to the city to work. Commuting, in turn, generates suburban nonfarm jobs in services, such as consumer services and retail trade, which are needed by the growing commuter population. Second, because cities provide dense markets to trade goods and services more efficiently, rural households close to cities can afford to specialize in particular economic activities (because of their comparative advantage), relying on the market for their other consumption and input needs (Fafchamps and Shilpi 2005; Dercon and Hoddinot 2005). This more extensive specialization is likely to boost productivity and income (Becker and Murphy 1992). Third, proximity to urban areas stimulates the nonfarm activities that are instrumental to agricultural trade (which is increased by urbanization), such as transport and marketing. Recent evidence from Asia provides strong support for the effect of cities in stimulating high-return nonfarm employment in nearby rural areas (Fafchamps and Shilpi 2003; Fafchamps and Wahba 2006; Deichmann and others 2008; Thanh and others 2008).

Remittances

Remittances sent to rural households of origin by rural-urban migrants constitute another potentially important *economic linkage* effect of urbanization on rural poverty. The vast majority of rural-urban migrants (between 80 percent and 90 percent) send remittances home, with varying proportions of income and frequency (Ellis 1998). To the extent that urbanization is (partly) fuelled by rural-urban migration, this growth may be associated with larger remittance flows to the rural place of origin. Stark (1980) and Stark and Lucas (1988) provide evidence in support of the positive effects of remittances in reducing resource constraints for rural households and providing insurance against adverse shocks (because their income is uncorrelated with the risk factors of agriculture).[4]

Rural land/labor ratio

Urbanization and rural poverty may also be linked through the changes in the rural labor supply that accompany the urbanization process. To the extent that rural-urban migration reduces the rural labor supply, this reduction would increase (reduce the decrease of) the land available per capita in rural areas. Given a fixed land supply and diminishing marginal returns to land, this increased availability should increase labor productivity in agriculture, creating some upward pressure on rural wages.[5] There is evidence in India of an

4. The migrant's family often provides economic support (monetary or in-kind) to the migrant during his initial stay in the urban area. This support, intended to cover the fixed costs of migration, can be interpreted as an investment whose main return is the counter urban-to-rural remittance flow that is received afterward (Stark 1980). This urban-to-rural remittance flow can somewhat reduce the net resources transferred to rural areas by urban workers.

5. To the extent that rural-urban migration is concentrated among the most productive rural workers, this may counter the productivity-enhancing effect of migration. However, the incentive for highly productive agricultural workers to move from rural to urban areas is partly softened because the skills that make them productive in agriculture are not easily transferable to the urban sector.

association between out-migration from rural areas and higher wages in the sending areas (Jha 2008).[6]

Rural Land Prices

The growth of cities can increase prices of agricultural land (owned by farmers) in nearby rural areas as a result of greater demand for agricultural land for residential purposes. This increase may generate increased income for landowners through sale or lease or through enhanced access to credit markets, where land acts as collateral. Some evidence from the US indicates that the expected (urban) development rents are a relatively large component of the agricultural land values in US counties that are near or that contain urban areas (Plantinga and others 2002). The consumption linkage channel of urbanization described above can also be expected to increase land rental prices as a result of a rise in the expected future stream of income from agriculture. The impact on rural poverty through this channel depends on how this increased income is distributed across the rural population. Typically, if land is very concentrated, this channel is likely to benefit a few landowners, potentially restricting access to waged agricultural employment for the landless population. Given the constraints on the reallocation of agricultural labor across sectors and the high labor intensity of agriculture, we would expect the net effect on rural poverty to be adverse (i.e., an increase in rural poverty) when land is highly concentrated (and vice versa).

Consumer Prices

Because the growth of a city is associated with lower consumer prices, surrounding rural consumers who have access to urban markets may benefit through higher real wages (Jacoby 2000). This effect may be due to increased competition among a larger number of producers in the growing urban area as well as thicker market effects in both factor and goods markets (Fujita and others 1999). However, the increased demand for agricultural goods may also lead to higher prices for these goods, especially if the supply is fairly inelastic. Therefore, the direction of the net effect of urbanization on consumer prices is a priori ambiguous.

The discussion above suggests that the total net effect of urbanization on rural poverty is poverty reducing. Moreover, the bulk of the effect is expected to be felt in rural areas that are relatively close to the growing urban area. This distance decay effect is consistent with recent research on the welfare impact of the expansion of a gold mine in Peru (Aragón and Rud 2009) and is important for our identification strategy, as explained below. The next sections will detail the methodology used to test these hypotheses by measuring the total net effect

6. In addition, an expanding city can benefit agriculture productivity in the surrounding rural areas through spillover effects in technology and marketing (Dore 1987; Allen 2009).

of urbanization on poverty in surrounding rural areas in the case of Indian districts.

Data

The data for the empirical analysis come from three main sources. For the district-level measures of poverty, we use data from three "thick" rounds of the National Sample Surveys (NSS: the Indian household survey) spanning the 1980s and the 1990s, the 38th (1983–84), 49th (1993–94) and 55th (1999–2000) rounds of the NSS.[7,8] These measures have been adjusted by Topalova (2010) in two ways. First, she uses the poverty lines (based on the state-level prices computed separately for rural and urban areas) proposed by Deaton (2003a, 2003b) instead of the standard Indian Planning Commission poverty lines, which are based on defective price indices. Second, she adjusts the consumption data from the 55th round to accommodate for a change in the survey design vis-à-vis the previous rounds (i.e., the recall period for certain goods).[9]

We are aware that the reliability of the district-level estimates of urban poverty is widely debated in light of the relatively small number of sampled households (Hasan and others 2007; Topalova 2010). Although we use the urban poverty measure at the district level, the results are also robust to using urban poverty at the level of the NSS regions, which are a census-based aggregation of several districts (there are approximately 60 NSS regions in India).[10] To the extent that the limited representativeness at the district level can be considered a classic nonsystematic measurement error in the dependent variable, it should not bias the estimate of the coefficients. It only suggests lower efficiency.

The other district-level data, such as population composition, come from the Indian districts database at the University of Maryland (which has been extracted from the original data in the Indian Census), which we update to 2001 using data from the Indian Census.[11] Data on town populations are available from various rounds of the Indian Census. In addition, for crop production volumes and values, we use the district-level database for India, available from

7. Although each survey was conducted over two years, we refer to them with the first of the two years.

8. The "thick" surveys are conducted approximately every five years and sample a higher share of households than do the "thin" surveys, thus allowing inferences at the district level. We do not use the other "thick" survey for the period, the 43rd round (1987–88), because we only use census data for the population variables, which do not have a natural match with the 1987 poverty data.

9. In particular, Topalova (2010) follows the adjustment made in Deaton (2003a and 2003b) and imputes the distribution of total per capita expenditure for each district from the households' expenditures on a subset of goods for which the new recall period questions were not used. The poverty and average consumption measures were derived from this corrected distribution of consumption from the detailed consumption schedule of the surveys.

10. Results are available upon request.

11. Available at www.bsos.umd.edu/socy/vanneman/districts/codebook/index.html.

the International Crops Research Institute for the Semi-Arid Tropics from 1980 to 1994 and recently updated to 1998 by Parthasarathy Rao and others (2004).[12]

The district classification was modified during the period of analysis because some districts were split into two units. Topalova (2010) created a consistent classification by aggregating the 2001 districts that originated from the district division of 1987. We conform to this reaggregation and modify the original population and demographic data accordingly.

Empirical strategy

We attempt to systematically assess whether and to what extent the urbanization in Indian districts during the 1983–1999 period affected rural poverty in these districts. We argue that districts are an appropriate spatial scale for such an analysis in India because all of the *location* and *economic linkage* channels described above are likely to display most of their effects within a district's boundaries. This argument is consistent with the theoretical discussion above suggesting that the effects of city growth are concentrated in surrounding rural areas. Specific evidence on India confirms that this relationship is likely to be the case.

First, intradistrict migration in India is a large component of total rural-urban migration. According to the census (Government of India 1991), 62 percent of the total stock of permanent internal migrants was intradistrict in 1991, although a share of this stock was composed of women migrating for marriage reasons.[13] This statistic does not deny the existence of long-distance migration in India, which increased during the 1990s (Jha 2008). However, long-distance rural-urban migration is primarily directed at a few growing metropolitan areas, such as Mumbai, Delhi, Bangalore, and Chennai, which are excluded from the analysis for this reason.[14] Notwithstanding the importance of intradistrict migration, in the empirical section, we test the robustness of the results against the relative size of the intradistrict migrant population.

Second, during the period of analysis, most agricultural goods markets do not appear to be well integrated at the national or even at the state level in India. The extremely poor transport infrastructure is one of the largest obstacles to trading bulky agricultural goods throughout the country (Atkin 2010). Furthermore, India maintains high internal food-trade barriers, including tariffs at state borders, licensing requirements for traders and district-level

12. The original source of these data is the Government of India, Directorate of Economic and Statistics, Ministry of Agriculture and Cooperation.

13. This finding is consistent with Topalova (2005), who finds limited labor mobility across Indian regions between 1983 and 2000.

14. In fact, Delhi and the urban Bangalore districts are automatically dropped because they do not have rural areas. In the following, we present robustness tests showing that the exclusion of Bangalore and Chennai increases precision but does not affect the main results of the analysis.

entry taxes (Das-Gupta 2006). Thus, a consistent share of agricultural trade tends to occur at a short distance, making districts a suitable spatial scale to capture a substantial part of the first two channels. Consistent with these ideas, some studies have tried to capture the demand-side effects on agriculture through district-level analyses (Parthasarathy Rao and others 2004).

There is also emerging evidence of increases in land prices in peri-urban and rural areas surrounding the urban agglomerates. The land values in these areas may be well above the discounted future stream of income from agricultural activity, inducing some landowners to sell the land (Jha 2008).[15]

We begin from the following basic specification:

$$H^R_{dt} = \alpha_d + \lambda_{st} + \beta P^U_{dt-j} + \chi X_{dt} + \varepsilon_{dt}, \tag{1}$$

where H^R_{dt} is a measure of rural poverty in district d at time t, α is the district fixed effect, λ is the state-year fixed effect,$P^U_{dt-j}$is the urban population of district d at time t-j (where$j \in [1, 2]$), and X is a vector of controls that includes other variables that are likely to have an independent impact on rural poverty, including, inter alia, the district's rural population to control for the growth of the district's overall population. The coefficient of interest here is β, which measures the collective impact of the *location* and *economic linkage* effects of urbanization on rural poverty.

We use a standard Foster Greer Thorbecke (FGT) measure of poverty for H^R_{dt}, the poverty headcount ratio. P^U_{dt-j}, the main independent variable, is computed as$P^U_{dt} = \sum_{i=1}^{N_d} u^d_{it-j}$, where u^d_{it-j} is the population of town i in district d at time t-j and N_d is the number of cities in district d. Because population figures from the census are available with only a 10-year frequency (e.g., 1981, 1991, 2001), the data for 1997 are estimated through nonlinear interpolation.[16]

The two sets of fixed effects included in (1) absorb some of the unobserved heterogeneity likely to bias the estimation. In particular, the district fixed effects absorb any time-invariant component at the district level, such as geographical position, climatic factors, and natural resources. The state-year fixed effects capture any state-specific time-variant shocks (including economic dynamics and policies). This specification, however, may not completely account for other sources of potential bias in the coefficient β, including endogeneity. For this

15. All of this evidence appears to be roughly consistent with Fafchamps and Shilpi (2003), who find that in Nepal, the effects of proximity on rural areas decrease beyond a four-hour radius (in travel time) around the cities. Using the boundaries of the Indian districts as in 1987, the average district size in our analysis is approximately 7,300 km^2. If we approximate the district with a circle, a city located in its center would be approximately 50 km from the boundary of the district. It is plausible that in several districts, this distance could be covered in three to four hours on rural Indian roads during the period considered.

16. Appendix S1 (available online at http://wber.oxfordjournals.org) contains a detailed description of all of the variables included in the empirical analysis.

reason, in the next section, we address a number of possible remaining concerns by means of several robustness tests and instrumental variable (IV) estimations.

Disentangling Economic Linkage from Location Effects

As mentioned above, it is important to empirically distinguish the *economic linkage* effects from the simple composition effect due to the migration of the poor from rural to urban areas (*location* effects) when identifying the impact of urbanization on rural welfare. To disentangle these effects, we start from the observation that a district's urban population grows more rapidly than the rural population because of two phenomena: intradistrict rural-urban migration or rural areas becoming urban (either because they are encompassed by an expanding urban area or because their population has grown sufficiently to upgrade from the status of village to that of town).[17] These two phenomena directly change the composition of both the rural and the urban populations, including their poverty rates. If the distribution of rural-urban migrants is skewed toward low-income individuals (i.e., the incidence of poverty is higher among migrants than nonmigrants) and if the poverty incidence in rural villages that become urban is higher than it is in the total rural population of the district, then rural-urban migration will directly reduce rural poverty. This example demonstrates the poverty-reducing *location* effect of urban growth.

To properly isolate the *location* effects of urbanization on rural poverty, we need representative data on the poverty profile of rural-urban migrants and the dwellers in areas that are rural at time t-1 and that become urban at time *t*. Unfortunately, these data are not available for the Indian context. Thus, we must find indirect proxies for this information.[18]

We use two sets of variables for this purpose. The first set is composed of three rural socio-demographic variables: the number of rural people in the 15–34 year age group, the share of literates in this group, and the share of the rural population classified as scheduled caste. The rationale behind the inclusion of these variables as proxies for the *location* effects of urbanization relies on the assumption that the poverty distribution of migrants can be expressed as a

17. This does not consider the possibility of interdistrict migration or of urban-rural migration. The latter is relatively unimportant in influencing the rural-urban split of the population in a country such as India. The stock of urban-rural migrants represented less than 1.4 percent of the total population in the majority of Indian districts in 1991, with a mean equal to 1.7 percent (based on the Indian districts database at the University of Maryland; see below). The interdistrict migration represents a substantial share of the total migration, particularly rural-urban migration. In 1991, it accounted for less than 34 percent of the total migration for the majority of Indian districts (with a mean equal to 37 percent). However, the empirical analysis below rejects the relevance of this type of migration in determining rural poverty.

18. In fact, the National Sample Survey contains socio-demographic information on intradistrict rural-urban migrants, but we cannot use this information for two reasons. First, the number is not large enough to be representative at the district level. Second, the survey only asks about migration in the last 365 days, whereas we would need information on migrants in the last five years. Although we cannot use these data in the regressions, we nevertheless refer below to the national-level expenditure comparison between rural-urban migrants and stayers (Singh 2009).

function of the migrants' age, literacy, and caste composition. Other things being equal, the incidence of poverty tends to be lower among young adults (i.e., 15–34 years old) because they represent the most productive age class.[19] Therefore, the higher the share of young adults in the total migrant population (relative to their share in the rural population), the lower the probability is that urbanization will directly reduce rural poverty. Because we do not observe the composition of the migrant population, we can only control for it indirectly through the composition of the actual rural population. This strategy relies on the plausible assumption that the change in the number of young adults in the rural population is inversely related to the change in their number among the rural-urban migrant population in the same period.[20] The same argument can be applied to literate migrants, whereas the reverse is true for scheduled caste members, whose poverty incidence is higher than that of the rest of the population.

The second type of variable capturing the *location* effects of urbanization is the urban poverty rate. Changes in this variable should indirectly reflect the poverty profile of rural-urban migrants in every period. This hypothesis follows from the fact that the probability that poor rural-urban migrants will become urban poor (after migrating) is higher than the same probability for nonpoor rural-urban migrants. Therefore, for any given level of urban economic growth (which is the main determinant of urban poverty) between *t* and *t-1*, urban poverty is more likely to decrease between *t* and *t-1*, with a lower share of rural poor migrating to urban areas during that period.

In addition, the urban poverty rate can control for some unobserved time varying district-specific shocks that can affect both rural poverty and the urban population. For example, there may be a localized shock that spurs the district's economic growth. Because economic growth is generally associated with urbanization, this growth can foster urbanization while simultaneously reducing rural poverty. This omitted variable problem implies a spurious negative association between the two variables. The data on income per capita at the district level are not available to us. However, because a district's economic growth is likely to be the main determinant of the evolution of urban poverty (as well as rural poverty) in that district, and to the extent that the

19. This pattern receives strong support in the empirical analysis that follows, which suggests that rural poverty is negatively associated with the rural young adult population and with the share of literates in this group.

20. This assumption is supported by the results of regressing the 1981–91 change in the urban population in the 15–34 year age group (ΔP^{U}_{15-34}) on the change in the rural population in the same age group (ΔP^{R}_{15-34} controlling for changes in the district's total population and the total population in 1981). The coefficient of $\Delta P^{R}_{15-34}$is not statistically different from -1, indicating that changes in the rural population are reflected in mirrored changes in the urban population through either rural-urban migration or a rural-to-urban change in the status of villages (results available from the authors upon request).

poverty-reducing effects of economic growth are similar across urban and rural areas, the urban poverty rate is a good proxy for a district's economic growth.

Because the main aim of this paper is to estimate the size and direction of the *economic linkage* effects of urbanization on rural poverty, we can now extend (1) to separately capture the *location* effects of urbanization:

$$H_{dt}^{R} = \alpha_d + \lambda_{st} + \beta' P_{dt-j}^{U} + \Gamma Z_{dt} + \gamma H_{dt}^{U} + \chi X_{dt} + \varepsilon_{dt}, \tag{2}$$

where Z is the vector of socio-demographic variables described above and H_{dt}^{U} is the urban poverty rate. To the extent that these additional variables capture the *location* effects, β' measures only the *economic linkage* effects of urbanization on rural poverty. This coefficient should thus quantify the collective effect of the urban-rural linkages described in section two. Because all of these linkages point toward the rural poverty-reducing effect of urban growth, β' is expected to be negative. The empirical analysis below attempts to disentangle the effects of some of these *economic linkage* channels. The difference between β in (1) and β' in (2) should be an indication of the size and direction of the *location* effects of urban growth on rural poverty.

There may be some concern that urban poverty could be endogenous in expression (2) because it is simultaneously determined with rural poverty. To relieve these concerns, we employ a variant of this specification without urban poverty that uses the number of urban nonpoor instead of the number of urban people as the main regressor:

$$H_{dt}^{R} = \alpha_d + \lambda_{st} + \beta'' P_{dt-j}^{U(NP)} + \Gamma Z_{dt} + \chi X_{dt} + \varepsilon_{dt}. \tag{2'}$$

Given the property of urban poverty discussed above, we would expect the coefficient of the number of urban nonpoor to more closely capture the *economic linkage* effects of urbanization on rural poverty (coefficient β'') than the β coefficient of the total urban population in (1), which captures the overall effect of urbanization on rural poverty. We have this expectation because the urban nonpoor variable would be much less affected by the quantity of poor rural-urban migrants (who are more likely to become urban poor, at least in the short term) than the total urban population variable. Again, the difference between the standardized coefficients of the total urban population β and of the urban nonpoor population β" provide a sense of the magnitude of the *location* effects.

TABLE 1. The Effects of Urbanization on Rural Poverty, OLS

	(1)	(2)	(3)	(4)	(5)
	Rural poverty (headcount ratio)				
Dep. variable					
Period	1983–99	1983–99	1983–99	1983–93	1983–93
Urban pop.$_{t-2}$ (millions)	−0.074***	−0.068***		−0.101	−0.109*
	(0.024)	(0.023)		(0.063)	(0.058)
Urb. pop. nonpoor.$_{t-2}$ (mil.)		−0.047***			
			(0.017)		
Rural pop. (millions)	−0.010	−0.004	−0.004		0.012
	(0.015)	(0.014)	(0.014)		(0.020)
Scheduled caste (share)		0.338	0.349		0.374
		(0.290)	(0.289)		(0.475)
Rural pop 15–34 age (share)		−3.456***	−3.458***		−3.189**
		(0.940)	(0.944)		(1.478)
Rural lit 15–34 (% in 15–34)		−0.732	−0.740		−3.757***
		(0.654)	(0.657)		(0.938)
Rural lit 15_34 x post-1993		0.241***	0.243***		0.012
		(0.082)	(0.083)		(0.020)
Urban poverty (hc. ratio)		0.316***	0.300***		0.382***
		(0.063)	(0.064)		(0.104)
Observations	973	973	973	667	667
R-sq. (within)	0.647	0.681	0.681	0.619	0.671
No. of districts	355	355	355	355	355

Source: Authors' analysis of data described in appendix S (available online at http://wber.oxfordjournals.org).

Note: All specifications include district and state-year fixed effects. Robust standard errors (Huber-White method) clustered at the district level in parentheses; *significant at 10%; **significant at 5%; ***significant at 1%; all explanatory variables are lagged two years except for urban poverty (contemporaneous).

RESULTS

The descriptive statistics for the main variables used in the analysis are presented in table S2.1 (available online at http://wber.oxfordjournals.org). Table 1 presents the results of specifications (2) and (2′) using fixed effects estimation. Our dataset includes observations of 363 districts for three different time periods: 1983, 1993, and 1999. We run the regressions applying a two-year lag to the measure of the urban population and to the other demographic controls for two primary reasons. First, the two-year lag reduced the risk of potential simultaneity bias. Second, the two-year lag allows us to limit the use of interpolation for the Census variables (both population and socio-demographic variables), which are recorded in 1981, 1991 and 2001, to the last period (1999) only.[21] The standard errors are robust to heteroscedasticity (using the Huber-White correction) and are clustered at the district level.

21. In any instance, the results are not sensitive to the change in the time lag (i.e., applying a one-year lag as well) (results available upon request).

The estimate in column 1 indicates that the growth of the urban population exerts a significant poverty-reducing effect on the rural areas in the same district. This result is obtained by including district and state-year effects as well as the total rural population and should capture the overall effect of urbanization on rural poverty. Column 2 adds the set of controls, which, as we have argued above, should capture most of the *location* effects of urbanization. The coefficient of the urban population is virtually unaffected by the inclusion of these controls, suggesting that most of the poverty-reducing effect of urban growth is due to *economic linkage* effects. This finding is consistent with the evidence that rural-urban migrants (including those within the same district) enjoy, on average, similar, albeit slightly higher, expenditures and education levels than the rural stayers in India (Singh 2009). This similarity is also the case in Tanzania (Beagle and others 2011). Even if our controls imprecisely capture the *location* effect of urbanization on rural poverty, this evidence suggests that our estimation would tend to underestimate the absolute magnitude of the *economic linkage* effect of urbanization on rural poverty.

This magnitude over the 1983–1999 period is not particularly strong, according to our estimate. An increase in the district's urban population of 200,000 (a 43 percent increase from the mean value) reduces the poverty rate by approximately 1.3 percentage points, on average. Over the period of analysis, the rural poverty rate decreased by approximately 20 percentage points, and the urban population increased by 400,000 in the average district. Therefore, urban growth is responsible for approximately 13 percent of the overall reduction in rural poverty in India during the 1983–99 period.

When we use the nonpoor urban population as the main regressor, the β_1 coefficient decreases by approximately one-third vis-à-vis the total urban population coefficient (column 3). However, when considering the urban nonpoor effect in proportionate terms, the two effects are basically identical, providing further confirmation that the *economic linkage* effects of urbanization on rural poverty drive the overall poverty-reducing impact of urbanization in rural areas.

The signs of the controls are as expected, except for the positive coefficient of the share of literates in the last period (i.e., post-1993), when a higher incidence of literates in the most productive part of the rural labor force was associated with higher levels of rural poverty. A higher share of young adults in the rural population decreases rural poverty, whereas a higher presence of scheduled caste increases it (although not significantly). This result suggests that the direct effect of the young adult population on poverty prevails over the indirect effect, which captures the rural-urban migration of young adults. The inclusion of the controls does not significantly change the urban population coefficient.

One possible concern with these results is that the demographic controls, including the urban population, are interpolated for the last period. To address this issue, we check the robustness of our results when we restrict the analysis to the first two periods covering the 1983–1993 time interval because no interpolation is needed in this case. This analysis is interesting in its own right

TABLE 2. The Effects of Urbanization on Rural Poverty, OLS, Further Robustness

	(1)	(2)	(3)	(4)	(5)
Dep. variable	Rural poverty (headcount ratio)				
Period	1983–99	1983–99	1983–99	1983–99	1983–99
Urban pop.t-2 (millions)	−0.077***		−0.068***		−0.039
	(0.025)		(0.024)		(0.057)
Urban share $_{t-2}$		−0.051			
		(0.157)			
Urban pop. of bord. districts (mln)			0.007		
			(0.057)		
Urban pop. cities >20 k (millions)$_{t-2}$				−0.053**	
				(0.026)	
Urban pop.t-2 (millions)					−0.007
					(0.010)
Basic controls	YES	YES	YES	YES	YES
Agricultural productivity	YES	NO	NO	NO	NO
Observations	762	973	965	961	973
R-sq. (within)	0.669	0.679	0.681	0.686	0.681
No. of districts	274	355	349	355	355

Source: Authors' analysis of data described in appendix S (available online at http://wber.oxfordjournals.org).

Note: All specifications include district and state-year fixed effects. Robust standard errors (Huber-White method) clustered at the district level in parentheses; *significant at 10%; **significant at 5%; ***significant at 1%; all explanatory variables are lagged two years except for urban poverty (contemporaneous). The basic controls included are Rural pop. (millions), Scheduled caste (share), Rural pop 15–34 age (share), Rural lit 15–34 (% in 15–34), Rural lit 15_34 x post-1993, Urban poverty (hc. ratio).

because it focuses only on the preliberalization period. Overall, the effect of urbanization on rural poverty is slightly stronger over this period than over the entire period (columns 4 and 5), although the difference in the coefficients is not statistically significant. Again, the bulk of the effects appear to be driven by *economic linkage* effects (cf. column 4 vs. 5). Furthermore, both the share of young adults in the rural population and the share of literates in the young adult population are associated with a reduction of rural poverty. This association supports the hypothesis of a differential impact of literacy on rural poverty over time: it is poverty reducing until 1993 and then is poverty increasing.

OLS: Further Robustness

The analysis thus far does not completely account for other sources of potential bias in the coefficient of interest β'. Table 2 presents the results of a battery of tests to check the robustness of the results to a number of these possible concerns.[22]

22. We show these tests only for the entire period (1983–99) and for the specification with the total urban population, but the results of these tests are analogous for the first period (1981–93) and the nonpoor urban population (results available upon request).

A first potential source of bias is related to the omission of agricultural productivity. As argued by some of the literature on structural transformation, agricultural productivity growth can drive both urbanization and rural poverty (e.g., Matsuyama 1992; Michaels and others 2012). To address this issue, we add the measure of agricultural land productivity to the list of controls in column 1. The variable is constructed as the sum of the total value of 22 different crops produced in a given district divided by the area cultivated with these crops. To estimate the value, we use India-wide crop-specific prices instead of district prices to minimize data gaps (of which there are several for the latter) and the potential endogeneity of the districts' prices to rural poverty.[23] The variable is lagged one year because the simultaneity bias should not be an issue in this case, but a contemporaneous specification is not possible because of the lack of data for 1999. The main results are robust to the inclusion of this measure, which turns out to be not significant.[24]

Second, a question remains regarding the extent to which the impact on rural poverty can be attributed to composition (urban versus rural) rather than simply to a scale effect. To check for this, we substitute the urban population with the share of the urban population in the total population as the main regressor (column 2). This specification is close to Ravallion and others (2007) and yields a negative, although not significant, urban coefficient. This result suggests that the (*economic linkage*) effects of urbanization on rural poverty are primarily driven by a scale effect rather than a composition effect.

Third, until now, the estimation strategy has made the implicit assumption that the towns in a district are centrally located. However, in practice, they are neither centrally located nor evenly distributed within a district. This assumption could therefore lead to a biased estimation. For example, if a district has a large city just outside of its border, the relationship between rural poverty and urbanization within this district could simply be driven by the population growth in the city. Although we include district-level fixed effects and state-specific time effects in the regressions, this source of omitted variables may persist because of the growth of the city in the surrounding district over time. To address this concern, we add to the set of controls a spatially lagged urbanization variable, the average of the urban population of the contiguous districts (column 3). The results are once again unaffected, and this extra control is not significant.

Fourth, a further possible bias of our analysis may be due to small villages that upgrade to towns in the census definition. To the extent that these

23. We are aware that the quantity of agricultural goods produced in a district might also be endogenous to rural poverty even after controlling for a district fixed effect. However, this concern should not be significant for two reasons: first, the degree of endogeneity should be small given the large number of exogenous factors that affect agricultural productivity (e.g., climate); second, the variable is an accessory control that enters only a few specifications with almost no effect.

24. This nonsignificant result is due to the negative and significant effect of the total cultivated area, that is, the denominator, which compensates for the negative effect of the total production value, that is, the numerator (results available upon request).

growing villages are systematically located in rural areas where poverty is decreasing (increasing) for reasons independent of urbanization, we might detect a spurious negative (positive) effect for the urban population on the poverty share. We therefore reestimate the models, excluding from the urban population the variable towns with fewer than 20,000 inhabitants, the size category that would contain most of the 'upgraded villages.' Again, the results of this regression are extremely similar, although slightly less precise (column 4).

Finally, we might think that the effect of the urban population on rural poverty may be nonlinear. In this case, our model would be misspecified. We test whether nonlinearity is the case by adding the square of the urban population as an explanatory variable whose coefficient, however, is close to zero and not significant (column 5). We also attempt different specifications, substituting the urban population variable with various variables corresponding to the sum of the urban population by the size class of the towns (we attempt a number of different size classifications). Again, all of these additional variables have statistically nonsignificant coefficients (results not shown but available from the authors upon request), leading us to conclude that the linear approximation is substantially adequate to identify the phenomenon under scrutiny.

IV Estimation

Despite addressing these concerns, the estimation of (2) could still be biased to the extent that the relationship between rural poverty and the urban population is characterized by reverse causation; the conditions in the rural sector affect urbanization, or there is a correlation of unobserved variables with the variable of interest.

In particular, we are concerned that rural poverty could drive rural-urban migration. Rural poverty could either act as a push factor (i.e., poorer people migrate in search of an escape from poverty), or in the presence of the high fixed costs of migration, it can act as a restraint to migration. If the former case dominates (i.e., poverty is primarily a push factor), the coefficient β' in (2) would have a downward bias, whereas the opposite is true if the latter effect of poverty on migration dominates. The findings by Ravallion and others (2007) that associate global rural-urban migration with a large reduction in the number of rural poor lends some credit to the prevalence of the former case. Kochar (2004) also provides indirect support for this hypothesis, showing that in India, landless households have the highest incidence of rural-urban migrants among rural households.[25]

We resort to an IV estimation to address the endogeneity bias. To that end, we need at least an additional variable to act as a valid instrument, one that is correlated with the district urban population and exogenous to the poverty-

25. His finding emerges in the context of the response of rural schooling decisions to the possibility of employment in urban areas, which tends to be the largest amongst landless households.

induced rural-urban migration flows. We identify three variables that could plausibly satisfy both conditions and are candidates for valid instruments.[26]

The first variable is based on the fixed coefficient approach (Freeman 1980; Card 2001; Ottaviano and Peri 2005) and uses national levels of the urban population and the lagged values of its distribution across districts. This instrument builds on the interaction between two sources of variation, which are exogenous to changes in the local characteristics influencing urban population growth at the district level: the initial distribution of urban population across districts and the national trend in urban population. Similar to Card (2001), Ottaviano and Peri (2005), and Cortés (2008), we define the instrument for district d in year t as the share of the urban population of district d in the total Indian urban population in 1971 multiplied by the total urban population in India at time t:

$$\hat{P}^{U}_{dt} = \frac{P^{U}_{d1971}}{\sum_{d} P^{U}_{d1971}} * \left(\sum\nolimits_{\mathrm{d}} \mathrm{P}^{\mathrm{U}}_{\mathrm{dt}}\right). \tag{3}$$

The predicted measure defined in (3) conveys the size of the urban population in each district that would have been observed if the distribution of the urban population across districts had not changed since 1971. The fact that the initial urban population distribution is referred to 10 years before the beginning of our analysis reinforces its exogeneity to changes in the district urban populations during the period under scrutiny. In a cross-sectional setting, this exogeneity would not necessarily hold true; the unobserved structural factors driving the urban population dynamics before 1971 could also affect rural poverty in later periods. In our panel analysis, however, all of the regressions include the district fixed effects, which absorb all of the time-invariant factors at the district level and therefore the long-term determinants of urban growth. The hypothesis of serial correlation in the rural poverty variable is also unlikely because India underwent many political, social, and economic changes in the 1970s and 1980s, which make it extremely unlikely that the district-specific dynamics in the 1950s and 1960s, for example, would be serially correlated with the rural poverty dynamics in the 1980s and 1990s (after conditioning on state-year fixed effects). However, the contemporaneous India-wide trend in urban population is safely exogenous to the district-specific changes in rural poverty.

Moreover, because the strategy adopted for the first instrument is not immune to criticism (e.g., Cortés 2008), we also employ two other instruments. The first additional instrument is the number of people who migrate to the urban areas of the district from states other than the state where the district is located. It is plausible to assume that rural poverty or the other push factors

26. It is worth noting that the IV estimation can also correct eventual biases arising from errors in the measurement of the urban population. This correction occurs if the measurement error of the instrument and that of the instrumented variable are independent.

for migration in the other states are uncorrelated with the district characteristics once we control for the state-year fixed effects. At the same time, the number of migrants coming to district towns from other states is part of the urban population of the district and thus has a positive association with our main explanatory variable. Nevertheless, a concern about the exogeneity of the instrument could arise from the fact that, within a given district, the time-varying urban pull factors can be correlated with time-varying rural pull factors, and the latter can, in turn, be correlated with rural poverty. For example, migration from other states to the urban areas of one district could be driven by an unobserved positive shock to the entire district, such as an increase in government funding. This shock would help to reduce poverty throughout the district, effectively invalidating the exclusion restriction of our instrument.

However, the first (and second) stage of the IV estimation includes all of the controls listed in the ordinary least squares (OLS) specification, and some of them are good proxies for the pull factors of migration to the district rural areas: the measure of agricultural productivity, the demographic characteristics of the rural population, and the interaction of time and state fixed effects. To further address these concerns, we add the number of migrants from other states to the *rural* areas as a control. This variable should capture any district or rural district shocks driving both migration and poverty. The results with this additional control are nearly identical to the others.[27] Therefore, we can assume that the second IV captures the effect of migration to district cities from other states, *conditioning out* the pull factors of district rural areas.

The last instrument is based on the recognition that the urban areas that are relatively more specialized in tradable sectors are more likely to reap the benefits of Indian economic liberalization, which greatly facilitated trade within and outside of India (Aghion and others 2008; Topalova 2010). Therefore, the cities that specialized in the tradable sector (proxied by manufacturing) before the liberalization shock were more likely to experience a positive trade shock, leading to faster population growth (primarily through immigration). Therefore, we develop an additional instrument based on the interaction of the manufacturing share in urban employment in 1981 with a postliberalization dummy (equal to one for all the years following 1993). The validity of the instrument is further reinforced by the inclusion of the rural manufacturing share in the control set. The rural manufacturing share controls for the possibility that the instrument could reflect the economic structure in the district's rural area, thus capturing some of the direct effect of liberalization on rural poverty.

Columns 1 to 5 of table 3 present the results of the estimations using these instruments. Because the inclusion of the three instruments together leads to a weaker first stage, we included the strongest ones—migrants and predicted

27. Not shown here; results available upon request.

TABLE 3. The Effects of Urbanization on Rural Poverty, IV Estimation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Rural poverty (headcount ratio)						
Dep. variable							
Period	1983–99	1983–99	1983–99	1983–99	1983–93	1983–93	1983–93
Urban pop. (millions)$_{t-2}$	−0.120***	−0.125***	−0.136**		−0.231**		
	(0.037)	(0.046)	(0.056)		(0.108)		
Urb. pop. non poor.$_{t-2}$ (mil.)				−0.089***			
				(0.032)			
Urban pop. (millions)$_{t}$						−0.251	
						(0.197)	
Urban pop. (millions)$_{t+10}$							−0.075
							(0.106)
Basic controls	NO	YES	YES	YES	YES	YES	YES
Observations	926	926	846	926	620	628	628
R-sq. (within)	0.006	0.101	0.111	0.101	0.135	0.189	0.195
Nr. of districts	311	311	284	311	310	314	314
First Stage							
Migrants from other states	3.433***	3.337***		3.796***	3.307***	3.241***	2.342***
	(0.488)	(0.442)		(0.522)	(0.444)	(0.444)	(0.285)
Predicted urban population	1.196***	1.133***	1.289***	1.761***	1.127***	1.142***	1.257***
	(0.153)	(0.146)	(0.184)	(0.207)	(0.147)	(0.150)	(0.247)
Manuf. Shr$_{1981}$ x post$_{1991}$			0.816**				
			(0.395)				
Kleibergen-Paap Wald F statistic	53.38	75.91	41.13	80.62	53.31	51.95	63.31
Hansen J stat.	0.0875	0.0639	0.0448	0.286	0.524	0.635	0.007
Chi-sq(1) P-val.	0.767	0.800	0.832	0.593	0.469	0.426	0.933

Source: Authors' analysis of data described in appendix S (available online at http://wber.oxfordjournals.org).

Note: All specifications include district and state-year fixed effects. Robust standard errors (Huber-White method) clustered at the district level in parentheses; *significant at 10%; **significant at 5%; ***significant at 1%; all explanatory variables are lagged two years except for urban poverty (contemporaneous) and the IVs of columns 6 and 7, which are contemporaneous to the endogenous variable (urban population). The basic controls included are Rural pop. (millions), Scheduled caste (share), Rural pop 15–34 age (share), Rural lit 15–34 (% in 15–34), Rural lit 15_34 x post-1993, Urban poverty (hc. ratio). In columns 6 and 7, the dependent variables for the years 1981 and 1991 are linearly extrapolated on the values for 1983 and 1987.

population—in all regressions except column 3, in which the migration instrument is substituted with the postliberalization instrument. In all cases, the first-stage coefficients substantiate the power of the instruments, with a Kleibergen-Paap Wald F-statistic well above the confidence threshold of the Stock and Yogo (2005) test for weak instruments. Furthermore, the test of the overidentifying restrictions (Hansen's J, reported in the last two rows of table 3) supports the exclusion restriction assumptions for all IV specifications. Analogous to OLS, the standard errors in the IV estimations are robust and clustered at the district level.[28]

The results from the second-stage regressions confirm those of the OLS regression, including the significant reducing impact of urbanization on rural poverty and the fact that most of this impact is driven by *economic linkage* effects. The main difference between these regressions concerns the absolute size of the urban population coefficient, which is almost twice as large as the OLS estimate (and over twice as large in the case of the 1983–93 period). According to the IV estimation, an increase in the urban population of 200,000 reduces rural poverty by between 2.4 and 2.6 percentage points, and urban growth was responsible for approximately one-quarter of the total rural poverty reduction during the 1983–99 period. The larger urban coefficient in the IV estimation is consistent with our theoretical expectations. An attenuation bias in the OLS regression could be due to a favorable shock in rural areas, which reduces rural poverty, rural-urban migration, and thus the urbanization rate. If not controlled for, this mechanism provides a source of downward bias in the absolute size of the coefficient. The larger IV coefficient could be due to measurement error in the endogenous variable.

Falsification Test

To further support the validity of our IVs, in columns 6 and 7 of table 3, we report a falsification test based on the assumption that we should not find an effect of *future* urban population on rural poverty in the two-stage specifications. On the contrary, if the coefficient of the future urban population is significant, this implies a district-specific trend that is correlated with both rural poverty change and urbanization, which makes the IV estimates inconsistent.

Such a test, however, cannot be directly implemented because the available urban population is lagged two years with respect to rural poverty in the data. As a consequence, there is a partial temporal overlap of the two variables, even if the urban population is referred to the subsequent period (e.g., rural poverty in 1983–1993 regressed on the urban population in 1991–2001). Therefore, we use the contemporaneous variable for the years 1981 and 1991 (plus 2001

28. To obtain the covariance matrix of orthogonality conditions of full rank, which allows the calculation of clustered standard errors, year-state dummies are "partialled out," and their coefficient is not calculated. By the Frisch-Waugh-Lovell theorem, in IV, the coefficients for the remaining regressors are the same as those that would be obtained if the variables were not partialled out (Baum and others 2008).

for the future urban population), and we extrapolate the poverty data for those years using the 38th round and the 43rd round of the NSS survey, conducted in 1983 and 1987, respectively. In this way, we are able to exploit the variation in rural poverty that does not overlap in time with the variation in the urban population. In all other aspects, including the control variables and instruments, the regressions are identical to those reported in column 5 of table 3.

The results from the contemporaneous specification are qualitatively similar to those shown in column 5, with an even larger coefficient (−0.25), which is significant at the 80 percent level (the lower significance is probably due to the interpolation, which adds some noise to the dependent variable). When we substitute the urban population with the *future* values (10 years later), the coefficient is much smaller (−0.07), with standard errors that are 50 percent higher than the coefficient value (0.11). This finding supports the assumption that the instruments are exogenous to a district-specific trend, potentially affecting both urban population growth and rural poverty dynamics.

First differences estimation

Given the importance of rigorously addressing the potential endogeneity bias in this type of analysis, we employ a different instrumentation strategy. We conduct an analysis of first differences instead of through fixed effects, regressing the change in rural poverty on the change in the urban population and the other control variables. In this way, we are able to exploit the determinants of urban growth as instruments. In particular, we use two determinants that have proven to be important in the literature (Glaeser and Shapiro 2001; Glaeser and Saiz 2004): the share of manufacturing employment in the urban adult population in 1971 and the historical (1931) urban population density. Manufacturing employment has been shown to be an important determinant of subsequent urban growth in societies at relatively early stages of industrialization, such as the United States in the first part of the 20th century (Glaeser and others 1995). Manufacturing specialization became negatively associated with urban growth as economies moved away from industries and toward tertiary sectors, as occurred in the United States at the end of the last century (Glaeser and Shapiro 2001). It is plausible that India, during the period of our analysis, could be considered an economy in the early days of the industrialization process, comparable to the United States in the first half of the 20th century. Hence, it could be supposed that manufacturing employment would be associated with subsequent urban growth in the Indian context. As we will see, these hypotheses are strongly confirmed by the data.

The second instrument in the first-difference estimation is the density of the urban population in 1931. The intuition here is that in a period of widespread urbanization across India, some districts were urbanizing relatively faster because their physical geography (e.g., terrain slope, water sources, railroads) was more suitable for urbanization. Because the effects of physical geography on urbanization are likely to be persistent over time and because these geographic

TABLE 4. The Effects of Urbanization on Rural Poverty, First Differences Estimation

	(1)	(2)	(3)	(4)
Method	OLS	OLS	IV	IV
Dep. Variable	$\Delta_{1993-83}$ Rural poverty (headcount ratio)			
$\Delta_{1991-81}$ Urban pop. (millions)	−0.089	−0.109*	−0.288**	−0.337**
	(0.061)	(0.060)	(0.125)	(0.141)
Basic controls	NO	YES	NO	YES
Observations	375	312	375	312
R-sq.	0.399	0.494	−0.028	0.095
First Stage				
Manuf. share in urb labor force in 1971			1.270***	1.230***
			(0.225)	(0.266)
Urban density in 1971			0.006***	0.004***
			(0.002)	(0.001)
Kleibergen-Paap rk Wald F statistic			20.62	14.34
Hansen J stat.			1.468	2.657
Chi-sq(1) P-val.			0.226	0.103

Source: Authors' analysis of data described in appendix S (available online at http://wber.oxfordjournals.org).

Note: All specifications include district and state-year fixed effects. Robust standard errors (Huber-White method) clustered at the district level in parentheses; *significant at 10%; **significant at 5%; ***significant at 1%; basic controls include the share of scheduled caste in rural population; the share of rural population aged 15–34 years; the share of literates in rural population aged 15–34 years, and the latter interacted with a post-1993 dummy. Urban population is instrumented through the number of urban immigrants from states other than where the district is located and through the level of urban population predicted using the district's share of the national urban population in 1971 and the changes in the national urban population.

factors do not change over time and were important for historical urbanization, the historical density of urban population is a good proxy for them.

Table 4 reports the results of the estimation of first difference. The results are strictly comparable to the fixed effects results for the 1983–93 period. The size of the IV urban population coefficient triples vis-à-vis its OLS counterpart (column 3 and 4 compared with column 1 and 2). This change is reassuringly similar to that experienced by the coefficient in the fixed effect regressions using a different set of instruments. The larger coefficients likely result because the instruments are now slightly weaker, although in the first-stage regression, their coefficients have the expected sign and are both significant. Even in this case, the Hansen test suggests that the instruments are correctly excluded from the second stage.

Unpacking the economic linkage *channels*

So far, we have established that urban growth reduces rural poverty within the same district and that most of this impact is driven by *economic linkage* channels. In this section, we attempt to test the extent to which some of these

channels explain the rural poverty reduction impact of urbanization. In section 2, we identified six main channels: consumption linkages, rural nonfarm employment, urban-rural remittances, rural land/labor ratio, rural land prices, and rural consumer prices. We can construct variables to proxy four of these channels, but the district data on land and consumer prices are not available separately for rural and urban areas. We add these variables to both the baseline (as defined in (2)) and the IV specifications to obtain a sense of the relative importance of these channels.

To capture the consumption linkage effect of urbanization, we use crop-specific shares of cultivated land. We interact this share at the beginning of the period of our analysis with the urban population variable to test the impact of urbanization across districts on the basis of their production dependence on specific crops. This measure relies on the assumption that a district's supply is a good proxy for urban demand. To identify the relevant crops, we first test for the relationship between urbanization and the main crops across Indian districts.[29] Among the crops that display a significant relationship, we then select the relatively large crops, those with a share of total cultivated land above 5 percent throughout the period. These cultivations are most likely to affect rural economic livelihood.

These crops are rice (which covers, on average, 15 percent of the total cultivated land) and pulses (8 percent). These are important expenditure items for Indian households (Subramanian and Deaton 1996). India is the world's largest consumer and producer of pulses and is one of the largest consumers and producers of rice. This choice is reinforced by the results of Jha and others (2005), who show that the rice market is not integrated within India. As shown below, our analysis suggests that urbanization is associated with an increase in the cultivation of rice and a decrease in the cultivation of pulses. Although the latter result is expected (because pulses are considered a relatively poor crop), the former result is surprising because rice is usually substituted with wheat and other food products as income rises. One reason that this substitution might not occur in India is that during the period of our analysis, the consumption linkage of urbanization appears to operate mainly through an income effect (which increases the quantities of the usual crops) rather than a substitution effect.

We use census data to construct the share of rural employment in nonagriculture/nonhousehold activities, which should capture the rural nonfarm employment channel.[30] Again, urban-rural remittance data are not available for

29. For each of the 24 such goods for which we have data, we regress the district's land cultivated with that crop over the urban population, all other control variables and the district and state-year fixed effects.

30. The census divides employment into four broad sectors: agricultural laborers, cultivators, household industry, and other nonagricultural workers. We use the last category to construct the variable, although the subsequent results are also very similar when using the total nonfarm employment share, obtained by adding the household industry category.

the period of this analysis; therefore, we use the number of intradistrict rural-urban migrants as a proxy for the remittance channel. We also include the share of rural-urban migrants in the total rural population to control for the possible influence of rural poverty on migration.

Finally, we use the total cultivated area over the rural population as a direct proxy for the rural land/labor ratio channel. To the extent that urbanization increases the demand for agricultural goods, thus raising the return on land cultivation, this ratio can also capture part of the consumption linkages channel.

Column 1 in table 5 shows the coefficient of the urban population obtained by running specification (2) for the reduced number of observations for which all of the new variables are available. Adding the set of new variables brings the urban population coefficient to zero (column 2).[31] All of the new variables have the expected sign, although rural nonfarm employment has a nonlinear effect on rural poverty. For relatively low shares of rural nonfarm employment (i.e., below 32 percent of total rural employment), increases in this share are associated with increases in rural poverty, whereas the opposite is true when nonfarm/nonhousehold employment is greater than the 32 percent threshold. This finding suggests that rural nonfarm activities tend to be more profitable than agriculture only when nonfarm activities represent a substantial part of the rural economy. When these activities are relatively marginal, they are likely to be confined to petty trading and other low-yield, nontradable services (such as construction and rickshaw pulling). Nonfarm activities appears to be relatively marginal in most Indian districts (in 1999, only 14 percent of the districts had a rural nonfarm share of employment above 32 percent). Most of the nonfarm employment involves casual employment (daily wage) and self-employment, which tend to be associated with lower incomes and lower stability than regular employment (Lanjouw and Murgai 2009). In 1999, the latter represented only one-quarter of the rural nonfarm employment in India and was the category that grew the least during our period of analysis.

The effect of internal rural-urban migrants is negative, as expected, although it is only significant at the 11 percent level, suggesting that a rise in the number of migrants is associated with a reduction in poverty through the remittance channel. Similarly, the cultivated area per rural population, which increases through urbanization, has a statistically significant poverty-reducing effect in rural areas. Finally, the shares of rice and pulse cultivated areas have a positive association with rural poverty (although it is not significant for pulses), consistent with the relatively small margins that are typical of these crops. However urbanization has a higher, albeit only weakly significant, poverty-reducing effect in districts with relatively large rice and pulse cultivations (i.e., consumer linkages are stronger).

31. In fact, the coefficient becomes positive and very small but with a large standard error.

TABLE 5. Disentangling the *Economic Linkage* Effects of Urbanization on Poverty

	(1) Rur pov	(2) Rur pov	(3) $Area_{PC}$	(4) Nonfarm shr	(5) Intra rur-urb mig	(6) Rice area	(7) Pulse area
Urban pop. (millions)$_{t-2}$	−0.068**	0.012					
	(0.028)	(0.057)					
Urban pop. (millions)			0.079*	0.016*	0.013***	19.763**	−15.227*
			(0.042)	(0.010)	(0.004)	(7.991)	(9.050)
Cultivated area per capita $_{t-1}$		−0.071*					
		(0.040)					
Nonfarm non-HH share $_{t-2}$		1.212***					
		(0.376)					
Nonfarm non-HH share $_{t-2}$ squared		−1.918***					
		(0.446)					
Intradistrict rural-urban migrants $_{t-2}$		−1.046					
		(0.673)					
Rice area share$_{t-2}$		0.552**					
		(0.252)					
Rice area x urb. pop. (mln)$_{t-2}$		−0.300					
		(0.209)					
Pulse area share$_{t-2}$		0.169					
		(0.187)					
Pulse area x urb. pop. (mln)$_{t-2}$		−0.644					
		(0.514)					
Basic controls	YES	YES	YES	YES	YES	YES	YES
Other controls	NO	NO	YES	YES	YES	YES	YES
Observations	734	734	746	746	746	746	746
Nr. of districts	270	270	270	270	270	270	270
R-sq. (within)	0.612	0.632	0.570	0.663	0.864	0.399	0.264

Source: Authors' analysis of data described in appendix S (available online at http://wber.oxfordjournals.org).

Note: Dependent variable is rural poverty; all specifications include district and state-year fixed effects. Robust standard errors (Huber-White method) clustered at the district level in parentheses; *significant at 10%; **significant at 5%; ***significant at 1%; basic controls include rural population; the share of scheduled caste in rural population, the share of rural population aged 15–34 years, the share of literates in rural population aged 15–34 years, and the latter interacted with a post-1993 dummy and the urban poverty rate. Other controls include all of the other extra regressors in column (2).

We repeat the same exercise using IV estimation. The results are almost identical for both the urban population and the variables' coefficients.[32]

To substantiate our claim that urbanization reduces poverty through these channels, columns 3 to 7 show the significant effect of the urban population variable on each of these variables. Using the same specification and controls as in column 2, we find that urbanization affects each variable positively and significantly, except for the share of the pulse area, which is affected negatively.

We check the relative contribution of each channel to the poverty-reducing effect of urbanization by adding each variable in turn to the specification in column 1. According to this analysis, approximately three-quarters of the poverty-reducing effect of urbanization is accounted for by consumer linkages. Intradistrict rural-urban migration accounts for less than one-fifth, and the rural land/labor ratio and the rural nonfarm employment account for 4 percent and 3 percent, respectively. The small contribution of the latter channel is particularly surprising because rural nonfarm employment is often an important cause of rural poverty reduction (Berdegue and others 2001; Lanjouw and Shariff 2002). This surprising result could be explained by considering that urbanization affects rural nonfarm employment positively and that the latter has an inverted-U effect on rural poverty in India.

Taken at face value, these results suggest that the other channels that we were not able to properly capture (i.e., rural consumer and land prices) are not likely to account for any of the poverty-reducing effect of urbanization on the rural areas within the district. This implication seems consistent with the ambiguous effects of these channels on rural poverty that were discussed in section 2.

Conclusions

Do the poor in rural areas benefit from the population growth in urban areas? If so, what is the size of the benefit? Despite the importance of these questions, little empirical evidence is available to provide adequate answers. We have attempted to address this gap by analyzing the effects of urbanization on rural poverty. Using data on Indian districts from 1983 to 1999, we find that urbanization has a significant poverty-reducing effect on the surrounding rural areas. The results are robust to the inclusion of a number of controls and to the use of different types of specifications. The results of the IV estimation suggest that the effect is causal and that the failure to control for causality downwardly biases the coefficient of urbanization. We find that an increase in the urban population of 200,000 determines a decrease in rural poverty in the same district of between 1.3 (lower bound) and 2.6 percentage points.

These figures represent between 13 percent and 25 percent of the overall reduction in rural poverty in India over the period. That amount is a substantial

32. Results available upon request.

contribution, but it is lower than the contribution of another important change to the rural sector, i.e., the state-led rural bank branch expansion, which can explain approximately half of the overall decrease in rural poverty in India between 1961 and 2000 (Burgess and Pande 2005).[33] However, the contribution of urbanization to rural poverty reduction is slightly higher than that of another important state rural policy in post-independence India—land reform, which explains approximately one-tenth of the rural poverty reduction between 1958 and 1992 (Besley and Burgess 2000).

Our analysis suggests that the poverty-reducing impact of urbanization occurs through *economic linkage* effects rather than through the direct movement of the rural poor to urban areas. This finding is not surprising given that rural-urban migrants appear to be, on average, less poor and more educated than rural nonmigrants. These *economic linkage* effects of urbanization on rural poverty are accounted for by four channels: consumer linkages (which explain most of these effects), urban-rural remittances, the changing rural land/labor ratio, and nonfarm employment.

These findings have a number of potentially important policy implications. First, they can help to reassess the role of public investment in urban areas for poverty reduction. In fact, it is a popular tenet that investments in developing countries should be concentrated in rural areas to reduce poverty because the poor in developing countries are primarily concentrated there (see, for instance, World Bank 2008). However, investments in rural areas are often onerous because substantial resources are needed to reach a population that is scattered among vast territories. To the extent that urbanization can have substantial poverty-reducing effects on rural areas, urban investments may become an important complement to rural investments in poverty-reduction strategies.

Second, our findings run counter to the popular myth that rural-urban migration may deplete rural areas, causing them to fall further behind. The relatively low rate of urbanization in India may be due to public policies that have not facilitated (and, in certain instances, have even constrained) rural-urban migration (Deshingkar and Start 2003). At the very least, this paper questions the appropriateness of this bias against rural-urban migration.

Although this paper has not addressed the issue of urban poverty, increasing urban populations imply that, in the future, urban poverty may become a main issue in its own right (Ravallion and others 2007). Further research is needed to assess whether the growth of the urban population entails a trade-off between rural and urban poverty reduction.

33. This expansion was spurred by a government policy implemented between 1977 and 1990, which forced Indian banks to open four branches in locations with no bank branches to qualify for a license to open a branch in a location that already had at least one bank.

References

Aghion, P., R. Burgess, S. Redding, and F. Zilibotti. 2008. "The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India." *American Economic Review* 98(4): 1397–1412.

Allen, R. 2009. *The British Industrial Revolution in Global Perspective*. Cambridge: Cambridge University Press.

Aragón, F., and J.P. Rud. 2009. "The Blessing of Natural Resources: Evidence from a Peruvian Gold Mine." *WP 2009-015, Banco Central de Reserva del Perú*.

Atkin, D. 2010. "Trade, Tastes and Nutrition in India." *Yale University Economic Growth Center Discussion Paper No. 986*.

Bairoch, P. 1988. *Cities and Economic Development: From the Dawn of History to the Present*. Chicago: The University of Chicago Press.

Baum, C., M. Schaffer, and S. Stillman. 2008. "IVREG2: Stata Module for Extended Instrumental Variables/2SLS and GMM Estimation." *Statistical Software Components from Boston College Department of Economics*.

Beegle, K., J. De Weerdt, and S. Dercon. 2011. "Migration and Economic Mobility in Tanzania: Evidence from a Tracking Survey." *Review of Economics and Statistics* 93(3): 1010–33.

Becker, G.S., and K.M. Murphy. 1992. "The Division of Labor, Coordination Costs, and Knowledge." *Quarterly Journal of Economics* 107(4): 1137–60.

Berdegue, J.A., E. Ramirez, T. Reardon, and G. Escobar. 2001. "Rural Nonfarm Employment and Incomes in Chile." *World Development* 29(3): 411–25.

Besley, T., and R. Burgess. 2000. "Land Reform, Poverty Reduction and Growth: Evidence from India." *Quarterly Journal of Economics* 117(4): 389–430.

Burgess, R., and R. Pande. 2005. "Do Rural Banks Matter? Evidence from the Indian Social Banking Experiment." *American Economic Review* 95(3): 780–795.

Card, D. 2001. "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration." *Journal of Labor Economics* 19(1): 22–64.

Chandra Mohan Reddy, Y.V.R. 2000. "A Study of Livestock Markets and Marketing of Livestock in Rangareddy District of Andhra Pradesh." MSc Dissertation, Department of Agricultural Economics, Acharya, N.G. Ranga Agricultural University, Hyderabad.

Cortés, P. 2008. "The Effect of Low-Skilled Immigration on U.S. Prices: Evidence from CPI Data." *Journal of Political Economy* 116(3): 381–422.

Das-Gupta, A. 2006. "Internal Trade Barriers in India: Fiscal Check-posts." *South Asia Economic Journal* 7(2): 231–54.

Datt, G., and M. Ravallion. 2002. "Is India's Economic Growth Leaving the Poor Behind?" *Journal of Economic Perspective* 16: 89–108.

Deaton, A. 2003a. "Adjusted Indian Poverty Estimates for 1999–2000." *Economic and Political Weekly* 38(4): 322–26.

———. 2003b. "Prices and Poverty in India, 1987–2000." *Economic and Political Weekly* 38(4): 362–68.

Deichmann, U., F. Shilpi, and R. Vakis. 2008. "Spatial Specialization and Farm-Non-Farm Linkages." *Policy Research Working Paper 4611: The World Bank*.

Deshingkar, P., and D. Start. 2003. "Seasonal Migration for Livelihoods in India: Coping, Accumulation and Exclusion." *ODI Working Paper 220*.

Dercon, S., and J. Hoddinott. 2005. "Livelihoods, Growth, and Links to Market Towns in 15 Ethiopian Villages." FCND discussion papers no. 194, IFPRI.

Dore, R. 1987. *Taking Japan Seriously: A Confucian Perspective on Leading Economic Issues*. Stanford, CA: Stanford University Press.

Duranton, G., and D. Puga. 2004. "Micro-foundations of Urban Agglomeration Economies." In *Handbook of Regional and Urban Economics,* ed. Henderson Vernon and Thisse Jacques-Francois, Vol. 4. Amsterdam: Elsevier.

Ellis, F. 1998. "Household Strategies and Rural Livelihood Diversification." *Journal of Development Studies* 35(1): 1–38.

Fafchamps, M., and F. Shilpi. 2003. "The Spatial Division of Labor in Nepal." *Journal of Development Studies* 39(6): 23–66.

———. 2005. "Cities and Specialization: Evidence from South Asia." *Economic Journal* 115: 477–504.

Fafchamps, M., and J Wahba. 2006. "Child Labor, Urban Proximity, and Household Composition." *Journal of Development Economics* 79(2): 374–97.

Foster, A., and M. Rosenzweig. 2004. "Agricultural Productivity Growth, Rural Economic Diversity and Economic Reforms: India 1970-2000." *Economic Development and Cultural Change* 52(3): 509–42.

Freeman, R.B. 1980. "An Empirical Analysis of the Fixed Coefficient 'Manpower Requirements' Model, 1960–1970." *Journal of Human Resources* 15(2): 176–99.

Fujita, M., P. Krugman, and A.J. Venables. 1999. *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, MA: MIT Press.

Glaeser, E.L., and A. Saiz. 2004. "The Rise of the Skilled City." *Brookings-Wharton Papers on Urban Affairs* 5: 47–94.

Glaeser, E.L., J.A. Scheinkman, and A. Shleifer. 1995. "Economic Growth in a Cross-Section of Cities." *Journal of Monetary Economics* 36(1): 117–43.

Glaeser, E.L., and J. Shapiro. 2001. "Is There a New Urbanism? The Growth of US Cities." *National Bureau of Economic Research Working paper 8357*.

Government of India. 1971. *Census of India 1971*. Delhi.

———. 1981. *Census of India 1981*. Delhi.

———. 1991. *Census of India 1991*. Delhi.

———. 2001. *Census of India 2001*. Delhi.

Hasan, R., D. Mitra, and B.P. Ural. 2007. "Trade Liberalization, Labor-Market Institutions and Poverty Reduction: Evidence from Indian States." *India Policy Forum* 3: 71–122.

Jacoby, H. 2000. "Access to Markets and the Benefits of Rural Roads." *Economic Journal* 110(465): 713–37.

Jacoby, H., and B. Minten. 2009. "On Measuring the Benefits of Lower Transport Costs." *Journal of Development Economics* 89: 28–38.

Jha, R., K.V.B. Murthy, and A. Sharma. 2005. "Market Integration in Wholesale Rice Markets in India." *ASARC Working Papers 2005-03*.

Jha, V. 2008. "Trickle down effects of Inter State Migration in a period of high growth in the Indian economy." Unpublished manuscript.

Kijima, Y., and P. Lanjouw. 2003. "Poverty in India During the 1990s: A Regional Perspective." *World Bank Policy Research Working Paper 3141*.

Kochar, A. 2004. "Urban Influences on Rural Schooling in India." *Journal of Development Economics* 74: 113– 136.

Kuznets, S. 1955. "Economic Growth and Income Inequality." *American Economic Review* 45(1): 1–28.

Lanjouw, P., and A. Shariff. 2002. "Rural Nonfarm Employment in India: Access, income, and poverty impact." *Working Paper Series no 81*. New Delhi: National Council of Applied Economic Research.

Lanjouw, P., and R. Murgai. 2009. "Poverty Decline, Agricultural Wages, and Non-Farm Employment in Rural India: 1983–2004." *Agricultural Economics* 40(2): 243–63.

Lewis, W.A. 1954. "Economic Development with Unlimited Supplies of Labour." *The Manchester School* 22: 139–91.

Matsuyama, K. 1992. "Agricultural Productivity, Comparative Advantage, and Economic Growth." *Journal of Economic Theory* 58: 317–34.

Michaels, G., F. Rauch, and S. Redding. 2012. "Urbanisation and Structural Transformation." *Quarterly Journal of Economics* 127(2): 535–86.

Ottaviano, G.I.P., and G. Peri. 2005. "Cities and Cultures." *Journal of Urban Economics* 58(2): 304–7.

Parthasarathy Rao, P., P.S. Birthal, P.K. Joshi, and D. Kar. 2004. "Agricultural Diversification in India and Role of Urbanisation." *MTID Discussion Paper No*. 77, IFPRI.

Plantinga, A.J., R.N. Lubowski, and R.N. Stavins. 2002. "The Effects of Potential Land Development on Agricultural Land Prices." *Journal of Urban Economics* 52(3): 561–81.

Ravallion, M., S. Chen, and P. Sangraula. 2007. "New evidence on the urbanisation of global poverty." *Population and Development Review* 33(4): 667–701.

Schaffer, M. 2007. "XTIVREG2: Stata module to perform extended IV/2SLS, GMM and AC/HAC, LIML and k-class regression for panel data models." Statistical Software Components from Boston College Department of Economics.

Singh, D.P., 2009. "Poverty and Migration: Does Moving Help?" In *Urban Poverty Report 2009*, ed. A. Kundu. Delhi: Delhi University Press.

Stark, O. 1980. "On the Role of Urban-rural Remittances in Rural Development." *Journal of Development Studies* 16(3): 369–74.

Stark, O., and R.E.B. Lucas. 1988. "Migration, Remittances and the Family." *Economic Development and Cultural Change* 36(3): 465–82.

Stock, J.H., and M. Yogo. 2005. "Testing for weak instruments in linear IV regression." In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. D.W.K. Andrews and J.H Stock. Cambridge: Cambridge University Press.

Subramanian, S., and A. Deaton. 1996. "The Demand for Food and Calories." *Journal of Political Economy* 104(11): 133–62.

Thanh, H.X., D.T. Thu Phuong, N. Thu Huong, and C. Tacoli. 2008. "Urbanisation and rural development in Vietnam's Mekong Delta." *IIED Working Paper* 14.

Topalova, P. 2010. "Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India." *American Economic Journal: Applied Economics* 2: 1–41.

United Nations. 2008. "World Urbanisation Prospects: The 2007 Revision Population Database." Available at http://esa.un.org/unup/, accessed May 23, 2008.

Williamson, J. 1990. *Coping with City Growth During the British Industrial Revolution*. Cambridge: Cambridge University Press.

World Bank. 2008. *World Development Report 2008: Agriculture for Development*. Washington, DC: The World Bank.

Learning versus Stealing: How Important Are Market-Share Reallocations to India's Productivity Growth?

Ann E. Harrison, Leslie A. Martin, and Shanthi Nataraj†

Recent trade theory emphasizes the role of market-share reallocations across firms ("stealing") in driving productivity growth, whereas previous literature focused on average productivity improvements ("learning"). We use comprehensive, firm-level data from India's organized manufacturing sector to show that market-share reallocations were briefly relevant to explain aggregate productivity gains following the beginning of India's trade reforms in 1991. However, aggregate productivity gains during the period from 1985 to 2004 were largely driven by improvements in average productivity. We show that India's trade, FDI, and licensing reforms are not associated with productivity gains stemming from market share reallocations. Instead, we find that most of the productivity improvements in Indian manufacturing occurred through "learning" and that this learning was linked to the reforms. In the Indian case, the evidence rejects the notion that market share reallocations are the mechanism through which trade reform increases aggregate productivity. Although a plausible response would be that India's labor laws do not easily permit market share reallocations, we show that restrictions on labor mobility cannot explain our results. JEL codes: F13, F14, F16, O24, O25.

Over the last two centuries, economists have frequently returned to the question of how nations gain from trade. Early studies focused primarily on aggregate productivity gains driven by interindustry specialization according to

†. Ann E. Harrison is a professor at the Wharton School of the University of Pennsylvania and a Research Associate at the National Bureau of Economic Research; her email address is annh@wharton.upenn.edu. Leslie A. Martin is a lecturer (assistant professor) in the Department of Economics at the University of Melbourne; her email address is Leslie.Martin@unimelb.edu.au. Shanthi Nataraj (corresponding author) is an Associate Economist at RAND Corporation; her email address is snataraj@rand.org. This material is based upon work supported by the National Science Foundation under Grant No. SES-0922332. We thank Mr. M.L. Philip, Mr. P.C. Nirala, Dr. Praveen Shukla, and Mr. M.M. Hasija at the Ministry of Statistics and Programme Implementation for their assistance in elucidating the Annual Survey of Industries data collection process; Rana Hasan, Karl Jandoc, and Steve O'Connell for discussions regarding the firm-level data; Pauline Grosjean and Ben Crost for providing us with their district-level concordances, which formed the basis for ours; and three anonymous referees as well as seminar participants at the IMF, the World Bank, and Georgetown University for their valuable comments. A supplementary appendix to this article is available at http://wber.oxfordjournals.org/.

THE WORLD BANK ECONOMIC REVIEW, VOL. 27, NO. 2, pp. 202–228 doi:10.1093/wber/lhs029
Advance Access Publication November 21, 2012

comparative advantage. In the 1980s, the importance of learning by doing and the role of trade in facilitating the exploitation of economies of scale were emphasized by Paul Krugman and others. Recently, productivity gains associated with the entry of more productive firms and the exit of less productive ones have generated significant interest. A related question is whether trade reforms lead to the redistribution of market share between incumbents with different productivity levels.

This most recent wave of trade theory stresses the importance of market-share reallocations in increasing aggregate productivity following trade liberalization (Bernard, Eaton, Jensen, and Kortum 2003; Melitz 2003). The "new new" trade theory emphasizes gains from trade in the presence of "heterogeneous firms." In heterogeneous firm models, firms of different productivities, sizes, and profit levels coexist. A trade reform that exposes firms to greater competition or enables more firms to sell in export markets will lead less productive firms to exit or to lose market share. In a Melitz (2003) world, the primary sources of productivity gains associated with trade reform are the exit of less productive firms and the expansion of more productive firms.

Although early heterogeneous firm models, such as the model by Melitz (2003), assumed that firms had exogenous, fixed productivity levels, recent research allows for changing productivity within the firm.[1] This may occur when some product lines are discontinued and other product lines are developed. Consequently, trade reforms can lead to changes in within-firm productivity as firms shift their focus to high-productivity products. These theories recall earlier literature that emphasized that trade could improve average productivity within surviving firms.[2]

The policy implications that arise from these explanations for the gains from trade differ significantly. In a world where market share reallocation away from less productive firms matters more than learning or product shifting within the same firm, it is crucial to facilitate firm entry, exit, and downsizing. In a world where there is learning or shifting of product types within a firm, it is also crucial to work within the enterprise to facilitate the learning that accompanies trade reform. However, few empirical studies quantify the relative importance of average productivity gains and gains from market-share reallocations in the wake of major trade liberalization.

In this paper, we use a comprehensive, firm-level dataset to examine the role of market-share reallocations in driving aggregate productivity growth in India's organized manufacturing sector from 1985 to 2004. The organized manufacturing sector in India consists of firms that are registered under sections 2m(i) and 2m(ii) of the Factories Act. All firms with 20 or more employees (10 if a power source is used) are required to register. In 1991, India

1. See Arkolakis and Muendler 2010; Bernard, Redding, and Schott 2010; Bernard, VanBeveren, and Vandenbussche 2010; Bernard, Redding, and Schott 2011; Eckel and Neary 2010; Feenstra and Ma 2007; Mayer, Melitz, and Ottaviano 2011; Nocke and Yeaple 2008.

2. See Corden 1974; Grossman and Helpman 1991; Helpman and Krugman 1985.

embarked on a series of reforms, including the liberalization of trade, licensing, and foreign direct investment (FDI) regulations. As part of the trade liberalization, nontariff barriers were removed from a number of product lines, and tariff levels were gradually reduced from extraordinarily high levels. The licensing regime, which required that firms seek permission from the "License Raj" to enter an industry and to change or expand production, was gradually dismantled. FDI restrictions that prohibited foreign firms from entering some sectors and restricted their participation in others were relaxed.

We begin by measuring aggregate productivity growth in the manufacturing sector and show that there were three distinct phases from 1985 to 2004. During this 20-year period, aggregate productivity (defined as output-weighted, mean firm productivity) grew by nearly 20 percent. From 1985 to 1990, the growth in aggregate productivity was driven by "learning," an increase in unweighted, average firm productivity. This measure of learning captures the change in productivity for the average firm and includes not only changes in productivity among surviving firms but also changes in average productivity that can be attributed to firm entry and exit. In the period immediately following the beginning of the reforms (1991–1994), the "stealing" of market share (the reallocation of market share from less productive to more productive firms) became more important than learning for driving aggregate productivity growth. In the longer run (1998–2004), learning again became more important for aggregate productivity growth, with a minor contribution from stealing (reallocation).

Overall, we find that for the organized manufacturing sector as a whole, market-share reallocations were an important source of productivity growth in the years immediately following the beginning of the 1991 reforms, but not during other periods. In other words, in the Indian case, the contribution of the reallocation of market shares is concentrated at a given point in time. One implication is that trade reforms lead to short-term, one-off market share reallocations. Our results suggest that in the longer term, opening up trade in India had more important effects on average productivity.

Our main results rely on the widely used decomposition suggested by Olley and Pakes (1996). This method identifies changes in average productivity and reallocation but does not disentangle the contributions of survivors, entrants, and exiters. Although firm identifiers are not available for the organized sector data during most of the period that we study, we construct a panel dataset by matching individual firms from one year of the survey to the next. We match firms using beginning- and end-of-period information on capital and other types of stocks, supplemented with other identifying information. This panel allows us to test how our main results change when we employ an alternative decomposition method, suggested by Melitz and Polanec (2010). We find that our results are robust to these different approaches when decomposing the sources of productivity change during the period.

We then use the Olley-Pakes decomposition to examine the extent to which individual policy reforms are associated with industry-level productivity gains.

In particular, we exploit variations in tariff cuts, FDI liberalization, and industrial licensing reforms across industries to examine the contribution of each reform to changes in industry-level total factor productivity (TFP). We find that the average decline in final goods tariffs during the study period implies a 3 percent increase in aggregate productivity, whereas the average decline in input tariffs implies a 22 percent increase in aggregate and average productivity. Moreover, FDI liberalization accounts for a 4.7 percent increase in average productivity. Consequently, the reduction in input tariffs and opening up to foreign investment are the most important policy changes associated with improved firm performance in our sample. The industrial licensing reforms, which promoted internal competition, are associated with productivity gains among large firms and in states and industries that were relatively less exposed to external competition prior to the reforms. We use our constructed panel to show that the trade and FDI liberalizations are associated with increased productivity within firms, even when controlling for unobservable firm heterogeneity. Our results also suggest that, overall, the reforms are associated with average productivity improvements rather than market share reallocation across firms.

Finally, we explore whether our results can be explained by regulatory barriers that prevented market share reallocations, such as restrictive labor laws. Our results suggest that labor laws and a legal framework that prevented firm adjustment cannot explain our findings. We also explore the extent to which the reforms had differential effects in states and industries that were previously exposed to trade and among firms of different initial sizes. We find that delicensing and FDI reforms had larger effects on productivity among firms that were relatively less exposed to trade. These results suggest some substitutability between external and internal competition: where states or industries are not already exposed to trade through proximity to ports or international trade, industrial reforms that promote competition have larger effects.

Our study was motivated by the emphasis of contemporary trade theory on the importance of market-share reallocation in increasing aggregate productivity. Although a number of papers have tested implications of this literature, few of these studies directly examine the impact of trade liberalization on market-share reallocation.[3] Existing evidence on the importance of reallocation in promoting productivity growth is mixed.[4]

Our study also contributes to the substantial body of work examining India's 1991 reforms. Topalova and Khandelwal (2011) establish that reductions in final goods and input tariffs increased productivity among approximately 4,000 large, publicly listed manufacturing firms. Sivadasan (2009) uses a pooled, cross-sectional dataset, but not a panel, for the early years of the

3. See, for example, Arkolakis 2010; Bernard et al. 2003; Bernard, Jensen, and Schott 2006; Berthou and Fontagne forthcoming; Eaton, Kortum, and Kramarz 2011; Helpman, Melitz, and Yeaple 2004; Manova and Zhang 2012.

4. See, for example, Tybout and Westbrook 1995; Pavcnik 2002; Trefler 2004; Fernandes 2007; Menezes-Filho and Muendler 2011.

reforms (1986–1994) and finds that the reduction in final goods tariffs and FDI liberalization increased productivity. He also documents that these reforms were associated with average productivity increases, but not reallocation, in the early 1990s. Nataraj (2011) compares the reactions of the organized and unorganized manufacturing sectors to trade liberalization and finds that although the reduction in final goods tariffs increased productivity significantly in the unorganized sector, the reduction in input tariffs was more important for productivity growth in the organized sector. Aghion, Burgess, Redding, and Zilibotti (2008) find that following the removal of licensing requirements, the number of factories and output in the organized sector increased, particularly in states with relatively less restrictive labor regulations.

Our study contributes to these two strands of the literature in several ways. First, we document that although market-share reallocations were important to overall productivity growth immediately following the implementation of the 1991 reforms, most of the productivity improvements in manufacturing during the period from 1985 to 2004 occurred because average productivity increased. Market-share reallocations, the focus of most of the heterogeneous trade literature, were not as important. One implication is that in the Indian case, theories that emphasize within-firm changes in response to trade policy changes are more relevant.

Second, we tie these different sources of productivity growth to the various reforms. Our constructed panel is the largest, most representative panel of Indian manufacturing firms that covers the period of the reforms, allowing us to isolate within-firm productivity improvements. In contrast to the earlier literature on gains from trade with heterogeneous firms, which assumed exogenous productivity draws at the firm level and emphasized the role of market-share reallocations, we cannot explain the increases in productivity with market-share reallocations using our policy measures. Instead, our constructed panel suggests that trade and FDI reforms raised average, within-firm productivity. One plausible mechanism is that trade reforms led to productivity improvements because firms focused on their most productive goods. However, evidence on this type of effect in the Indian case is mixed: Goldberg, Khandelwal, Pavcnik, and Topalova (2010a) find that lower input tariffs accounted for approximately one-third of the increase in new products created by Indian firms, but the same authors (2010b) test for evidence of product rationalization in the Indian case and reject this possibility.

The rest of this paper is organized as follows. Section I provides a brief background on the Indian reforms; section II describes the data and outlines the construction of the panel of firms; section III discusses the empirical framework and presents results; and section IV concludes.

I. The 1991 Reforms

Prior to 1991, India imposed high tariffs and nontariff barriers on most goods. FDI (foreign ownership) was capped at 40 percent for most industries, and

large manufacturing firms were required to obtain operating licenses. During the 1980s, India removed licensing requirements from approximately one-third of industries but retained its trade and FDI restrictions.

A combination of economic and political shocks created a balance of payments crisis in 1991, and the IMF agreed to assist the Indian government under certain conditions (Hasan, Mitra, and Ramaswamy 2007; Topalova and Khandelwal 2011). Major policy changes, including FDI and tariff liberalization, exchange rate liberalization, and the removal of the requirement for operating licenses in most industries, were announced in July 1991, and many of these policy changes were formalized in India's Eighth Five-Year Plan (1992–97).

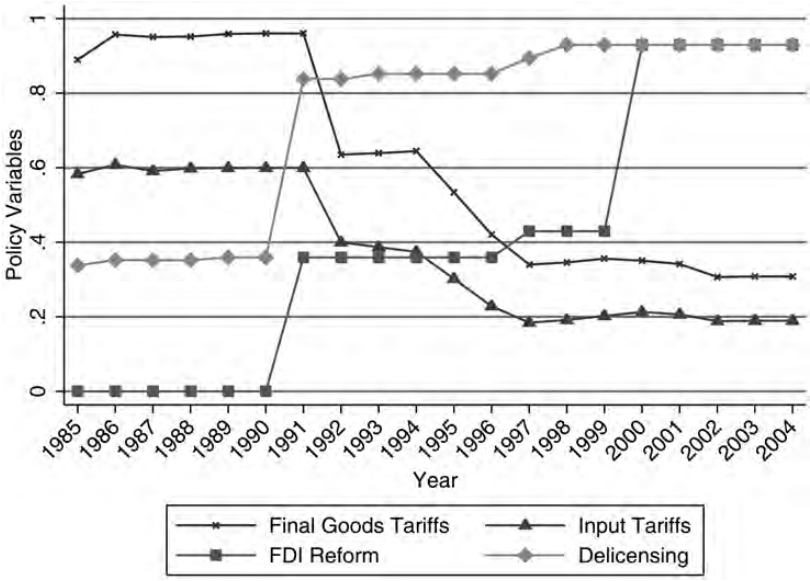
The average final goods tariff rate on manufactured products fell from 95 to 35 percent between 1991 and 1997, and tariffs were harmonized across industries (industries with the highest prereform tariffs received the largest tariff cuts). India continued to lower its tariffs after the Eighth Five-Year Plan ended in 1997, although the reductions were no longer as uniform. Input tariffs also fell significantly during this period. The supplementary appendix (available at http://wber.oxfordjournals.org/) illustrates the tariff changes in more detail (figure S1.1).

In addition, India dismantled its "License Raj" during this period. Under the licensing regime, every firm with more than 50 employees (100 employees where a power source is not used) and a certain amount of assets was required to obtain an operating license. The license specified the amount of output a firm could produce, the types of goods it could make, and a number of other conditions. Approximately one-third of industries were delicensed (the requirement for a license was dropped) in 1985, and most remaining industries were delicensed as part of the 1991 reforms (Aghion et al. 2008). Several additional industries were delicensed during the following decade.

Beginning in 1991, majority FDI shares were also allowed in a number of industries with "automatic" approval (Sivadasan 2009). Approximately one-third of industries were opened to FDI in 1991. A few additional industries were liberalized by 1997, and in 2000, the government indicated that all industries would be eligible for automatic FDI approval, except those requiring an industrial license or meeting several other conditions (figure 1).

The occurrence of most of these policy changes as part of an externally required reform package reduced the likelihood that industries were selected into the reforms on the basis of political factors. Furthermore, to the extent that industries with certain characteristics may have been more likely to be liberalized, we use a fixed-effects estimation strategy that should address any time-invariant characteristics that may have affected selection. However, if the reforms are correlated with prereform trends in industry characteristics, our results may be biased. To evaluate the potential extent of this bias, we examine the correlations between changes during the reforms (1990–2004) and prereform trends for a number of industry characteristics (1985–1989). In the supplementary appendix (table S1.1), we show that there are no statistically significant correlations between prereform trends in industry characteristics and future reforms.

FIGURE 1. Trade, FDI, and Licensing Reforms



Mean values of policy variables from 1985 to 2004. Final goods and input tariff variables are fractions, with one representing an ad valorem tariff of 100 percent. FDI Reform is a dummy variable equal to one if any products within the industry are liberalized and zero otherwise. Delicensing is a dummy variable equal to one if any products within the industry are delicensed and zero otherwise. *Source*: Authors' calculations based on various publications of the Government of India, including the Customs Tariff Working Schedules, the Trade Analysis and Information System database, and Aghion et al. (2008).

Moreover, the supplementary appendix (table S1.9) shows that our results are robust to limiting our analysis to the period ending in 1997. Because the initial reforms were developed in the wake of the 1991 crisis, they are even less likely to be subject to potential selection issues than are reforms in later years.

II. DATA

Our primary dataset consists of firm-level surveys from the Annual Survey of Industries (ASI).[5] We obtained data for all available years between 1985 and 2004. Data were not available for 1995. Furthermore, the way in which input data were collected and made available for 1996 and 1997 made it impossible to construct certain key variables for those two years that were consistent with the other years. Therefore, we restrict our analysis to firm-level data for the

5. The ASI covers accounting years that ended on any day during the fiscal year. The 1985–86 survey (which we refer to as the 1985 survey) indicates the factory's accounting year that ended on any day between April 1, 1985, and March 31, 1986.

remaining 17 years between 1985 and 2004 (1985 through 1994 and 1998 through 2004).

The sampling universe for the ASI is all firms that are registered under sections 2m(i) and 2m(ii) of the Factories Act as well as firms registered under the Bidi & Cigar Workers Act and a number of utility and service providers. These firms constitute the "organized" sector; they account for approximately 80 percent of output but only 20 percent of employment in Indian manufacturing. We only include manufacturing firms in our analysis.[6] Large firms are considered part of the "census" sector and are surveyed every year. Smaller firms are considered part of the "sample" sector and are sampled every few years.[7] In the population-level analysis (but not the panel analysis), we apply the multiplier weights provided by the ASI. Each unit surveyed is generally a factory (plant); however, if an owner has two factories in the same state, sector (census versus sample) and industry, a joint return can be furnished. In the population of firms, fewer than 2 percent of the observations report having more than one factory. We will use the term "firm" to mean one observation in our dataset.

The key variables we construct from the ASI data are output, material input, labor, and capital.[8] We drop closed firms, and we only include firms with positive values of the key variables. To address a few extreme outliers, we also trim the top 0.5 percent of the output and material input values. We deflate output using industry-specific wholesale price indices (WPI) from the Government of India's Handbook of Industrial Statistics. Similarly, we construct material input deflators using the WPI along with India's 1993–94 Input-Output Transactions Table. Labor is measured as the total number of individuals employed by the firm, and capital is measured by deflating the book value of capital by the WPI for machinery. Summary statistics for the population are presented in table 1.[9]

6. Our sample is therefore representative of all manufacturing firms that have registered with each state's Chief Inspector of Factories. All firms that have 20 or more employees (10 or more employees if a power source is used) are required to register. In practice, a significant fraction of ASI firms report fewer than 10 employees. These firms may be registered for various reasons, including the possibility that they formerly had more than 10 employees but shrank, that they plan to grow in the future, and that registering may be a signal to creditors or other business partners.

7. The division between the two sectors depends on firm size. It changed several times from 1985 to 2004.

8. Output includes the ex-factory value of products, the increase in the stock of semifinished goods, and the value of own construction. Material input includes materials and fuel.

9. Sampling weights are applied to the summary statistics in the first column of table 1; hence, the results are representative of the overall organized sector. The second column shows results for the firms that were sampled without applying sampling weights. Because larger firms are surveyed more frequently than smaller firms, the mean and median values of output, capital, material inputs, and labor are much larger in the sampled population than in the estimated population. The term "firm-years" indicates the total number of firm-level observations over all of the years in our dataset, whereas "census-firm-years" indicates the total number of observations of "census" sector firms over all years.

TABLE 1. Summary Statistics for the Firm-Level Data

	Estimated population	Sampled Firms	Panel
Firm-years, no. obs.	1,410,725	580,941	414,074
Firms per year, mean	82,983	34,173	24,357
Census firm-years, no. obs.	277,178	277,178	247,777
Census firms per year, mean	16,304	16,304	14,575
Unique firm series			147,695
Output, mean (million Rs.)	18.7	32.0	41.8
Output, median (million Rs.)	2.6	3.6	5.3
Capital, mean (million Rs.)	6.9	12.8	17.0
Capital, median (million Rs.)	0.4	0.5	0.8
Materials, mean (million Rs.)	12.5	21.1	27.3
Materials, median (million Rs.)	1.9	2.6	3.9
Labor, mean (no. employees)	74.1	133.4	171.9
Labor, median (no. employees)	21	31	43
In panel, as fraction of total in sampled population:			
Output			0.93
Capital			0.95
Labor			0.92
Firm-years > 100 employees			0.94
Firm-years > 200 employees			0.96
Firm-years			0.71
Census firm-years			0.89

Source: Authors' calculations based on ASI data.

Summary statistics for the estimated population (using sampling weights), for the sampled population (not using sampling weights), and for firms that appear for two or more years in the panel. Only open firms with positive values of key variables are included. The term "firm-years" indicates the total number of observations, whereas "census firm-years" indicates the number of observations in the census sector. Mean and median values are averages across all years used in the analysis (1985–1994 and 1998–2004). Output, material inputs, and capital have been deflated to 1985 values and are expressed in millions of rupees. Fractions of output, capital, and so forth that appear in the panel are reported relative to the sampled (rather than the estimated) population.

Creating a Panel

The ASI data provide unique firm identifiers beginning in 1998. It has not previously been possible to track firms prior to 1998 and to follow them during the most significant reform period. We overcame this challenge by matching individual firms from one year of the survey to the next between 1985 and 1998. To construct our panel, we searched for firms that had matching open and close values between one year and the next (e.g., we searched for a match between the close value in 1985 and the open value in 1986) for one of several variables. To link firms for 1995, a year for which firm-level data have not been released, and for other years in which individual firms may not have been sampled, we matched firms on the basis of a number of static characteristics as well as growth projections. We then combined this constructed panel with the

actual panel provided by the ASI from 1998 to 2004. Details on panel construction are provided in the supplementary appendix.

Because we observe each firm's age, we are confident that we can correctly identify survivors and entrants in our panel. However, given the substantial fraction of firms that are not surveyed every year, we are more reserved about our ability to identify exiting firms. The rates of exit that we observe in our panel are significantly higher than the rates that we extrapolate on the basis of the observed distributions of firm age. Therefore, when estimating productivity, we avoid methods that rely on accurately identifying firm exit and instead employ an index number method that is robust to potentially spurious exit.

Summary statistics for the panel (to which we do not apply sampling multipliers) are presented in the final column of table 1. Observations of firms in the "census" sector (indicated by the number of census-firm years) account for 60 percent of total firm-year observations in the panel, 48 percent of firm-year observations in the full sample of firms, and only 20 percent of firm-year observations in the estimated population. The panel should not be considered representative of the population; rather, it is a selection of relatively large firms. Nonetheless, 71 percent of firm-year observations appearing in the sample, representing 93 percent of total deflated output over the entire period and 92 percent of the labor force, are captured for at least two years in the panel (table 1, bottom rows).

A key contribution of our panel is that we are confident in the firms that we are able to match over time. This contribution allows us to examine the impacts of the reforms on within-firm learning. These impacts have not previously been examined for such a large subsample of the organized manufacturing sector.

Policy Variables

We matched applied tariff data from the Government of India's Customs Tariff Working Schedules and the Trade Analysis and Information System database with India's three-digit National Industrial Classification (NIC-87) codes using the concordance developed by Debroy and Santhanam (1993). We then calculated average final goods tariff rates within each of approximately 140 NIC codes. We also calculated input tariffs using India's Input-Output Transactions Table, following the method suggested by Amiti and Konings (2007).[10]

To capture the effects of the delicensing and FDI reforms, we used data from Aghion et al. (2008), supplemented by information from Press Notes from the Ministry of Commerce and Industry. Both the delicensing and FDI reform variables are equal to one if any products in a three-digit industry have been liberalized and are equal to zero otherwise. Figure 1 shows changes in these policy variables over time.

10. For example, if the footwear industry derives 80 percent of its inputs from the leather industry and 20 percent from the textile industry, the input tariff for the footwear industry is 0.8 times the final goods tariff for the leather industry plus 0.2 times the final goods tariff for the textile industry. In our baseline measure of input tariffs, we use both traded and nontraded inputs, assigning tariff rates of zero to nontraded inputs.

III. Empirical Framework and Results

We measure TFP for firm i in industry j at time t using a chain-linked, index number method. This measure is well suited for our data because approaches such as semiparametric methods (Olley and Pakes 1996, for example), which rely on panel data, cannot be used for the population of firms. Although we explore the robustness of our results to using a panel with linked firms over time below, in much of the analysis, we need an approach that allows us to exploit the population of firms. Another problem with the Olley-Pakes methodology and other semiparametric approaches is that they require an accurate identification of exit. Therefore, we employ the method suggested by Aw, Chen, and Roberts (2001):

$$TFP_{ijt} = \left(q_{ijt} - \overline{q_{jt}}\right) + \sum_{r=2}^{t}\left(\overline{q_{jr}} - \overline{q_{jr-1}}\right) - \left[\sum_{z=1}^{Z}\frac{1}{2}\left(\zeta_{ijt}^{z} + \overline{\zeta_{jt}^{z}}\right)\left(z_{ijt} - \overline{z_{jt}}\right) + \sum_{r=2}^{t}\sum_{z=1}^{Z}\frac{1}{2}\left(\overline{\zeta_{jr}^{z}} + \overline{\zeta_{jr-1}^{z}}\right)\left(\overline{z_{jr}} - \overline{z_{jr-1}}\right)\right] \quad (1)$$

where q_{ijt} is the log of output, ζ_{ijt}^{z} is the revenue share of input z, and z_{ijt} is the log of input z. A firm's TFP is the deviation of its output from average output in that year and the difference in the average output in that year from the base year, minus the deviation of the firm's inputs from average inputs in that year, along with the difference in average inputs in that year from the base year. Inputs include labor, capital, and material inputs, which are measured and deflated as discussed in section II. Bars over variables indicate average values within a particular industry and year. Revenue shares for labor and material inputs are calculated as the share of each input in total revenue. Capital's revenue share is assumed to be one minus the sum of the other two shares.

Decomposing All-India TFP Growth

We begin by examining productivity changes for the population of firms from 1985 to 2004. To do so, we first calculate aggregate TFP in year t, Φ_t^{AGG}, as the sum of each firm's productivity, ϕ_{it}, weighted by its market share, ψ_{it}. Olley and Pakes (1996) show that this measure of aggregate TFP can be decomposed into two components:

$$\Phi_t^{AGG} = \sum_i \psi_{it}\phi_{it} = \overline{\phi_t} + \sum_i \left[\psi_{it} - \overline{\psi_t}\right]\left[\phi_{it} - \overline{\phi_t}\right] = \Phi_t^{U} + R_t \quad (2)$$

where $\overline{\phi_t}$ and $\overline{\psi_t}$ are unweighted average productivity and market share, respectively. The first component, Φ_t^{U}, is unweighted average productivity. The

second component, R_t, measures the covariance between firm productivity and market share. Changes in this measure represent a reallocation of market share between firms with different productivity levels.

We first construct these measures at the all-India level. To make the results representative of the population of firms and consistent over time, we premultiply each observation by the sampling weight provided by the ASI. Furthermore, to make the results more comparable with our later regression results, we only consider firms in state-industry groups (collections of firms in a particular state and three-digit industry) that exist over the entire period.

Following Pavcnik (2002), we normalize productivity values to zero in 1985. Hence, changes in productivity levels can be interpreted as growth since 1985. Between 1985 and 2004, aggregate productivity grew by 18 percent (figure 2). This increase in productivity implies an annual increase of slightly less than 1 percent per year, within the range found in previous studies.[11]

When we consider the time period as a whole, nearly all of this increase can be attributed to growth in average productivity, rather than reallocation. However, figure 2 suggests that there are three distinct phases between 1985 and 2004. First, from 1985 to 1990, average productivity rose by over 8 percent, while the reallocation component fell by more than 6 percent, indicating that more productive firms lost market share relative to less productive firms. Beginning in 1991, this trend was reversed: average productivity fell, whereas reallocation productivity rose sharply. By 1998, however, average productivity improvements were once again the more important driver of aggregate productivity growth. Reallocation productivity remained at approximately the level it reached between 1992 and 1993, but it did not increase further.

Our results suggest that market-share reallocations played an important role in aggregate productivity growth, but only during the few years immediately following the implementation of the 1991 reforms. Over the longer time horizon, average productivity improvements remained more important for explaining the increase in aggregate TFP.

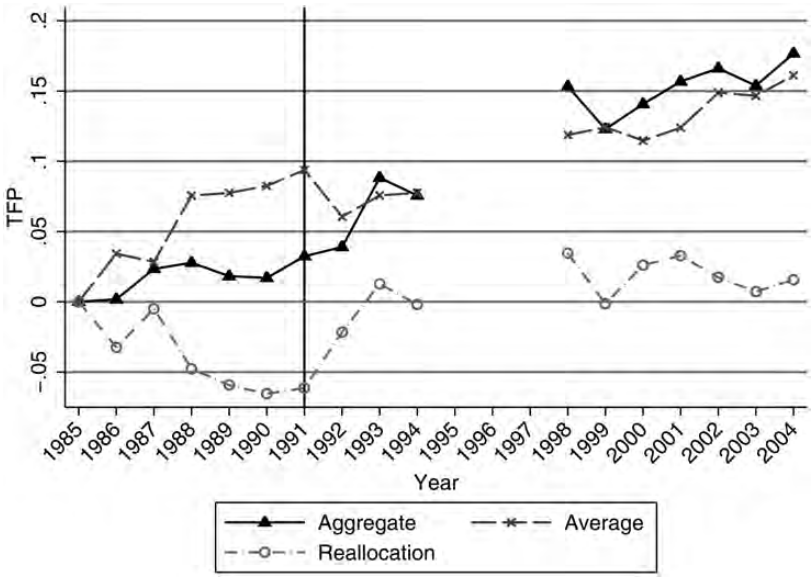
Industry-Level TFP Changes and Policy Reforms

To what extent can the increase in productivity be attributed to trade and other policy reforms that occurred during the 1990s? To answer this question, we exploit variation in policies across industries to examine whether changes in the individual components of productivity are systematically related to specific reforms.

To use policy variation across industries, we recreate our aggregate, average, and reallocation TFP measures at the state-industry level (recall that a state-industry group is the collection of firms in a particular state and three-digit

11. There has been an extensive debate about TFP growth in the organized Indian manufacturing sector, particularly during the 1980s. Goldar (December 7, 2002) provides a summary of a number of TFP growth estimates and discusses many of the measurement issues involved. Our TFP estimates are based on a gross output (rather than value added) production function. Value-added TFP growth rates tend to be much higher than gross output growth rates.

FIGURE 2. All-India Total Factor Productivity



TFP decompositions for the population of firms, conducted at the all-India level, using the Olley and Pakes method. "Aggregate" indicates market-share-weighted mean productivity, "Average" indicates unweighted mean productivity, and "Reallocation" indicates the covariance between market share and productivity. *Source*: Authors' calculations based on ASI data.

industry). We use the state-industry level because this level of disaggregation allows us to consider variations in policies and other characteristics across both industries and states and because the ASI survey is designed to be representative at this level. We weight each group by the total number of firms that appear in that group across all years. This method ensures that the results are comparable to the all-India results because larger state-industry groups are given more weight.[12] We exploit the fact that the trade, licensing, and FDI reforms occurred differentially across industries to isolate the impacts of each policy on each productivity measure. Consider the relationship between our outcomes of interest and the

12. This weighting scheme ensures that average productivity is nearly the same at the state-industry and all-India levels. The reallocation component at the state-industry level follows the same general pattern as that observed in figure 2, but it is lower across most years. The reason is that at this level, we can only measure reallocation *within* state-industry groups. For example, suppose that the steel industry in Maharashtra is more productive than the chemical industry in Gujarat, and all firms in the former state-industry group increase output by 10 percent, whereas all firms in the latter state-industry group reduce output by 10 percent. The all-India reallocation measure will increase, but the state-industry reallocation measure will not. It would be ideal to capture market-share reallocations between and within state-industry groups, but our identification strategy relies on the variation in reforms across industries and over time and thus does not allow us to use an all-India measure of productivity. The supplementary appendix shows that our results are robust to performing the baseline regression analysis without including weights (table S1.12).

reforms:

$$\widehat{Y}_{jst} = \beta_1 \tau_{j,t-1} + \beta_2 \tau^I_{j,t-1} + \beta_3 Delic_{j,t-1} + \beta_4 FDI_{j,t-1} + \alpha_{js} + \alpha_t + \varepsilon_{jst} \qquad (3)$$

where $\widehat{Y}_{jst}$ is the estimated aggregate TFP ($\widehat{\Phi}^{AGG}_{jst}$), average TFP ($\widehat{\Phi}^{U}_{jst}$), or reallocation ($\widehat{R}_{jst}$) for industry j and state s at time t; $\tau_{j,t-1}$ and $\tau^I_{j,t-1}$ are final goods tariffs and input tariffs; $Delic_{j,t-1}$ is a dummy variable equal to one if any products in an industry are delicensed and zero otherwise; $FDI_{j,t-1}$ is a dummy variable equal to one if any products in an industry are FDI liberalized and zero otherwise; and α_{js} and α_t are state-industry and year dummy variables, respectively. Because our firm data are annual and policy changes occurred throughout the year, we lag all policy variables by one year. We employ a fixed-effects estimator to estimate equation 3[13] and cluster all standard errors at the state-industry level.[14] We use a balanced panel of state-industries to avoid confounding within-group effects with the entry and exit of certain industries in particular states, and we weight all observations using the total number of firms in each state-industry group over all years.

Table 2 presents baseline results for 1986 to 2004.[15] Column (1) suggests that a 10-percentage-point reduction in final goods tariffs yields a 0.51 percent increase in aggregate productivity, whereas a 10-percentage-point reduction in input tariffs yields a 5.7 percent increase in aggregate productivity. Columns (2) and (3) present results for the average and reallocation components of productivity, respectively. Column (2) indicates that 10-percentage-point declines in final goods and input tariffs raise average productivity by 0.44 and 5.5 percent, respectively, although the coefficient on final goods tariffs is not statistically significant at the 10 percent level. FDI liberalization increases average productivity by 5 percent. The results are similar in magnitude for the impact of the reforms on both aggregate and average productivity. This similarity largely results because the reforms primarily affected average productivity, which we refer to in this paper as "learning."

In contrast, column (3) shows that input tariffs, final goods tariffs, and delicensing changes do not significantly affect productivity gains through reallocation of market share. The only statistically significant result, for FDI reform, indicates that liberalization lowers rather than raises reallocation productivity. One potential reason for this puzzling result is that prior to the FDI reform, foreign investors only invested in the most productive Indian firms. By introducing an "automatic" approval for majority FDI ownership, the reform decreased the fixed cost of foreign investment and may therefore have encouraged

13. This is similar to including a full set of state-industry interactions.

14. The supplementary appendix (table S1.14) shows that our results are robust to clustering standard errors at the industry-year level.

15. We exclude 1985 because we do not have lagged policy variables for this year.

TABLE 2. Productivity Decompositions and Policy Changes: Baseline Results

	Aggregate (1)	Average (2)	Reallocation (3)
Final Goods Tariff	−.051	−.044	−.007
	(.026)**	(.030)	(.015)
Input Tariff	−.567	−.546	−.021
	(.104)***	(.116)***	(.061)
FDI Reform	.021	.050	−.030
	(.013)	(.014)***	(.010)***
Delicensed	−.006	.005	−.011
	(.017)	(.017)	(.011)
Obs.	17106	17106	17106
R^2	.082	.077	.014

Source: Authors' analysis based on data sources discussed in the text.

Each observation is a state-industry. Dependent variable names are given at the top of each column. "Aggregate" indicates market-share-weighted mean productivity, "Average" indicates unweighted mean productivity, and "Reallocation" indicates the covariance between market share and productivity. All specifications are fixed-effects analyses at the state-industry level and include year dummies. Each observation is weighted by the total number of firms in the state-industry across all years, and standard errors are clustered at the state-industry level.

investment in less productive firms, allowing them to increase their market shares.

In table 3, we estimate the extent to which policy reforms can explain overall productivity growth by multiplying the coefficients from table 2 by the average policy changes. The results suggest that trade liberalization, particularly the decline in input tariffs, is largely responsible for aggregate and average productivity growth. A 60-percentage-point decline in final goods tariffs implies an aggregate productivity increase of 3 percent and an average productivity increase of 2.6 percent (although the related regression coefficient is not statistically significant), and a 40-percentage-point decline in input tariffs implies aggregate and average productivity increases of approximately 22 percent. FDI liberalization also plays a role, implying a 4.7 percent increase in average productivity.[16] The variation in policies across industries cannot explain the gains in reallocation productivity that were observed in the initial years following the reforms. However, the policies explain the gains in average productivity, which was the more important driver of aggregate productivity growth during this period.

Firm-Level Regressions

Our results using the population of enterprises aggregated to the state-industry level suggest that average productivity ("learning") played a more important role in explaining aggregate productivity increases in India during the sample

16. In fact, the average policy changes can explain somewhat more than the total increase in productivity during this time period. In the regression framework, the coefficients on several year dummies are negative, implying that in the absence of policy reforms, productivity would have fallen.

TABLE 3. Productivity Increases Implied by Policy Changes

	Final Goods Tariffs	Input Tariffs	FDI Liberalization	Delicensing
Aggregate (%)	**3.0**	**22.1**	2.1	−0.4
Average (%)	2.6	**21.3**	**4.7**	0.3
Reallocation (%)	0.4	0.8	**−2.8**	−0.6

Source: Authors' analysis based on data sources discussed in the text.

Implied increases in aggregate, average, and reallocation productivity. Results are based on regression coefficients and average policy changes. Bold font indicates that the underlying regression results are statistically significant at the 10 percent level.

period than did reallocation of market share ("stealing"). In this section, we use firm-level data to confirm that learning was a key component of productivity change. If learning is important, we would expect our policy variables to explain productivity within the larger firm-level population and for the smaller constructed panel.

We now use the constructed panel to examine the average productivity results in more detail. We estimate the following equation at the firm level:

$$\widehat{\phi}_{ijst} = \beta_1 \tau_{j,t-1} + \beta_2 \tau^J_{j,t-1} + \beta_3 Delic_{j,t-1} + \beta_4 FDI_{j,t-1} + \alpha_i + \alpha_t + \varepsilon_{ijst}. \quad (4)$$

As discussed in section II, our constructed panel makes it difficult to accurately identify firm exit. However, we are confident in the firms that we are able to match. Hence, the fixed-effects estimator shown in equation 4 should allow us to identify within-firm changes in productivity.

Table 4 presents the results. In column (1), we include all firms that were used in the state-industry level analysis (the estimated population). This specification includes industry and year dummy variables. As we expect, the coefficients on the policy variables are similar to the average productivity results at the state-industry level.

Column (2) presents results for firms that appear in the panel for at least two years, without sampling multipliers. This specification includes industry and year dummy variables but not firm fixed effects, and it shows that the results for firms in the panel are similar to the overall results. The coefficients on input tariffs and FDI reforms remain the largest and most significant, indicating that reductions in input tariffs and the FDI reform made the largest contributions to increased productivity during the sample period.

Column (3) controls for a number of firm characteristics and shows that public firms and young firms (less than three years old) are relatively unproductive, whereas firms furnishing joint returns for multiple factories ("multiplant" firms) are relatively more productive.[17] The poor productivity performance of

17. The multiplant dummy should be cautiously interpreted because true multiplant firms may not be able or may not choose to submit joint returns (see section II).

TABLE 4. Firm-Level Productivity

	Population (1)	Panel Firms (2)	Panel Firms (3)	Firm FE (4)	OP (5)	Large Firms FE (6)
Final Goods Tariff	−.046	−.038	−.037	−.042	−.050	−.054
	(.027)*	(.019)**	(.019)*	(.008)***	(.007)***	(.011)***
Input Tariff	−.486	−.532	−.519	−.141	−.085	−.156
	(.108)***	(.088)***	(.087)***	(.034)***	(.029)***	(.053)***
FDI Reform	.045	.055	.053	.031	.027	.014
	(.015)***	(.012)***	(.011)***	(.004)***	(.004)***	(.005)***
Delicensed	−.005	−.002	−.002	−.002	.005	.027
	(.016)	(.013)	(.013)	(.005)	(.005)	(.007)***
Public			−.167			
			(.014)***			
Multiplant			.039			
			(.011)***			
Young			−.092			
			(.005)***			
Small			−.073			
			(.010)***			
Midsize			−.016			
			(.009)*			
Obs.	1322803	388430	388430	388430	384003	63062
R^2	.065	.057	.064	.002	.004	.046
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	No	No	No
Firm FE	No	No	No	Yes	Yes	Yes

Source: Authors' analysis based on data sources discussed in the text.

Each observation is a firm. The dependent variable is TFP, calculated following Aw et al. (2001) in all columns except (5), in which TFP is calculated following Olley and Pakes (1996). Column (1) includes all firms that were part of the state-industry level analysis; columns (2)–(5) only include firms that appear in the panel for at least two years. Column (6) only includes the largest firms in each year. Columns (1)–(3) include industry and time dummies, and standard errors are clustered at the state-industry level. Columns (4)–(6) include year dummies and firm fixed effects, and standard errors are clustered at the firm level. "FE" indicates fixed effects.

public sector enterprises is consistent with the evidence for other countries, including China (Du, Harrison, and Jefferson 2011) and Indonesia (Bartel and Harrison 2005). We also created dummy variables for firms in three size categories: less than 20 employees (small), 20 to 99 employees (medium), and more than 100 employees (large, the omitted category). The results indicate a positive correlation between size and productivity.

In column (4), we present results with firm fixed effects, thus isolating within-firm changes in productivity. The coefficient on final goods tariffs remains similar. Although the coefficients on input tariffs and FDI reform are attenuated, they remain statistically significant at the 1 percent level. These results confirm that trade and FDI liberalizations are associated with substantial increases in productivity within individual firms. However, the results also indicate that the changing composition of firms and unobserved, firm-level characteristics may play important roles in the observed average productivity gains for the population of firms.

As discussed above, our baseline productivity measure does not depend on accurately identifying exit; the difficulty in identifying exit is a limitation of our panel. Nevertheless, in column (5), we assume that our assignment of exit is accurate, and we use the Olley and Pakes method to estimate productivity. The coefficients are similar to those in column (4), although the coefficient on final goods tariffs is slightly larger in magnitude, whereas the coefficients on input tariffs and FDI are slightly smaller.

To our knowledge, our constructed panel is the largest and most representative of the organized sector to date, consisting of approximately 25,000 firms per year. Our study complements the work of Topalova and Khandelwal (2011), who examine the impacts of India's trade liberalization on approximately 4,000 large firms. Although we cannot replicate their analysis, in column (6), we restrict our analysis to the largest 4,000 firms that appear in the panel in any given year. The results for the largest firms and for the whole panel are similar. The coefficient on delicensing, which was insignificant across all firms, becomes positive and statistically significant for the largest firms. These results suggest that delicensing was particularly important for spurring productivity increases among the largest firms. Additionally, our study extends Sivadasan (2009) in several important ways. First, we create a panel of firms that allows us to confirm that the impacts of the trade and FDI liberalization are important even after controlling for firm fixed effects and to directly show that "learning" is an important mechanism for understanding productivity increases over this 20-year period. Second, we examine long-run impacts on productivity by extending our analysis through 2004. Third, we distinguish between the effects of final goods and input tariffs and demonstrate that input tariffs play a larger role in boosting productivity.

State, Industry, and Firm Characteristics

We now explore whether the effects of the reforms vary across states or industries with different prereform characteristics and among firms of different sizes using the population of firms.

First, we consider whether the impact of liberalization on firm productivity is influenced by exposure to trade. We use three measures to proxy for trade exposure. First, we construct a dummy variable that is equal to one if a state-industry group is located in a state with a port and that is equal to zero otherwise. Second, we calculate each industry's share of imports in output in 1990 using data from the COMTRADE database, and we create a dummy variable for import exposure that is equal to one if the industry has an import share above the median and that is equal to zero otherwise. Third, we construct a similar measure for export share.

The delicensing and FDI reforms have larger effects on productivity among firms that are relatively less exposed to trade (table 5). In states without ports and in nonimporting industries, delicensing is associated with a 4 to 5 percent increase in average productivity. FDI reform is associated with a 7.7 percent increase in average productivity among nonexporting industries. The effect of the FDI reform on average productivity is attenuated in exporting industries, whereas the effect of the delicensing reforms is actually reversed in importing industries and in states with ports. These results suggest some degree of substitutability between external competition and internal competition: where states or industries are not already exposed to trade through proximity to ports, import competition, or exposure to foreign markets, industrial reforms that encourage competition have larger effects.

Next, we consider the role of labor regulations using two state-level measures: (1) the measure developed by Besley and Burgess (2004), which is based on state amendments to the Industrial Disputes Act, and (2) data from the Ministry of Labor on how often firm requests to close down or lay off workers are granted. We interact each measure with our reform variables and show (tables S1.2 and S1.3 in the supplementary appendix) that the effects of the policy reforms are largely similar across states, regardless of labor regulations. However, FDI reform is associated with a 7.4 percent increase in average productivity in states where it is difficult to lay off workers but only a 2.6 percent increase in average productivity in states where it is easy to lay off workers. This difference suggests that in states where it is difficult for firms to achieve an optimal input mix by laying off workers, they may be able to increase their productivity through other means, such as attracting FDI. In other words, FDI reform matters more when existing rigidities make it difficult for firms to optimize their production.

Finally, we consider whether productivity changes may differ across firms of different sizes. Using the population data but harnessing information from the panel, we classify firms into three categories, small (<20 employees), medium

TABLE 5. Productivity Decompositions and Policy Changes: Trade Exposure

	Aggregate (1)	Aggregate (2)	Aggregate (3)	Average (4)	Average (5)	Average (6)	Reallocation (7)	Reallocation (8)	Reallocation (9)
Final Goods Tariff	−.036	−.027	.013	−.065	.018	.033	.029	−.045	−.021
	(.035)	(.019)	(.025)	(.045)	(.032)	(.030)	(.026)	(.026)*	(.023)
Input Tariff	−.600	−.538	−.622	−.502	−.616	−.587	−.098	.078	−.035
	(.113)***	(.109)***	(.102)***	(.123)***	(.111)***	(.115)***	(.069)	(.065)	(.072)
FDI Reform	.026	.028	.044	.053	.054	.077	−.027	−.026	−.033
	(.017)	(.017)*	(.019)**	(.020)***	(.019)***	(.018)***	(.012)**	(.013)**	(.013)**
Delicensed	.007	.042	−.013	.043	.046	.010	−.036	−.004	−.023
	(.019)	(.018)**	(.025)	(.021)**	(.019)**	(.024)	(.015)**	(.014)	(.014)*
Port in State X Final Goods Tariff	−.0002			.0003			−.0005		
	(.0005)			(.0006)			(.0003)		
Port in State X Input Tariff	.0006			−.0005			.001		
	(.0009)			(.001)			(.0006)*		
Port in State X FDI Reform	−.007			−.004			−.004		
	(.017)			(.020)			(.015)		
Port in State X Delicensed	−.020			−.058			.038		
	(.027)			(.027)**			(.019)**		
Importing Industry X Final Goods Tariff		−.0003			−.001			.0008	
		(.0005)			(.0005)**			(.0003)**	
Importing Industry X Input Tariff		−.001			.0007			−.002	

(Continued)

TABLE 5. Continued

	Aggregate (1)	Aggregate (2)	Aggregate (3)	Average (4)	Average (5)	Average (6)	Reallocation (7)	Reallocation (8)	Reallocation (9)
		(.0009)			(.001)			(.0006)***	
Importing Industry X FDI Reform		−.013			−.004			−.009	
		(.016)			(.019)			(.015)	
Importing Industry X Delicensed		−.160			−.125			−.035	
		(.044)***			(.039)***			(.025)	
Exporting Industry X Final Goods Tariff			−.001			−.001			.0002
			(.0004)**			(.0005)**			(.0003)
Exporting Industry X Input Tariff			.001			.001			.00007
			(.0008)			(.001)			(.0007)
Exporting Industry X FDI Reform			−.038			−.048			.009
			(.017)**			(.018)***			(.014)
Exporting Industry X Delicensed			.011			−.013			.024
			(.031)			(.030)			(.020)
Obs.	17106	17106	17106	17106	17106	17106	17106	17106	17106
R^2	.083	.091	.085	.078	.083	.081	.015	.016	.014

Source: Authors' analysis based on data sources discussed in the text.

Each observation is a state-industry. Dependent variable names are given at the top of each column. "Aggregate" indicates market-share-weighted mean productivity, "Average" indicates unweighted mean productivity, and "Reallocation" indicates the covariance between market share and productivity. All specifications are fixed-effects analyses at the state-industry level and include year dummies. Each observation is weighted by the total number of firms in the state-industry across all years, and standard errors are clustered at the state-industry level. "Port in state" is a dummy variable equal to one if the state-industry group is located in a state with a port and zero otherwise. "Importing" ("Exporting") is a dummy variable equal to one if the industry's prereform share of imports (exports) in total output was greater than the median and zero otherwise.

(20–99 employees), and large (>100 employees), on the basis of the size of the firm when we first observed it. In the supplementary appendix (table S1.4), we show that across all types of firms, policy changes continue to drive average productivity, but not reallocation of market share across firms. However, the effects of the reforms vary by firm size. For example, FDI liberalization is most important for large firms; the reform is associated with a 7.5 percent (9.1 percent) increase in aggregate (average) productivity for firms with 100 or more employees, approximately twice the magnitude of the average effect. Although the delicensing reforms are not associated with overall productivity increases, they are associated with a 4.6 percent increase in aggregate productivity among large firms and a 3.9 percent increase in average productivity among mid-sized firms. This heterogeneity is consistent with the fact that only firms with 50 or more employees and a certain amount of assets were required to obtain operating licenses prior to reform.

Robustness Tests

We test the robustness of our baseline results in a number of ways. The results are presented in the supplementary appendix (tables S1.5 through S1.14). First, we examine whether our results are robust to constructing TFP in different ways: (1) winsorizing our baseline measure, (2) using a variation of our baseline measure that employs cost shares instead of revenue shares, and (3) using ordinary least squares. Next, we examine several other modifications of the baseline specification: using an alternative measure of capital; restricting the analysis to the initial years of the reforms, during which policy changes were less likely to be influenced by political considerations; using an alternative measure of input tariffs; removing outlier values in tariff changes; weighting all state-industry groups equally; including state-by-year dummy variables; and clustering standard errors at the industry-year level. The appendix shows that although there is some variation in the magnitude and significance of results, they are robust to each of these tests.

Finally, to test the robustness of our productivity decomposition, we use an alternative method suggested by Melitz and Polanec (2010). We divide the panel of firms in any two consecutive periods, $t-1$ and t, into firms present in both periods (*survivors*), firms that exit after period $t-1$ (*exiters*), and firms that enter in period t (*entrants*). In period $t-1$, only exiters and survivors are present; $S_{X,t-1}$ denotes the market share associated with exiters. In period t, only entrants and survivors are present; $S_{E,t}$ denotes the market share associated with entrants. Melitz and Polanec show that the change in aggregate productivity from period $t-1$ to t can be decomposed as follows:

$$\Phi_t^{AGG} - \Phi_{t-1}^{AGG} = \left[\Phi_{S,t}^{U} - \Phi_{S,t-1}^{U}\right] + \left[R_{S,t} - R_{S,t-1}\right] + S_{E,t}\left[\Phi_{E,t}^{AGG} - \Phi_{S,t}^{AGG}\right] + S_{X,t-1}\left[\Phi_{S,t-1}^{AGG} - \Phi_{X,t-1}^{AGG}\right]. \quad (5)$$

The first and second terms on the right-hand side represent changes in within-firm productivity and the covariance between productivity and market share of firms that survive from $t - 1$ to t. The third term represents the contribution of firms that enter in period t, weighted by the market share of entrants, $S_{E,t}$. Similarly, the last term represents the contribution of firms that exit in period $t - 1$, weighted by the market share of exiters, $S_{x,t-1}$. Using this approach, we calculate the change in TFP between period $t - 1$ and period t and then add the change in TFP to the existing level of TFP in period $t - 1$. TFP is normalized to zero in 1985. In this analysis, we do not use the sampling multipliers.

To use this method, we must assign every firm in our panel to the category of survivor, entrant, or exiter in every year. Given the nature of our panel data, this method requires two relatively strong assumptions. First, to address the fact that we do not directly calculate TFP for 1995–1997, we impute missing values for TFP and output for each series that bridges these years using linear interpolation. We perform a similar linear interpolation of TFP and output for individual firms for which we have bridged over another year.[18]

Second, we must make some assumptions regarding firm exit. When we observe a potential exiter, it is unclear whether the firm actually exited, whether it still existed but was not surveyed in the following year, or whether it was surveyed but we failed to match it.[19] We address this challenge by estimating the "true" rate of exit for each cohort of firms (e.g., firms established between 1974 and 1976) on the basis of the number of surviving firms that we observe. In each year, we consider the potential pool of exiting firms (i.e., firms we do not observe in any subsequent year), and we assign exit status to the number of firms that we estimate to have exited from each cohort.[20] The remaining firms are assigned to the group of survivors.

The baseline results are robust to this alternative decomposition method (table 6).[21] A 10-percentage-point decline in input tariffs is associated with a 4.1 percent increase in aggregate productivity and a 4.8 percent increase in average productivity. FDI liberalization also increases aggregate productivity by

18. For example, if a firm was surveyed in 1992 and 1994 and we are able to link that firm across those years, we use a linear interpolation to estimate TFP and output for that firm in 1993.

19. In the actual panel from 1998 to 2004, the third case is not a concern, although we are still unable to distinguish between the first two cases in many instances.

20. To determine an appropriate method for assigning exit, we examined the distribution of TFP for potential exiters compared to that of survivors for two years (1999 and 2000) in which the observed exit rates are relatively close to estimated exit rates, indicating that the pool of potential exiters is likely to be representative of "true" exiters. We also examined TFP distributions in two years (1995 and 2004) when the observed exit rate is significantly higher than the estimated true exit rate, indicating that many true survivors are classified as exiters. In both cases, the distributions of potential exiters are slightly left-shifted, indicating that exiters are, on average, less productive than survivors. However, the two distributions of potential exiters are similarly left-shifted relative to the distributions of survivors, suggesting that the pool of potential exiters is fairly representative of the actual exiters. Therefore, we assign exit by selecting a random sample of firms from the pool of potential exiters.

21. In this case, the number of state-industry observations is larger because of the imputation of TFP in the panel.

TABLE 6. Robustness Test: Alternative Decomposition

	Aggregate (1)	AvgSurv (2)	ReallocSurv (3)	Entrants (4)	Exiters (5)
Final Goods Tariff	−.022	−.023	−.007	.002	.006
	(.022)	(.026)	(.019)	(.006)	(.008)
Input Tariff	−.408	−.476	.093	−.008	−.016
	(.096)***	(.106)***	(.082)	(.022)	(.033)
FDI Reform	.039	.044	−.011	−.0006	.006
	(.013)***	(.014)***	(.012)	(.002)	(.004)
Delicensed	−.010	.008	−.014	−.001	−.003
	(.015)	(.017)	(.015)	(.004)	(.005)
Obs.	19328	19328	19328	19328	19328
R^2	.059	.068	.035	.07	.003

Source: Authors' analysis based on data sources discussed in the text.

The decomposition is performed using the method suggested by Melitz and Polanec (2010), with false exit addressed as discussed in the text. Each observation is a state-industry. Dependent variable names are given at the top of each column. "Aggregate" indicates market-share-weighted mean productivity, "AvgSurv" and "ReallocSurv" indicate unweighted mean productivity and the covariance between market share and productivity for surviving firms, respectively, and "Entrants" and "Exiters" indicate the contributions of entering and exiting firms. All specifications are fixed-effects analyses at the state-industry level and include year dummies. Each observation is weighted by the total number of firms in the state-industry across all years, and standard errors are clustered at the state-industry level.

3.9 percent and average productivity by 4.4 percent. The policy reforms are not associated with reallocation among survivors, entrants, or exiters.

IV. CONCLUSION

In the Indian case, we show that market share reallocations were important drivers of productivity growth only at the beginning of the trade reforms in 1991. Over the longer 20-year period from 1985 to 2004, average productivity improvements played a larger role in determining aggregate productivity growth.

In contrast to the earlier trade literature on heterogeneous firms, such as Melitz (2003), we do not find a link between India's tariff liberalization and market-share reallocations. Instead, we find that both the trade and FDI reforms increase average firm productivity. Our constructed panel allows us to verify that even after controlling for unobservable, firm-specific characteristics, the trade and FDI reforms are associated with increased within-firm productivity. Although the delicensing reforms do not affect productivity in the organized manufacturing sector as a whole, they are linked to productivity gains among large firms and among firms not previously exposed to trade. One potential reason that we do not find a link between trade reform and reallocations could be that India's rigid labor laws prevent reallocation among firms. However, we

find that the reforms have similar effects in states with different degrees of labor market rigidities.

Our findings, which suggest that "learning" is more important than "stealing" over the 1985 through 2004 period, are consistent with the most recent literature on heterogeneous firms (see, for example, Bernard, Redding, and Schott 2011), which suggests that firms exposed to increased competition from trade may focus on higher productivity product lines. This finding implies that much of the productivity increase associated with trade reform is likely to manifest as within-firm increases rather than productivity gains associated with shifting market shares toward more efficient enterprises. Unfortunately, given the nature of our data during the years in which the major reforms occurred, we are unable to confirm that product shifting occurred, and the existing evidence from a sample of large firms is mixed (Goldberg, Khandelwal, Pavcnik, and Topalova 2010a, 2010b). Exploring the specific channels through which individual firms increase their productivity in India remains an important avenue for future research.

References

Aghion, P., R. Burgess, S. Redding, and F. Zilibotti. 2008. "The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India." *The American Economic Review* 98 (4): 1397–1412.

Amiti, M., and J. Konings. 2007. "Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia." *The American Economic Review* 97 (5): 1611–1638.

Arkolakis, C. 2010. "Market Penetration Costs and the New Consumers Margin in International Trade." *Journal of Political Economy* 118 (6): 1151–1199.

Arkolakis, C., and M.-A. Muendler. 2010. "The Extensive Margin of Exporting Products: A Firm-level Analysis." NBER Working Paper 16641. National Bureau of Economic Research, Cambridge, MA.

Aw, B. Y., X. Chen, and M. J. Roberts. 2001. "Firm-level Evidence on Productivity Differentials and Turnover in Taiwanese Manufacturing." *Journal of Development Economics* 66 (1): 51–86.

Bartel, A. P., and A. E. Harrison. 2005. "Ownership versus Environment: Disentangling the Sources of Public-Sector Inefficiency." *The Review of Economics and Statistics* 87 (1): 135–147.

Bernard, A. B., J. B. Jensen, and P. K. Schott. 2006. "Trade Costs, Firms and Productivity." *Journal of Monetary Economics* 53: 917–937.

Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum. 2003. "Plants and Productivity in International Trade." *The American Economic Review* 93 (4): 1268–1290.

Bernard, A. B., I. Van Beveren, and H. Vandenbussche. 2010. "Multi-Product Exporters, Carry-Along Trade and the Margins of Trade." National Bank of Belgium Working Paper Research Series No. 203. National Bank of Belgium, Brussels, Belgium.

Bernard, A. B., S. J. Redding, and P. K. Schott. 2010. "Multiple-Product Firms and Product Switching." *American Economic Review* 100 (1): 70–97.

———. 2011. "Multiproduct Firms and Trade Liberalization." *The Quarterly Journal of Economics* 126: 1271–1318.

Berthou, A., and L. Fontagne. Forthcoming. "How do Multi-Product Exporters React to a Change in Trade Costs?" *Scandinavian Journal of Economics*.

Besley, T., and R. Burgess. 2004. "Can Labor Regulation Hinder Economic Performance? Evidence from India." *The Quarterly Journal of Economics* 119 (1): 91–134.

Corden, W. M. 1974. *Trade Policy and Economic Welfare*. Oxford: Clarendon Press.

Debroy, B., and A. T. Santhanam. 1993. "Matching Trade Codes with Industrial Codes." *Foreign Trade Bulletin* 24 (1): 5–27.

Du, L., A. Harrison, and G. Jefferson. 2011. "Do Institutions Matter for FDI Spillovers? The Implications of China's 'Special Characteristics'." NBER Working Paper 16767. National Bureau of Economic Research, Cambridge, MA.

Eaton, J., S. Kortum, and F. Kramarz. 2011. "An Anatomy of International Trade: Evidence from French Firms." *Econometrica* 79 (5): 1453–1498.

Eckel, C., and J. P. Neary. 2010. "Multi-Product Firms and Flexible Manufacturing in the Global Economy." *The Review of Economic Studies* 77: 188–217.

Feenstra, R., and H. Ma. 2007. "Optimal Choice of Product Scope for Multiproduct Firms under Monopolistic Competition." NBER Working Paper 13703. National Bureau of Economic Research, Cambridge, MA.

Fernandes, A. M. 2007. "Trade Policy, Trade Volumes and Plant-Level Productivity in Colombian Manufacturing Industries." *Journal of International Economics* 71 (1): 52–71.

Goldar, B. 2002. "TFP Growth in Indian Manufacturing in 1980s." *Economic and Political Weekly* December 7: 4966–4968.

Goldberg, P. K., A. Khandelwal, N. Pavcnik, and P. Topalova. 2010a. "Imported Intermediate Inputs and Domestic Product Growth: Evidence from India." *The Quarterly Journal of Economics* 125 (4): 1727–1767.

———. 2010b. "Multiproduct Firms and Product Turnover in the Developing World: Evidence from India." *The Review of Economics and Statistics* 92 (4): 1042–1049.

Grossman, G., and E. Helpman. 1991. *Innovation and Growth in the Global Economy*. Cambridge, MA: MIT Press.

Hasan, R., D. Mitra, and K. V. Ramaswamy. 2007. "Trade Reforms, Labor Regulations, and Labor Demand Elasticities: Empirical Evidence from India." *The Review of Economics and Statistics* 89 (3): 466–481.

Helpman, E., and P. R. Krugman. 1985. *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*. Cambridge, MA: MIT Press.

Helpman, E., M. J. Melitz, and S. R. Yeaple. 2004. "Export Versus FDI with Heterogeneous Firms." *The American Economic Review* 94 (1): 300–316.

Manova, K., and Z. Zhang. 2012. "Export Prices across Firms and Destinations." *The Quarterly Journal of Economics* 127 (1): 379–436.

Mayer, T., M. J. Melitz, and G. I. P. Ottaviano. 2011. "Market Size, Competition, and the Product Mix of Exporters." NBER Working Paper 16959. National Bureau of Economic Research, Cambridge, MA.

Melitz, M. J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695–1725.

Melitz, M. J., and S. Polanec. 2010. "Dynamic Olley-Pakes Decomposition with Entry and Exit." INNODRIVE Working Paper No. 11. INNODRIVE, Vaasa, Finland.

Menezes-Filho, N. A., and M.-A. Muendler. 2011. "Labor Reallocation in Response to Trade Reform." NBER Working Paper No. 17372. National Bureau of Economic Research, Cambridge, MA.

Nataraj, S. 2011. "The Impact of Trade Liberalization on Productivity: Evidence from India's Formal and Informal Manufacturing Sectors." *Journal of International Economics* 85: 292–301.

Nocke, V., and S. Yeaple. 2008. "Globalization and the Size Distribution of Multiproduct Firms." Centre for Economic Policy Research Discussion Paper DP6948. Centre for Economic Policy Research, London, UK.

Olley, G. S., and A. Pakes. 1996. "The Dynamics of Producticity in the Telecommunications Equipment Industry." *Econometrica* 64 (6): 1263–1297.

Pavcnik, N. 2002. "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants." *Review of Economic Studies* 69 (1): 245–276.

Sivadasan, J. 2009. "Barriers to Competition and Productivity: Evidence from India." *The B.E. Journal of Economic Analysis & Policy, 2009* 9 (1): 42.

Topalova, P., and A. Khandelwal. 2011. "Trade Liberalization and Firm Productivity: The Case of India." *The Review of Economics and Statistics* 93 (3): 995–1009.

Trefler, D. 2004. "The Long and Short of the Canada-U.S. Free Trade Agreement." *The American Economic Review* 94 (4): 870–895.

Tybout, J. R., and M. D. Westbrook. 1995. "Trade Liberalization and the Dimensions of Efficiency Change in Mexican Manufacturing Industries." *Journal of International Economics* 39: 53–78.

Structural Change and Cross-Country Growth Empirics

Markus Eberhardt and Francis Teal

One of the most striking features of economic growth is the process of structural change whereby the share of agriculture in GDP decreases as countries develop. The cross-country growth literature typically estimates an aggregate homogeneous production function or convergence regression model that abstracts from the process of structural change. In this paper, we investigate the extent to which assumptions about aggregation and homogeneity matter for inferences regarding the nature of technology differences across countries. Using a unique World Bank dataset, we estimate production functions for agriculture and manufacturing in a panel of 40 developing and developed countries for the period from 1963 to 1992. We empirically model dimensions of heterogeneity across countries, allowing for different choices of technology within both sectors. We argue that heterogeneity is important within sectors across countries implying that an analysis of aggregate data will not produce useful measures of the nature of the technology or productivity. We show that many of the puzzling elements in aggregate cross-country empirics can be explained by inappropriate aggregation across heterogeneous sectors. JEL codes: O47, O11, C23

The early literature on developing countries distinguished between the processes of economic development and economic growth. Economic development was considered a process of structural transformation by which, in Arthur

Markus Eberhardt (markus.eberhardt@nottingham.ac.uk, corresponding author) is a lecturer in economics at the University of Nottingham and a research associate at the Centre for the Study of African Economies (CSAE), Department of Economics, University of Oxford. Francis Teal (francis.teal@economics.ox.ac.uk) is a university reader in economics at the University of Oxford, deputy director of CSAE, and a research fellow at the Institute for the Study of Labor (IZA), Bonn. This research was supported by the UK Economic and Social Research Council [grant numbers PTA-031-2004-00345 and PTA-026-27-2048 to M.E.] and the Bill and Melinda Gates Foundation [to M.E.]. The authors thank Anindya Banerjee, Alberto Behar, Steve Bond, Josep Carrion-i-Silvestre, Areendam Chanda, Hashem Pesaran, Måns Söderbom, Ron Smith, Dietz Vollrath, three anonymous referees, and seminar/session attendants at Oxford, Manchester, and Birmingham as well as the CSAE Annual Conference 2009, the 13th Applied Economics Meeting, the 16th International Panel Data Conference, and the 7th Annual Meeting of the Irish Society of New Economists for useful comments and suggestions. The usual disclaimers apply. The supplementary appendix to this article is available at http://wber.oxfordjournals.org/.

THE WORLD BANK ECONOMIC REVIEW, VOL. 27, NO. 2, pp. 229–271 doi:10.1093/wber/lhs020
Advance Access Publication October 28, 2012

Lewis' frequently cited phrase, an economy that was "previously saving and investing 4 or 5 percent of its national income or less, converts itself into an economy where voluntary savings is running at about 12 to 15 percent of national income" (Lewis 1954: 155). An acceleration in the investment rate was only one part of this process of structural transformation; of equal importance was the process by which an economy moves from dependence on subsistence agriculture to one in which a modern industrial sector absorbs an increasing proportion of the labor force (e.g., Jorgensen 1961; Ranis and Fei 1961; Robinson 1971). In contrast to these models of "development for backward economies" (Jorgensen 1961: 309), where duality between the modern and traditional sectors was a key feature of the model, was the analysis of economic growth in developed economies.[1] Here, the processes of factor accumulation and technical progress occur in an economy that is already developed, in the sense that it has a modern industrial sector and agriculture has ceased to be a major part of the economy (e.g. Solow 1956; Swan 1956).

Since the early 1990s, the literature on economic development and economic growth has yielded a wide array of models with increasing interaction between theory and empirics (Durlauf and Quah 1999; Easterly 2002; Durlauf, Johnson, and Temple 2005). The applied literature continues to be dominated by an empirical version of the aggregate Solow-Swan model (Temple 2005), with much of the debate focusing on the roles of factor accumulation versus technical progress (Young 1995; Klenow and Rodriguez-Clare 1997a, b; Easterly and Levine 2001; Baier, Dwyer, and Tamura 2006). Although some new theoretical and empirical work has used a dual economy approach (e.g., Vollrath 2009a, b; Lin, 2011; McMillan and Rodrik 2011; Page 2012), this model is largely absent from textbooks on economic growth and has not been the central focus for most empirical analyses (Temple 2005). A primary reason for this focus has been the availability of data. The Penn World Table (PWT) dataset (most recently, Heston, Summers, and Aten 2011) and the Barro-Lee data on human capital (most recently, Barro and Lee 2010) have supplied macro data that facilitate the estimation of the aggregate human capital-augmented Solow-Swan model. However, a team at the World Bank has developed comparable sectoral data for agriculture and manufacturing (Crego, Larson, Butzer, and Mundlak 1998) that allow for a closer matching between a dual economy framework and the data, which we seek to exploit in this paper.

We estimate production functions for the manufacturing and agriculture sectors and contrast the results with those from 'stylized' aggregate production functions where we construct all variables by adding up the sectoral values in each country. In addition, we follow the standard approach in the literature

1. We refer to 'dual economy models' as representing economies with two stylized sectors of production (agriculture and manufacturing). 'Technology' and 'technology parameters' refer to the coefficients on capital and labor in the production function model (elasticities with respect to capital and labor), not Total Factor Productivity (TFP) or its growth rate (technical/technological progress).

using data from the PWT to estimate aggregate functions. Our findings indicate that technological differences across countries and sectors are important and that aggregate specifications are likely to produce misleading inferences regarding total factor productivity (TFP).

The remainder of the paper is organized as follows: section I provides motivations for technology heterogeneity across sectors and countries. In section II, we introduce an empirical specification for our dual economy framework, discuss the data, and briefly review the empirical methods and estimators employed. Section III reports and discusses empirical findings at the sector level. Section IV presents empirical findings from stylized and PWT aggregate data as well as evidence for technology heterogeneity. Summary remarks and conclusions are provided in section V.

I. Technology Heterogeneity

In the following sections, we sketch our theoretical arguments for technology heterogeneity across sectors of production and across countries, building on the dual economy and new growth literatures.

Technology Heterogeneity across Sectors

From a technical point of view, an aggregate production function only offers an appropriate construct in a cross-country empirical framework if the economies under investigation do not display large differences in sectoral structure (Temple 2005), because a single production function framework assumes common production technology across all firms facing the same factor prices. Consider two distinct sectors, assuming marginal labor product equalization and capital homogeneity across sectors, and Cobb-Douglas-type production technology. Then, if technology parameters differ between sectors, aggregated production technology cannot be of the (standard) Cobb-Douglas form (Stoker 1993; Temple and Wößmann 2006). Thus, finding different technology parameters across sectoral production functions is potentially a serious challenge to treating production in the form of an aggregated function.

An alternative motivation for focusing on sector-level rather than aggregate growth across countries is the following: it is common practice in applied work to exclude oil-producing countries from any aggregate growth analysis because "the bulk of recorded GDP for these countries represents the extraction of existing resources, not value added" (Mankiw, Romer, and Weil 1992: 413). The underlying argument is that sectoral 'distortions,' such as resource wealth, justify the exclusion of these observations. Therefore, it could be argued that given the large share of agriculture in GDP for countries such as Malawi (25 to 50 percent over the period between 1970 and 2000), India (25 to 46 percent), or Malaysia (8 to 30 percent), these countries should be excluded from any aggregate growth analysis because a significant share of their aggregate GDP is

derived from a single resource, namely, land.[2] A sector-level analysis mitigates this problem because manufacturing and agriculture are clearly more homogeneous sectors than any aggregate construct.

Technology Heterogeneity across Countries

A theoretical justification for heterogeneous technology parameters across countries can be found in the 'new growth' literature. This strand of the literature on theories of economic growth argues that production functions differ across countries and seeks to determine the sources of this heterogeneity (Durlauf, Kourtellos, and Minkin 2001). As Brock and Durlauf (2001: 8/9) remark, "the assumption of parameter homogeneity seems particularly inappropriate when one is studying complex heterogeneous objects such as countries." Azariadis and Drazen's (1990) model can be considered the 'grandfather' for many of the theoretical attempts to allow countries to possess technologies that differ from one another or over time. Other theoretical studies lead to interpretations of multiple equilibria as factor parameter heterogeneity in the production function (e.g., Murphy, Shleifer, and Vishny 1989; Durlauf 1993; Banerjee and Newman 1993). The 'appropriate technology' literature provides a further challenge to the assumption of a common technology, arguing that different technologies are appropriate for different factor endowments (see Basu and Weil 1998): global R&D leaders develop productivity-enhancing technologies that are suitable for their own capital-labor ratios and that cannot be used effectively by poorer countries; therefore, the latter do not develop. Empirical evidence that lends support to this hypothesis can be found in Clark (2007) and Jerzmanowski (2007). A simpler justification for heterogeneous production functions is offered by Durlauf, Kourtellos, and Minkin (2001), who suggest that the Solow model was never intended to be valid in a homogeneous specification for all countries but that it might be a good way to investigate each country, that is, if we allow for parameter differences across countries.

Formal insights for empirical modeling can be gained from the micro production framework introduced in Mundlak (1988) and applied to macro data for agriculture in Mundlak, Larson, and Butzer (1999) and Mundlak, Butzer, and Larson (2012). In these studies, the technology of production available to individual firms is a collection of possible techniques, each with its own production function, with optimal output over implemented techniques defined as

$$Y^* \equiv F(X^*, s) = \varphi(s) \tag{1}$$

2. The quoted shares are from the WDI database (World Bank 2008). For comparison, the maximum share of oil revenue in GDP, computed as the difference between 'industry share in GDP' and 'manufacturing share in GDP' from the same database, yields the following ranges for some of the countries mentioned by Mankiw, Romer, and Weil (1992): Iran (12 to 51 percent), Kuwait (15 to 81 percent), Gabon (28 to 60 percent), and Saudi Arabia (29 to 67 percent).

where X^* and Y^* represent (optimal) inputs and output aggregated over implemented techniques, and s is a vector of state variables determining both optimal input choice $X^*$and implemented technique $F(\cdot)$.[3] In each period,[4] firms face the economic problem of choosing inputs and the appropriate production technique. This joint determination of inputs and technique makes it difficult to identify parameter coefficients in an empirical equivalent of equation (1) unless additional structure is imposed on the problem. Adopting a number of simplifying assumptions, Mundlak, Butzer, and Larson (2012) provide the following approximation for their empirical model of output and inputs (i.e., production/supply and factor demand functions), explicitly including the exogenous state variables s

$$y_{it} = x_{it}\beta(s) + s_{it}\gamma + m_{0it} + u_{0it} \tag{2}$$

$$x_{jit} = s_{it}\gamma + m_{0it} + \varepsilon_{jit} \tag{3}$$

where subscript j refers to the specific observed input to production x, and y is observed output;[5] m_{0it} represents a firm-specific productivity shock at time t that is observed by the firm, thus influencing its input choice, but is unknown to the econometrician. A large body of microeconometric literature (for a recent survey, see Eberhardt and Helmers 2010) has attempted to address the resulting 'transmission bias' first highlighted by Marschak and Andrews (1944). Mundlak, Butzer, and Larson (2012) simplify this productivity shock by requiring that it be decomposable into firm- and time-specific effects, $m_{0it} = m_{0i} + m_{0t}$ (similarly for the input equations). This setup further highlights two 'technology shifters': first, the state variables affect output directly and indirectly through the selection of inputs, acting as input/output shifters; second, the state variables directly influence the technology parameters β. The state variables act as technology shifters in the sense that, conditional on s, (i) different countries might have different β coefficients, and/or (ii) at different points in time, the same country might have different β coefficients. The presence of the state variables in the equations for y and x prevents the straightforward application of instrumental variables.[6]

Following some simplifying assumptions regarding aggregation (see Mundlak, 1988), the above framework is extended to apply at the country level.

3. Crucially, all changes in X^* are instigated by the state variables, and with the exception of error, it is deemed 'meaningless' to think of any other factors driving inputs (Mundlak, Larson, and Butzer 1999).

4. For simplicity, the exposition in Mundlak, Butzer, and Larson (2012) is limited to a static model.

5. u_{0it} and ε_{jit} are white noise.

6. Mundlak, Butzer, and Larson (2012) refer to the presence of state variables in both equations as technology 'heterogeneity.' Our use of the term differs from theirs because we refer to $\beta_i \neq \beta$ as technology heterogeneity.

Empirical testing in the case of the cross-country production function for agriculture is conducted with the following set of state variables: proxies for human capital, level of development, institutions, peak agricultural yield, and a number of indicators for prices and price variability.[7] Using the simplifying assumption $\beta(s) = \beta$, where β is referred to as a 'sample-dependent constant,' the model is estimated using ordinary least squares (OLS) following a within-country-time transformation of the variables (i.e., applying the two-way fixed effects estimator). The authors refer to the results from this regression as 'core technology.'[8] Further empirical analysis in this paper and in a related study by one of the co-authors (Butzer, 2011) investigates parameter constancy over time and parameter heterogeneity across countries by splitting the data into two periods and two country groups. Our own empirical approach discussed below builds on the theoretical model by Mundlak (1988) but allows for more flexibility in the empirical implementation than Mundlak, Butzer, and Larson (2012).

II. An Empirical Model of a Dual Economy

In the following section, we first present a general, empirical specification for our sector-specific analysis of agriculture and manufacturing that shows how recent developments in the econometric modeling of production functions link to the framework proposed by Mundlak. Next, we review a number of empirical estimators, focusing on those arising from the recent panel time series literature, before we briefly discuss the data.

Empirical Specification

Our empirical framework adopts a 'common factor' representation for a standard log-linearized Cobb-Douglas production function model. Each sector/level of aggregation is modeled separately. For ease of notation, we do not identify this multiplicity in our general model. Let

$$y_{it} = \boldsymbol{\beta}_i' \boldsymbol{x}_{it} + u_{it} \qquad u_{it} = \alpha_i + \boldsymbol{\lambda}_i' \boldsymbol{f}_t + \varepsilon_{it} \tag{4}$$

$$x_{mit} = \pi_{mi} + \delta_{mi}' g_{mt} + \varphi_{1mi} f_{1mt} + \ldots + \varphi_{nmi} f_{nmt} + \upsilon_{mit} \tag{5}$$

$$\text{where} \qquad \boldsymbol{f}_t = \boldsymbol{\tau} + \boldsymbol{\rho}' \boldsymbol{f}_{t-1} + \boldsymbol{\omega}_t \quad \text{and} \quad \boldsymbol{g}_t = \boldsymbol{\mu} + \boldsymbol{\kappa}' \boldsymbol{g}_{t-1} + \boldsymbol{\upsilon}_t \tag{6}$$

7. The between-country regressions further include time-invariant proxies for countries' physical environment.

8. Between-time and between-country estimates are also provided, but the 2FE results are the focus of attention.

for $i = 1, \ldots, N$ countries, $t = 1, \ldots, T$ time periods, and $m = 1, \ldots, k$ inputs.[9] Equation (4) represents the production function, with y as sectoral or aggregated value-added and $\boldsymbol{x}$ as a set of inputs: labor, physical capital stock, and a measure for natural capital stock (arable and permanent crop land) in the agriculture specification (all variables are transformed to log values). We consider additional inputs (human capital, livestock, and fertilizer) as robustness checks for our general findings (see supplemental appendix S4, available at http:/wber.oxfordjournals.org/). The output elasticities associated with each input (β_i) are allowed to differ across countries.[10]

For unobserved TFP, we employ the combination of a country-specific TFP level (α_i) and a set of common factors ($\boldsymbol{f}_t$) with country-specific factor loadings ($\boldsymbol{\lambda}_i$). TFP is therefore in the spirit of a 'measure of our ignorance' (Abramowitz 1956), driven by latent processes that are either difficult to measure or that are truly unobservable. Equation (6) provides some structure for these unobserved common processes that are modeled as simple AR(1) processes with drift terms. We do not exclude the possibility of unit root processes ($\boldsymbol{\rho} = 1$, $\boldsymbol{\kappa} = 1$) leading to nonstationary observables and unobservables. Note that the potential for spurious regression results arises in this setup if the empirical equation is misspecified.

Equation (5) details the evolution of the set of inputs, that is, the input demand functions. Crucially, some of the same processes determining the evolution of inputs are assumed to drive TFP in the production function equation.[11] Economically, this assumption implies that the processes that make up TFP (e.g., knowledge, innovation, absorptive capacity) affect choices of inputs, including the accumulation of capital stock, the evolution of the labor force, and (in the agriculture equation) the area of land under cultivation, while at the same time affecting the production of output directly. Thus, technical progress affects both production and the choice of productive inputs. Econometrically, this setup leads to endogeneity whereby the regressors are correlated with the unobservables, making it difficult to identify $\boldsymbol{\beta}_i$ separately from $\boldsymbol{\lambda}_i$ and $\boldsymbol{\varphi}_i$ (Kapetanios, Pesaran, and Yamagata 2011). The nature of macroeconomic variables in a globalized world, where economies are strongly connected to each other and latent forces drive all of the outcomes, provides a conceptual justification for the pervasive character of unobserved common factors. The presence of these latent factors makes it difficult to argue for the validity of traditional approaches to causal interpretation of cross-country empirical analyses. Instrumental variable estimation in cross-section growth regressions or Arellano and Bond-type (1991) lag-instrumentation within pooled panel models become invalid in the face of common factors and/or heterogeneous equilibrium relationships (Pesaran and Smith 1995; Lee, Pesaran, and Smith 1997).

9. Further, $f_{\cdot mt}$ is a subset of f_t, and the error terms $\boldsymbol{\varepsilon}_{it}$, $\boldsymbol{v}_{mit}$, $\boldsymbol{\omega}_t$ and $\boldsymbol{v}_t$ are white noise.
10. Heterogeneity over time will be addressed in section IV.
11. Others, namely, g_t, are specific to the input evolution.

TABLE 1. Estimators and Assumptions about the Data Generating Process

		Impact of Unobservables:	
		Common	Idiosyncratic
Production Technology:	Common	POLS, 2FE, GMM*, PMG*	CCEP, CPMG*
	Idiosyncratic	MG, FDMG	CMG

This framework can be viewed as an empirical version of theoretical Mundlak model developed above. Equations (4) and (5) capture the jointness property that is made explicit in their empirical model by the inclusion of a set of 'state variables,' which affect inputs and output in an identical fashion: γ in equations (2) and (3). Conversely, our framework allows underlying unobserved factors to affect inputs and output differentially via the country-specific factor loadings $\boldsymbol{\lambda}_i$ and $\boldsymbol{\varphi}_i$.[12] These factors are conceptually similar to the state variables in the Mundlak model; they represent any variable or process that might affect both factor choice and TFP. The empirical implementation of our model, however, differs from that of Mundlak. We allow the data to identify the different choices for the β coefficients. The evolution of the factors is fairly general, including nonstationarity, and the setup provides for global shocks (strong factors) as well as local spillovers (weak factors). The productivity shock term m_{0it} is accounted for by a fixed effect α_i (m_{0i}) and the common factor structure ($m_{0t} = \boldsymbol{\lambda}'\boldsymbol{f}_t$).[13] Finally, we allow for technology heterogeneity β_i across countries and analyze whether parameter constancy holds over time ($\boldsymbol{\beta}_{it} = \boldsymbol{\beta}_i$). The parameter constancy tests will provide further insights into the 'core technology' by highlighting whether technology parameters are likely to be functions of unobservable processes (in our case, $\boldsymbol{f}_t$, in the Mundlak, Butzer, and Larson 2012 notation, s). Our empirical implementation is focused on recent panel time series estimators that address nonstationarity, parameter heterogeneity, and cross-section dependence. The following section introduces these methods in more detail.

Empirical Implementation

Our empirical setup incorporates a large degree of flexibility concerning the impact of observable and unobservable inputs on output. Empirical implementation will necessarily lead to different degrees of restrictions on this flexibility, which will then be formally tested: the emphasis is on a comparison of different empirical estimators allowing for or restricting the heterogeneity in the observables and unobservables outlined above. The two-by-two matrix in table 1

12. A detailed review of the important contribution of factor models to empirical macroeconometrics is beyond the scope of this study. See Stock and Watson (2002), Bai and Ng (2008), and Onatski (2009) for details.

13. The shock can never be truly idiosyncratic; m_{0it} differs for each country i at each point in time t. We consider this assumption reasonable given the interconnectedness of economies.

indicates the assumptions that are implicit in the various estimators implemented below.[14] For the estimators marked with stars, we confine the results to the supplemental appendix to save space.[15]

The panel time series econometric approach is given particular attention in this study for a number of reasons (for a detailed discussion, see Eberhardt and Teal 2011). First, we know that many macro variables are potentially nonstationary (Nelson and Plosser 1982; Granger 1997; Pedroni 2007), a property that cannot be rejected for the variables in our data (see supplemental appendix S1). When variables are nonstationary, standard regression output must be treated with extreme caution because results are potentially spurious. However, we can establish long-run equilibrium relationships in the data, provided variables (and unobserved processes) are cointegrated. The practical indication of cointegration is when regressions yield stationary residuals, whereas nonstationary residuals indicate a potentially spurious regression. Panel time series estimators can address this concern over spurious regression, and below, we investigate the residuals of each empirical model using panel unit root tests. Second, panel time series methods allow for parameter heterogeneity across countries, which, as discussed above, is a central interest in our analysis. Third, panel time series methods can address the problems arising from cross-section correlation. Whether this is the result of common economic shocks or local spillover effects, cross-section correlation can potentially induce serious bias in the estimates because the impact assigned to an observed covariate in reality confounds its impact with that of the unobserved processes. Although the panel time series approach does not allow us to quantify their impact, common shocks and local spillovers can be accommodated in the empirical analysis to obtain unbiased technology coefficients for the observable inputs. Below, we will employ diagnostic tests to analyze each model's residuals for the presence or absence of cross-section dependence.

We introduce the Common Correlated Effects (CCE) estimators developed in Pesaran (2006) and extended to nonstationary variables in Kapetanios, Pesaran, and Yamagata (2011) in some more detail because relatively few applied studies employ these estimators (e.g., Holly, Pesaran, and Yamagata,

14. Abbreviations: POLS, Pooled OLS; 2FE, 2-way Fixed Effects; GMM, Arellano and Bond (1991) Difference GMM and Blundell and Bond (1998) System GMM; MG, Pesaran and Smith (1995) Mean Group estimator (with linear country trends); FDMG, ditto with variables in first difference and country drifts; PMG, Pesaran, Shin, and Smith (1999) Pooled Mean Group estimator; CPMG, ditto augmented with cross-section averages following Binder and Offermanns (2007); CCEP/CMG, Pesaran (2006) Common Correlated Effects estimators. Note that our POLS model is augmented with T-1 year dummies.

15. GMM, PMG, and CPMG estimation was based on an error correction model specification; see Pesaran, Shin, and Smith (1999) for details. Further discussion of the empirical setup and results is available on request.

2010; Moscone and Tosetti, 2010; Cavalcanti, Mohaddes, and Mehdi, 2011; Eberhardt, Helmers, and Strauss, forthcoming).[16]

The CCE estimators augment the regression equation with cross-section averages of the dependent ($\bar{y}_t$) and independent variables ($\bar{\boldsymbol{x}}_t$) to account for the presence of unobserved common factors with heterogeneous impact. For the Mean Group version (CMG), the individual country regression is specified as

$$y_{it} = a_i + \boldsymbol{b}_i'\boldsymbol{x}_{it} + c_{0i}\bar{y}_t + \sum_{m=1}^{k} c_{mi}\bar{\boldsymbol{x}}_{mt} + e_{it} \tag{7}$$

In a second step, the parameter estimates $\hat{b}_i$ are averaged across countries similar to the practice in the Pesaran and Smith (1995) Mean Group (MG) estimator.[17] The pooled version (CCEP) is specified as

$$y_{it} = a_i + \boldsymbol{b}'\boldsymbol{x}_{it} + \sum_{j=1}^{N} c_{0i}\left(\bar{y}_t D_j\right) + \sum_{m=1}^{k}\sum_{j=1}^{N} c_{mi}\left(\bar{\boldsymbol{x}}_{mt} D_j\right) + e_{it} \tag{8}$$

where D_j represents country dummies.[18] The CMG is thus a simple extension to the Pesaran and Smith (1995) MG estimator based on country-specific OLS regressions, whereas the CCEP is a standard fixed effects estimator augmented with additional regression terms.

To obtain insight into the mechanics of this approach, consider the cross-section average of our model in equation (4). As the cross-section dimension N increases, given $\bar{\varepsilon}_t = 0$, we obtain

$$\bar{y}_t = \bar{\alpha} + \bar{\boldsymbol{\beta}}'\bar{\boldsymbol{x}}_t + \bar{\boldsymbol{\lambda}}'\boldsymbol{f}_t \quad \Leftrightarrow \quad \boldsymbol{f}_t = \bar{\boldsymbol{\lambda}}^{-1}\left(\bar{y}_t - \bar{\alpha} - \bar{\boldsymbol{\beta}}'\bar{\boldsymbol{x}}_t\right) \tag{9}$$

This simple derivation provides a powerful insight: working with the cross-sectional means of y and $\boldsymbol{x}$ can account for the impact of unobserved common factors (TFP) in the production process.[19] Given the assumed heterogeneity in

16. We abstain from discussing the standard panel estimators here in great detail and refer to the articles by Coakley, Fuertes, and Smith (2006), Bond and Eberhardt (2009), and Bond (2002) for more information. We also investigate the Pooled Mean Group (PMG) estimator by Pesaran, Shin, and Smith (1999) as well as a simple extension to the PMG in which we include cross-section averages of the dependent and independent variables (CPMG), as suggested in Binder and Offermanns (2007).

17. Although $\bar{y}_t$ and e_{it} are not independent, their correlation goes to zero as N becomes larger.

18. Thus, in the MG version, we have N individual country regressions with $2k + 2$ RHS variables, and in the pooled version, there is a single regression equation with $k + N\,(k + 2)$ RHS variables.

19. Most conservatively, the CCE estimators require $\bar{\boldsymbol{\lambda}} \neq 0$: the impact of each factor is, on average, non-zero (Coakley, Fuertes, and Smith 2006). Alternative scenarios (see Pesaran 2006; Kapetanios, Pesaran, and Yamagata 2011) allow for this assumption to be dropped in certain situations, but for the sake of generality, we maintain it here.

the impact of unobserved factors across countries ($\boldsymbol{\lambda}_i$), the estimator is implemented in the manner detailed above, which allows for each country i to have different parameter estimates for $\bar{y}_t$ and the $\bar{\boldsymbol{x}}_t$ and, thus, implicitly for $\boldsymbol{f}_t$. Simulation studies (Pesaran 2006, Coakley, Fuertes, and Smith 2006; Kapetanios, Pesaran, and Yamagata 2011; Pesaran and Tosetti 2011) have shown that this approach works well even when the cross-section dimension N is small, when variables are nonstationary, cointegrated, or not integrated, in the presence of local spillovers and global/local business cycles and when the relationship is subject to structural breaks.[20] In the present study, we implement two versions of the CCE estimators in the sector-level regressions: estimators in a standard form as described above and estimators in a variant form that includes the cross-section averages of the input and output variables from both sectors. This variant specification allows for cross-section dependence across sectors, albeit at the cost of a reduction in degrees of freedom. It is conceivable that the evolution of the agricultural sector in developing countries influences that of the wider economy in general and the manufacturing sector in particular, such that this extension is sensible in the dual economy context.

This completes our discussion of the empirical implementation within each sector/level of aggregation. We highlight the direct link between the issues that these estimators seek to address and the problem of identifying the technology parameters of interest raised in the previous section. Heterogeneity in the impact of observables and unobservables across countries can be directly interpreted as differences in the production technology and a differential TFP evolution across countries. The above discussion suggests that, from an economic theory standpoint, there are reasons to prefer a more flexible empirical approach. Empirically, however, we do not impose this more flexible approach on our data. We compare models with differing degrees of parameter heterogeneity and use established econometric diagnostics (tests for residual stationarity and cross-section independence) to identify the models that are rejected and those that are supported by the data.

Data

Descriptive statistics and a more detailed discussion of the data can be found in the appendix. We conduct our empirical analysis with four datasets:

20. An alternative approach to empirically implementing equation (4) is to estimate factors, factor loadings, and slope coefficients jointly, as in the estimators developed in Bai and Kao (2006) and Bai, Kao, and Ng (2009). Computational complexity aside, two recent theoretical contributions support the Pesaran (2006) approach adopted in this study. Theoretical work by Westerlund and Urbain (2011: 17f) compares the two approaches and concludes that "one is unlikely to do better than when using the relatively simple CA [cross-sectional average augmentation] approach." Similarly, a study by Bailey, Kapetanios, and Pesaran (2012: 25) concludes that the methods used to determine the number of strong factors on which the approach by Bai and co-authors relies are "invalid and will select the wrong number of factors, even asymptotically."

(i) for the agricultural sector, building on the sectoral investment series collected by Crego, Larson, Butzer, and Mundlak (1998) and output from the WDI (World Bank, 2008) as well as sectoral labor and land data from FAO (2007);
(ii) for the manufacturing sector, building on the sectoral investment series collected by Crego, Larson, Butzer, and Mundlak (1998), output data from the WDI, and labor data from UNIDO (2004);
(iii) for a stylized aggregate economy made up of the aggregated data for the agriculture and manufacturing sectors;[21]
(iv) for the aggregate economy, building on data provided by the PWT (we use version 6.2, Heston, Summers, and Aten 2006).

The capital stocks in the agriculture, manufacturing, and PWT samples are constructed from investment series following the perpetual inventory method (see Klenow and Rodriguez-Clare, 1997b). For the aggregated sample, we simply added up the sectoral capital stocks. A comparison across sectors and with the stylized aggregate sector is possible because of the efforts by Crego, Larson, Butzer, and Mundlak (1998) in providing sectoral investment data for agriculture and manufacturing. All monetary values in the sectoral and stylized aggregated datasets are transformed into US dollar values for the year 1990 (in the capital stock case, this transformation is applied to the investment data), following Martin and Mitra (2002). In light of concerns that the stylized aggregate economy data might not offer a sound representation of true aggregate economy data, we have adopted the PWT data, which measure monetary values in international dollars (purchasing power parity adjusted), as a benchmark for comparison. Despite a number of vocal critics (e.g., Johnson, Larson, Papageorgiou, and Subramanian, 2009), the PWT data are undoubtedly the most popular macro dataset for cross-country empirical analysis.[22]

Our sample is an unbalanced panel[23] for 1963 to 1992, consisting of 40 developing and developed countries with a total of 918 observations (average $T = 23$). Our aim is to compare estimates across the four datasets, which requires us to match the same sample, thus reducing the number of

21. We sum the values for value-added, capital stock, and labor, then transform the former two into per capita values, and finally take logarithms.

22. We are, of course, aware that the difference in deflation between our sectoral and stylized aggregated data, on the one hand, and PWT, on the other hand, makes them conceptually very different measures of growth and development. The aggregated data emphasize tradable goods production, whereas the PWT data equally emphasize tradable and non-tradable goods and services. However, we believe that these differences are comparatively unimportant for the purposes of estimation and inference in comparison to the distortions introduced by neglecting the sectoral makeup and technology heterogeneity of economies at different stages of economic development.

23. We do not account for missing observations in any way. The preferred empirical specifications presented below are based on heterogeneous parameter models, in which (arguably) the lack of balance (25 percent of observations in the balanced panel are missing) is less relevant than in the homogeneous models because of the explicit averaging of estimates.

observations to the smallest common denominator. Only eight countries in our sample are in Africa, whereas approximately half are present-day 'industrialized economies.' However, these numbers are deceiving if one recalls that structural change and development in many of these industrialized economies has primarily been achieved during our period of study. For example, prior to 1964, GDP per capita was higher in Ghana than in South Korea. In 1970, the share of agricultural value-added in GDP for Finland, Ireland, Portugal, and South Korea amounted to 13 percent, 16 percent, 31 percent, and 26 percent, respectively, whereas the 1992 figures were 5 percent, 8 percent, 7 percent, and 8 percent. This is strong evidence of economies undergoing structural change. A detailed description of our sample is available in table A1 and descriptive statistics for each sample are provided in table A2.

III. Empirical Results

Panel unit root and cross-section dependence tests for our data are available in the supplemental appendix (S1, S2) of the paper. We adopt the Pesaran (2007) CIPS panel unit root test to analyze the time series properties of each variable series. The results provide strong indication that variables in log levels for the agriculture and manufacturing data as well as the two aggregate economy representations are nonstationary.

A number of formal and informal tests were conducted to investigate cross-section correlation in the data. The results (see supplemental appendix S2) show very high average absolute correlation coefficients for the data in log levels and in the data represented as growth rates. Formal tests for cross-section dependence (Pesaran, 2004; Moscone and Tosetti, 2009) reject cross-section independence in virtually all variable series tested.

Below, we discuss the empirical results from sectoral production function regressions for agriculture and manufacturing, first assuming technology parameter homogeneity and then allowing for differential technology across countries. For all regression models, we report residual diagnostic tests, including the Pesaran (2007) panel unit root test (we summarize results using I(0) for stationary residuals, I(1) for nonstationary residuals, and I(1)/I(0) for ambiguous results), and the Pesaran (2004) cross-section dependence (CD) test (H_0: cross-section independence), which we use to build our judgment for a preferred empirical model. Residual nonstationarity invalidates the inferential tools (for example, *t*-statistics) employed (Kao, 1999) and indicates that regression results are potentially spurious. In the same way that serial dependence indicates dynamic misspecification, residual cross-section dependence violates the assumption that the error terms are independent and identically distributed (iid). This suggests that the specific model tested fails to adequately address the correlation of inputs, output, and unobservables

across different countries, induced by, for example, common shocks or local spillover effects.[24]

Note that our empirical regressions express all variables in per-worker terms (in logs). The inclusion of the log labor variable therefore indicates the deviation from constant returns to scale (i.e., $\hat{\beta}_L + \hat{\beta}_K[+\hat{\beta}_N] - 1$): a positive (negative) significant coefficient on log labor indicates increasing (decreasing) returns; an insignificant coefficient indicates constant returns. The coefficient on labor in the regression is thus not the output elasticity with respect to labor, which we also report in a lower panel of each table ('Implied $\hat{\beta}_L$')[25] along with the returns to scale ('Implied RS'). This setup allows for an easy imposition of constant returns (CRS) by dropping the log labor variable from the model. In each table, Panel (A) shows results with no restrictions on returns to scale, whereas Panel (B) imposes CRS.

Pooled Models

Table 2 presents the empirical results for agriculture and manufacturing. Beginning with agriculture, the empirical estimates for models [1] and [2] neglecting cross-section dependence are quite similar, with the capital coefficient of about .63 and statistically significant decreasing returns to scale. The land coefficient is insignificant in all pooled specifications, except in the 2FE model, where it carries a negative sign. Diagnostic tests indicate that the residuals in these models are cross-sectionally dependent and that the standard POLS and 2FE models yield nonstationary residuals and, thus, might represent spurious regressions. The two CCEP models yield stationary and cross-sectionally independent residuals, capital coefficients of approximately .5 and insignificant land coefficients. There is no substantial change in these results when CRS (Panel (B)) is imposed, with the exception of the 2FE estimates, where the land variable (previously negative and significant) is now insignificant and the capital coefficient has become further inflated. Land is still insignificant, but in models [3] and [4] it now has a plausible coefficient estimate.

In the manufacturing data, the models ignoring cross-section dependence in [5] and [6] yield increasing returns to scale and capital coefficients in excess of .85. Residuals again display nonstationarity; however, the CD tests now imply that they are cross-sectionally independent. Surprisingly, the standard CCEP

24. If the correlation is caused by the same factors as those present in the inputs, the situation is altogether more serious than mere lack of efficiency, namely, that $\boldsymbol{\beta}$ might be unidentified. Residual diagnostics and their importance for empirical modeling are discussed in more detail in Eberhardt and Teal (2011) and Banerjee, Eberhardt, and Reade (2010).

25. This computation is based on statistically significant parameters only: $\hat{\beta}_L = 1 - [\hat{\beta}_K(+\hat{\beta}_N)] + \hat{\beta}_{RS}$, where $\hat{\beta}_{RS}$ is the log labor coefficient discussed above. If any of $\hat{\beta}_K, \hat{\beta}_N$ or $\hat{\beta}_{RS}$ is insignificant, it is omitted from this calculation; if all parameters are insignificant, we report 'not applicable' (n/a).

TABLE 2. Pooled Regression Models for Agriculture and Manufacturing

	PANEL (A) UNRESTRICTED RETURNS TO SCALE							
	Agriculture				*Manufacturing*			
	[1] POLS	[2] 2FE	[3] CCEP	[4] CCEP[b]	[5] POLS	[6] 2FE	[7] CCEP	[8] CCEP[b]
Estimates								
log labor	−0.060	−0.199	−0.266	−0.142	0.043	0.081	0.082	0.002
$\hat{\beta}_L + \hat{\beta}_K (+ \hat{\beta}_N) - 1$	[7.20]**	[9.60]**	[2.13]*	[0.55]	[3.53]**	[4.35]**	[1.53]	[0.03]
log capital pw	0.618	0.661	0.480	0.531	0.897	0.845	0.472	0.469
$\hat{\beta}_K$	[73.80]**	[43.62]**	[9.87]**	[5.92]**	[55.38]**	[32.69]**	[7.62]**	[5.34]**
log land pw	0.011	−0.160	−0.165	0.052				
$\hat{\beta}_N$	[1.02]	[4.93]**	[0.98]	[0.20]				
Diagnostics								
implied RS †	DRS	DRS	DRS	CRS	IRS	IRS	CRS	CRS
implied $\hat{\beta}_L$ ‡	0.322	0.300	0.254	0.469	0.147	0.236	0.528	0.532
ê integrated ◇	I(1)	I(1)	I(0)	I(0)	I(1)	I(1)	I(0)	I(0)
CD test *p*-value #	0.00	0.00	0.45	0.38	0.19	0.34	0.00	0.93
R-squared	0.94	0.86	1.00	1.00	0.84	0.67	1.00	1.00
RMSE	0.446	0.127	0.095	0.086	0.439	0.128	0.090	0.066

(Continued)

TABLE 2. Continued

	PANEL (B) CONSTANT RETURNS TO SCALE IMPOSED							
	Agriculture				*Manufacturing*			
	[1] POLS	[2] 2FE	[3] CCEP	[4] CCEP[b]	[5] POLS	[6] 2FE	[7] CCEP	[8] CCEP[b]
Estimates								
log capital pw	0.644	0.725	0.496	0.526	0.919	0.860	0.490	0.500
$\hat{\beta}_K$	[85.46]**	[48.87]**	[11.22]**	[6.70]**	[70.80]**	[34.01]**	[13.55]**	[8.38]**
Diagnostics								
log land pw	0.008	−0.007	0.092	0.126				
$\hat{\beta}_N$	[0.66]	[0.20]	[1.24]	[1.02]				
implied $\hat{\beta}_L$ ‡	0.356	0.275	0.504	0.474	0.081	0.140	0.510	0.500
ê integrated ◇	I(1)	I(0)/I(1)	I(0)	I(0)	I(1)	I(1)	I(0)	I(0)
CD test *p*-value #	0.00	0.00	0.87	0.52	0.02	0.00	0.00	0.00
R-squared	0.94	0.85	1.00	1.00	0.84	0.66	1.00	1.00
RMSE	0.457	0.132	0.098	0.089	0.444	0.129	0.094	0.074

Source: Authors' analysis based on data sources discussed in the text.

Note: $N = 40$ countries, 918 observations, average $T = 23$. Dependent variable: value-added per worker (in logs). All variables are suitably transformed in the 2FE equations. Estimators: POLS, pooled OLS; 2FE, Two-way Fixed Effects; CCEP, Common Correlated Effects, Pooled version. We omit reporting the estimates on the intercept term. Absolute *t*-statistics reported in brackets are constructed using White heteroskedasticity-robust standard errors. For CCEP in [3], [4], [7], and [8], we report results on the basis of bootstrapped standard errors (100 replications). Time dummies are included explicitly in [1] and [5] or implicitly in [2] and [6]. Augmentation with cross-section averages in [3], [4], [7], and [8] (estimates not reported).

b The model includes cross-section averages for both the agricultural and manufacturing sector variables. † The log labor estimate represents $\hat{\beta}_L + \hat{\beta}_K (+\hat{\beta}_N) - 1$ and the returns to scale are based on the significance of this estimate: DRS, decreasing returns; IRS, increasing returns; CRS, constant returns to scale. ‡ Based on returns to scale and significant parameter estimates—see the main text. ◇ Order of integration of regression residuals is determined using Pesaran (2007) CIPS (full results available on request), H_0: nonstationary residuals. # Pesaran (2004) CD-test, H_0: cross-sectionally independent residuals. RMSE: root mean squared error. In common with most long T panel data the R-squared statistic is not very meaningful in this context and we therefore report the RSME.

* significant at the 5 percent level, ** significant at the 1 percent level

model in [7], with a capital coefficient of approximately .5 (as in agriculture data), does not pass the cross-section correlation test. However, further accounting for correlations across sectors in [8] yields favorable diagnostics and a similar capital coefficient. Following the imposition of CRS, all models reject cross-section independence, whereas parameter estimates are more or less identical to those in the unrestricted models. Based on these pooled regression results, the diagnostic tests (stationary and cross-section independent residuals) favor the CRS CCEP results in [3] and [4] for the agriculture data, whereas in the manufacturing data the unrestricted CCEP model in [8], which accounts for cross-sectoral impact, emerges as the preferred specification. Results for the other empirical models cannot be readily interpreted in the standard manner because of the presence of nonstationary and/or correlated residuals.[26]

In sum, relying on diagnostic testing, the alternative CCEP estimator emerges as the preferred estimator for both the agriculture and manufacturing samples. For agriculture, the imposition of CRS seems valid, whereas for manufacturing, the data reject this restriction. Across preferred specifications, the mean capital coefficients for agriculture and manufacturing are quite similar, approximately .5. Our shift to heterogeneous technology models, discussed in the next section, will allow us to determine whether these results are representative of the underlying technology. Although the CCEP imposes common technology coefficients, theory and simulations (Pesaran, 2006) have shown that if technology differs results reflect the mean coefficient across countries. However, outliers might exert undue influence on this mean. Therefore, our heterogeneous parameter models account for this possibility and report outlier-robust average coefficients.[27]

Averaged Country Regressions

Table 3 presents the robust means for each regressor across N country regressions for the unrestricted (Panel (A)) and CRS models (Panel (B)), respectively. The *t*-statistics reported for each average estimate test whether the average parameter is statistically different from zero, following Pesaran and Smith (1995). In addition, we report the share of countries for which the country results rejected CRS as well as the share of countries for which linear country trends are statistically significant (at the 10 percent level, respectively).

Beginning with the unrestricted models in Panel (A), we observe that MG and FDMG estimates for the agriculture and manufacturing equations are very imprecise. Furthermore, in the agriculture model, MG yields decreasing returns

26. The implication is that these empirical results are potentially spurious. We conduct a number of robustness checks adding further covariates in the agriculture equations (livestock per worker, fertilizer per worker) in the pooled regression framework. Results (available on request) do not change from those presented above. We also conduct robustness checks to include human capital in the estimation equation of both sectors. The results are presented in supplemental appendix S4 (see also discussion below).

27. We use robust regression to produce a robust estimate of the mean; see Hamilton (1992) and Eberhardt (2012) for details.

TABLE 3. Heterogeneous Parameter Models for Agriculture and Manufacturing (Robust Means)

	PANEL (A) UNRESTRICTED RETURNS TO SCALE							
	Agriculture				*Manufacturing*			
	[1] MG	[2] FDMG	[3] CMG	[4] CMG[b]	[5] MG	[6] FDMG	[7] CMG	[8] CMG[b]
Estimates								
log labor	−1.935	−0.474	−0.682	−0.068	−0.132	−0.127	0.069	0.003
$\hat{\beta}_L + \hat{\beta}_K (+ \hat{\beta}_N) - 1$	[2.43]*	[0.53]	[1.05]	[0.08]	[0.92]	[1.15]	[0.78]	[0.03]
log capital pw	−0.084	0.133	0.496	0.360	0.195	0.179	0.525	0.284
$\hat{\beta}_K$	[0.42]	[0.58]	[2.25]*	[1.37]	[1.32]	[1.12]	[6.46]**	[3.35]**
log land pw	−0.430	−0.269	−0.445	−0.129				
$\hat{\beta}_N$	[1.46]	[0.96]	[1.44]	[0.50]				
country trend/drift	0.015	0.010			0.015	0.018		
	[1.55]	[1.06]			[2.70]**	[3.31]**		
Diagnostics								
implied RS †	DRS	CRS	CRS	CRS	CRS	CRS	CRS	CRS
implied $\hat{\beta}_L$ ‡	n/a	n/a	0.504	n/a	n/a	n/a	0.475	0.717
reject CRS	0.38	0.20	0.23	0.23	0.50	0.13	0.38	0.25
sign. trends/drifts	0.40	0.18			0.40	0.20		
ê integrated ◇	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)
CD test *p*-value #	0.00	0.00	0.49	0.75	0.00	0.00	0.02	0.18
RMSE	0.081	0.094	0.069	0.059	0.080	0.077	0.068	0.047
Observations	918	872	918	918	918	872	918	918

PANEL (B) CONSTANT RETURNS TO SCALE IMPOSED

	Agriculture				*Manufacturing*			
	[1] MG	[2] FDMG	[3] CMG	[4] CMG[b]	[5] MG	[6] FDMG	[7] CMG	[8] CMG[b]
Estimates								
log capital pw	−0.050	0.300	0.538	0.620	0.291	0.346	0.509	0.413
$\hat{\beta}_K$	[0.29]	[2.22]*	[4.55]**	[2.98]**	[2.60]**	[3.64]**	[6.19]**	[6.37]**
log land pw	0.260	0.031	0.082	0.073				
$\hat{\beta}_N$	[1.03]	[0.20]	[0.47]	[0.38]				
country trend/drift	0.016	0.014			0.012	0.013		
	[2.71]**	[3.09]**			[2.72]**	[3.61]**		
Diagnostics								
implied $\hat{\beta}_L$ ‡	n/a	0.700	0.462	0.380	0.709	0.654	0.491	0.588
sign. trends/drifts	0.45	0.13			0.55	0.23		
ê integrated ◇	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)
CD test *p*-value #	0.00	0.00	0.93	0.73	0.00	0.00	0.00	0.00
RMSE	0.087	0.096	0.076	0.068	0.088	0.078	0.080	0.059
Observations	918	872	918	918	918	872	918	918

Source: Authors' analysis based on data sources discussed in the text.

Note: $N = 40$ countries, average $T = 23$ (21.8 for FDMG). Dependent variable: value-added per worker (in logs). All variables are suitably transformed in the FD equations. Estimators: MG, Mean Group; FDMG, MG with variables in first difference; CMG, Common Correlated Effects, Mean Group version. We report outlier-robust means; estimates on intercept terms are omitted. Absolute t-statistics are in brackets following Pesaran and Smith (1995). Estimates on cross-section averages in [3], [4], [7], and [8] are not reported.

b The model includes cross-section averages for both the agricultural and manufacturing sector variables. † Returns to scale are based on the significance of the log labor estimate. ‡ Based on returns to scale and significant parameter estimates—see the main text. 'reject CRS' and 'sign. trends/drifts' report the share of countries where CRS is rejected and where country trends/drifts are statistically significant (in both cases, applying a 10 percent level of significance). ◇ Order of integration of regression residuals, determined using Pesaran (2007) CIPS (full results available on request), H_0: nonstationary residuals. # Pesaran (2004) CD-test, H_0: cross-sectionally independent residuals. RMSE: root mean squared error.

* significant at the 5 percent level, ** significant at the 1 percent level

to scale that are nonsensical in magnitude. Simulations for nonstationary and cross-sectionally dependent data (Coakley, Fuertes and Smith, 2006; Bond and Eberhardt, 2009) show that MG estimates are severely affected by their failure to account for cross-section dependence, and this is the likely cause of these results. Standard CMG in agriculture and manufacturing yield similar capital coefficients of approximately .5, whereas the alternative CMG results provide somewhat lower estimates, approximately .3 (these models allow for agriculture sectors to influence manufacturing sectors and vice-versa). Diagnostics are sound in the case of the two CMG results in agriculture, but only for the alternative CMG estimator in manufacturing (cross-sectionally dependent residuals in model [7]). Panel (B) shows how the imposition of constant returns affects the results: MG and FDMG in both sectors are generally more sensible, but the diagnostic tests suggest cross-section correlation in the residuals that might indicate serious misspecification. The two CMG estimates for agriculture are now more similar. Land coefficients are still insignificant, but positive. Manufacturing results for the standard CMG remain virtually unchanged from the unrestricted model; however, diagnostic tests still indicate cross-sectionally dependent residuals. The same caveat applies to the alternative CMG for manufacturing.

In sum, the diagnostic tests support the use of the CRS versions of the CMG estimators for agricultural data and the unrestricted returns to scale version of the 'alternative' CMG estimator for the manufacturing data. These preferred models suggest that average technology differs across sectors, with a manufacturing capital coefficient of approximately .3 and an agriculture capital coefficient of approximately .5.[28]

The results for the land coefficient, where our preferred estimates indicate a positive, albeit statistically insignificant average coefficient, warrant additional comment. Given the relative persistence of the area under cultivation, the short time series dimension of the data might be responsible for this outcome. Any form of land quality adjustment would require time-varying information on land quality, which is not available at an annual rate over a long time

28. We further implement alternative specifications for both sectors that include the level and squared human capital terms (average years of schooling in the adult population) as additional covariates (see supplemental appendix S4). In the agriculture data, augmentation with human capital does not lead to statistically significant results (not reported). Manufacturing results for the MG and FDMG mirror those in the unaugmented models presented above. For the standard CMG models, we find capital coefficients somewhat below those in the unaugmented models but within each other's 95 percent confidence intervals (we do not estimate the 'alternative CMG estimator' with human capital because we encounter a dimensionality problem due to the large number of covariates). Average education coefficients are significant and indicate high returns to education in manufacturing: 11 percent and 12 percent in the unrestricted and CRS models, respectively.

horizon.[29] Time-invariant adjustments are accounted for by the country-specific intercepts.

Because of the aim of our study, we do not put too much emphasis on providing the best estimate for the 'true' sectoral technology coefficients. Instead, we highlight the discrepancy between these sectoral results and the results obtained when analyzing aggregate economy data.

IV. Aggregation Versus Heterogeneity

In this section, we provide practical evidence that the use of an aggregate production function will lead to severely biased technology estimates. We then provide some insights into the nature of technology heterogeneity across sectors and countries.

Aggregation Bias: Empirical Evidence

To investigate the impact of aggregation across heterogeneous sectors with technology furthermore differing across countries, we create a stylized 'aggregated economy' from our data on agriculture and manufacturing. To avoid the suggestion that our results might be critically distorted by this overly simplistic design, we compare them with those obtained from a matched sample of aggregate economy data from the PWT. Pre-estimation testing reveals that both datasets utilized in this section consist of nonstationary series that are cross-sectionally correlated; the results are provided in the supplemental appendix (S1, S2).[30]

We begin our discussion with the results for the pooled models in table 4. Across all specifications, the estimated capital coefficients in the stylized aggregated data far exceed those derived from the respective agriculture and manufacturing samples in table 2. Furthermore, the patterns across estimators are replicated one-to-one in the PWT data, which also yield excessively high capital coefficients across all models. All models suffer from cross-sectional dependence in the residuals. There are also indications that the residuals in the CCEP model for the aggregated data are nonstationary (those in the two other specifications in levels are always nonstationary). We also investigate the impact of human capital (via a proxy variable, average years of schooling

29. It can be argued that the CCE approach accounts for the induced bias for systematic distortion of the land variable. In Eberhardt, Helmers, and Strauss (forthcoming), we suggest that similar 'mismeasurement' of research and development investments leading to 'expensing' and 'double-counting' bias can be addressed in a common factor approach to the Griliches knowledge production function.

30. The supplemental appendix (S3) also contains details of an extensive simulation exercise in which we formulate a number of production technologies for agriculture and manufacturing, reflecting our insights into the effects of parameter heterogeneity, variable nonstationarity, and cross-section dependence and analyze stylized aggregate data constructed from these two sectors. This exercise suggests that, more than any other feature, the introduction of common factors (even different ones across sectors) creates the largest problems in the aggregate empirical results.

TABLE 4. Pooled Regression Models for Aggregated and PWT Data

	PANEL (A) UNRESTRICTED RETURNS TO SCALE					
	Aggregated data			*Penn World Table data*		
	[1] POLS	[2] 2FE	[3] CCEP	[4] POLS	[5] 2FE	[6] CCEP
Estimates						
log labor	0.010	−0.082	−0.054	0.035	−0.131	−0.097
$\hat{\beta}_L + \hat{\beta}_K (+ \hat{\beta}_N) - 1$	[1.32]	[3.75]**	[0.78]	[7.57]**	[4.57]**	[0.76]
log capital pw	0.828	0.798	0.657	0.742	0.704	0.631
$\hat{\beta}_K$	[107.55]**	[66.20]**	[19.43]**	[113.76]**	[51.43]**	[13.71]**
Diagnostics						
implied RS †	CRS	DRS	CRS	IRS	DRS	CRS
implied $\hat{\beta}_L$ ‡	0.172	0.120	0.343	0.293	0.165	0.369
ê integrated ◇	I(1)	I(1)	I(0)/I(1)	I(1)	I(1)	I(0)
CD test *p*-value #	0.40	0.00	0.04	0.10	0.00	0.00
R-squared	0.96	0.89	1.00	0.96	0.82	1.00
RMSE	0.358	0.109	0.078	0.195	0.095	0.061
observations	918	918	918	912	912	912

	PANEL (B) CONSTANT RETURNS TO SCALE IMPOSED					
	Aggregated data			*Penn World Table data*		
	[1] POLS	[2] 2FE	[3] CCEP	[4] POLS	[5] 2FE	[6] CCEP
Estimates						
log capital pw	0.825	0.824	0.666	0.730	0.745	0.651
$\hat{\beta}_K$	[120.48]**	[73.01]**	[20.85]**	[130.30]**	[63.41]**	[19.33]**
Diagnostics						
implied $\hat{\beta}_L$ ‡	0.175	0.176	0.334	0.270	0.255	0.349
ê integrated ◇	I(1)	I(1)	I(0)/I(1)	I(1)	I(1)	I(0)
CD test *p*-value #	0.31	0.30	0.06	0.00	0.00	0.00
R-squared	0.96	0.88	1.00	0.96	0.82	1.00
RMSE	0.358	0.109	0.086	0.202	0.097	0.069
observations	918	918	918	912	912	912

Source: Authors' analysis based on data sources discussed in the text.

Note: See table 2 for definitions and further details on diagnostic testing.

* significant at the 5 percent level, ** significant at the 1 percent level

TABLE 5. Heterogeneous Parameter Models for Aggregated and PWT Data (Robust Means)

	PANEL (A) UNRESTRICTED RETURNS TO SCALE					
	Aggregated data			*Penn World Table data*		
	[1] MG	[2] FDMG	[3] CMG	[4] MG	[5] FDMG	[6] CMG
Estimates						
log labor	−0.154	−0.079	0.117	−1.152	−1.681	−0.389
$\hat{\beta}_L + \hat{\beta}_K (+ \hat{\beta}_N) - 1$	[0.36]	[0.25]	[0.62]	[1.23]	[2.28]*	[1.03]
log capital pw	0.220	0.297	0.609	0.655	1.004	0.753
$\hat{\beta}_K$	[1.17]	[1.66]	[6.11]**	[4.22]**	[5.38]**	[5.26]**
country trend/drift	0.025	0.020		0.010	−0.010	
	[2.73]**	[2.42]*		[0.90]	[1.88]	
Diagnostics						
implied RS †	CRS	CRS	CRS	CRS	DRS	CRS
implied $\hat{\beta}_L$ ‡	n/a	n/a	0.391	0.345	n/a	0.247
reject CRS	0.60	0.23	0.38	0.68	0.33	0.53
sign. trends/drifts	0.55	0.33		0.43	0.18	
ê integrated ◇	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)
CD test *p*-value #	0.00	0.00	0.00	0.00	0.00	0.16
RMSE	0.081	0.094	0.051	0.080	0.077	0.041
observations	918	872	918	918	872	918

(Continued)

TABLE 5. Continued

	PANEL (B) UNRESTRICTED RETURNS TO SCALE					
	Aggregated data			*Penn World Table data*		
	[1] MG	[2] FDMG	[3] CMG	[4] MG	[5] FDMG	[6] CMG
Estimates						
log capital pw	0.293	0.202	0.725	0.619	0.923	0.811
$\hat{\beta}_K$	[1.92]	[1.90]	[10.95]**	[6.36]**	[6.01]**	[12.09]**
country trend/drift	0.014			0.002	−0.007	
	[2.93]**			[0.50]	[1.97]*	
Diagnostics						
implied $\hat{\beta}_L$ ‡	n/a	n/a	0.275	0.381	0.077	0.189
sign. trends/drifts	0.48	0.28		0.48	0.25	
ê integrated ◇	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)
CD test *p*-value #	0.00	0.00	0.05	0.00	0.00	0.00
RMSE	0.074	0.064	0.067	0.061	0.044	0.059
observations	918	872	918	912	866	912

Source: Authors' analysis based on data sources discussed in the text.

Note: See table 3 for definitions and further details on diagnostic testing.

* significant at the 5 percent level, ** significant at the 1 percent level.

attained in the population over 15 years of age) in these aggregate economy data models, but as the results in the supplemental appendix (S4) reveal, the basic bias remains.

In the results from averaged country regressions in table 5, the MG and FDMG models indicate differences between the aggregated and PWT data. The capital coefficients in the MG model are estimated very imprecisely but seem to center at approximately .3, whereas in the FDMG model, they are considerably higher, approximately .7 to .9. The results for the conceptually superior CMG, however, are very consistent between the two samples and across unrestricted and CRS models, with capital coefficients of approximately .7. Residual testing suggests that all specifications yield stationary residuals. Cross-section correlation tests reject independence in all but the PWT data unrestricted CMG residual series.

For ease of comparison, table 6 provides an overview of the preferred empirical results at the sectoral and aggregate data level, assuming common technology (top panel) or technology differences across countries (bottom panel).[31] Thus, across a large number of empirical specifications, we have found a systematic difference between the results for the sectoral data, on the one hand, and the results for the stylized aggregated and aggregate economy data, on the other hand. Theoretical work by Hsiao, Shen, and Fujiki (2005) provides insight into potential causes of this phenomenon. These authors find that if variable series are nonstationary and cointegrated at the 'micro unit' level (in their empirical illustration, in Japanese prefectures), then aggregation will only yield stable macro relations if all technology parameters are the same across units or if the weights used to construct the aggregate economy series from the micro units stay the same over time. In terms of our empirical question, time-invariant weights would imply the absence of any structural change in the economy over time, which clearly is not given here.

Technology Heterogeneity

Our empirical analysis has been based on the theoretical model first developed in Mundlak (1988). As the empirical implementations in Mundlak, Larson, and Butzer (1999) and Mundlak, Butzer, and Larson (2012), we have had to make simplifying assumptions to take this model to the data. By assuming parameter constancy over time, we have had to impose the same restriction on the parameter coefficients in the time series dimension as these studies. Our empirical model has, however, provided for more flexibility in the cross-section dimension, where we have allowed for parameter heterogeneity across countries

31. As a further robustness check, we ran regressions where, rather than aggregating the data, we forced manufacturing and agriculture production to follow the same technology using cross-equation restrictions. Results (available on request) did not differ qualitatively from the aggregated results presented above. Additionally, we estimated dynamic pooled models, introducing the PMG and CPMG estimators (for the results, see supplemental appendix S4). All of these results confirm the patterns across the sectoral and aggregated data described above.

TABLE 6 Comparison of Preferred Models

	PANEL (A) HOMOGENEOUS TECHNOLOGY			
	Sectoral Data		*Aggregate Data*	
	Agri [1] CCEP[b]	Manu [2] CCEP[b]	Stylized [3] CCEP	PWT [4] CCEP
Estimates				
log labor		0.002		−0.097
$\hat{\beta}_L + \hat{\beta}_K (+ \hat{\beta}_N) - 1$		[0.03]		[0.76]
log capital pw	0.526	0.469	0.666	0.631
$\hat{\beta}_K$	[6.70]**	[5.34]**	[20.85]**	[13.71]**
log land pw	0.126			
$\hat{\beta}_N$	[1.02]			
Diagnostics				
implied $\hat{\beta}_L$ ‡	0.474	0.532	0.334	0.369
ê integrated ◇	I(0)	I(0)	I(0)/I(1)	I(0)
CD test *p*-value #	0.52	0.93	0.06	0.00
RMSE	0.089	0.066	0.086	0.061
observations	918	918	918	912
	PANEL (B) HETEROGENEOUS TECHNOLOGY			
	Sectoral Data		*Aggregate Data*	
	Agri [1] CMG[b]	Manu [2] CMG[b]	Stylized [3] CMG	PWT [4] CMG
Estimates				
log labor		0.003		−0.389
$\hat{\beta}_L + \hat{\beta}_K (+ \hat{\beta}_N) - 1$		[0.03]		[1.03]
log capital pw	0.620	0.284	0.725	0.753
$\hat{\beta}_K$	[2.98]**	[3.35]**	[10.95]**	[5.26]**
log land pw	0.073			
$\hat{\beta}_N$	[0.38]			
Diagnostics				
implied $\hat{\beta}_L$ ‡	0.380	0.717	0.275	0.247
reject CRS (10%)		0.25		0.53
ê integrated ◇	I(0)	I(0)	I(0)	I(0)
CD test *p*-value #	0.73	0.18	0.05	0.16
RMSE	0.068	0.047	0.067	0.041
observations	918	918	918	912

Source: Authors' analysis based on data sources discussed in the text.

Note: See tables 2 and 3 for definitions and further details on diagnostic testing. In the agricultural regressions where the CCEP and CCEP[b] both had sound diagnostics (and very similar coefficient estimates), we report results for the CCEP[b] because it allows for greater flexibility.

* significant at the 5 percent level, ** significant at the 1 percent level

within each of the sectors. In the following, we critically review these modeling choices. First, we discuss our insights into technology heterogeneity across countries, and then, we provide evidence for parameter constancy.

From the empirical results in table 2, all pooled specifications, except for the CCEP estimators, yield residual series that are nonstationary. Therefore, we cannot rule out that the estimated coefficients are spurious. In addition the unrestricted POLS and 2FE models for agriculture as well as all POLS and 2FE models where the constant return to scale restriction has been imposed (a restriction rejected by the data) result in cross-sectionally dependent residual series. In contrast, the preferred heterogeneous parameter models for agriculture and manufacturing in table 6 do not suffer from nonstationary or cross-sectionally correlated residuals (or both). In conclusion, it appears that the data for both sectors reject the crucial assumptions underlying a pooled regression model (well-behaved residuals) and cannot reject those underlying a heterogeneous one. We interpret this evidence for misspecification in the pooled models as an indication of heterogeneous production technology within each sector of production.[32]

Given this finding for heterogeneity, one would naturally want to investigate the patterns of parameter heterogeneity across countries. With the specific data at our disposal (unbalanced panel, average $T = 23$), a closer analysis of whether we can identify discernible patterns must be interpreted with caution, and we view our results below as merely indicative. Previous empirical analysis averaging individual country regressions has frequently observed that although country estimates are widely dispersed and, at times, economically implausible, averages represent very plausible estimates (Boyd and Smith 2002; Baltagi *et al.* 2003). Pedroni (2007: 440) calls for caution when interpreting the estimates for any individual country because the "long-run signals contained in [limited] years of data may be relatively weak," whereas the cross-section averages will amplify the signal patterns sufficiently. Abstracting from the presence of common factors, Boyd and Smith (2002) discuss this issue somewhat more formally. Arguing for omitted variable bias in the country regression, assume a simple data generating process

$$y_{it} = \beta_i x_{it} + w_{it} + u_{it} \tag{10}$$

where w represents all variables omitted from the empirical model. Here, w is assumed to be correlated with the included regressor x in a particular country i and over a particular period of time T, indicated by the parameter subscript iT:

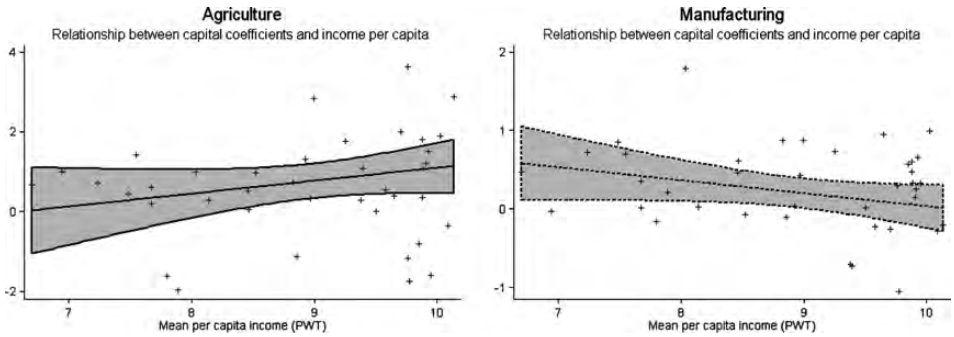
$$w_{it} = b_{iT} x_{it} + v_{it} \tag{11}$$

In a single country regression of y on x, we obtain

$$\mathrm{E}(\hat{\beta}_i) = \beta_i + b_{iT} \tag{12}$$

32. The importance of correctly specified technology heterogeneity in the presence of nonstationary processes is discussed in detail in Eberhardt and Teal (2011: 139f).

FIGURE 1. Investigating Technology Heterogeneity and Income



Note: These graphs investigate the issue of slope heterogeneity across countries. We plot the CMG country estimates for the capital coefficient β_K from the preferred heterogeneous agriculture and manufacturing models, corresponding to the models presented in columns [1] and [2] of table 6, Panel (B). The shaded areas represent the 90 percent confidence intervals of a linear regression of the respective capital coefficients on mean income per capita, where means are computed from aggregate PWT data over the entire 1963 to 1992 time horizon. Robust regressions of these relationships yield the following (statistically insignificant) slope parameters (standard errors in square brackets): .108 [.217] and −.079 [.087] for agriculture and manufacturing, respectively. For both plots we exclude outliers on the basis of weights computed from these robust regressions. Any coefficient with a weight less than .5 is excluded from the graph (for agriculture, five countries; for manufacturing, one country). *Source*: Authors' analysis based on data sources discussed in the text.

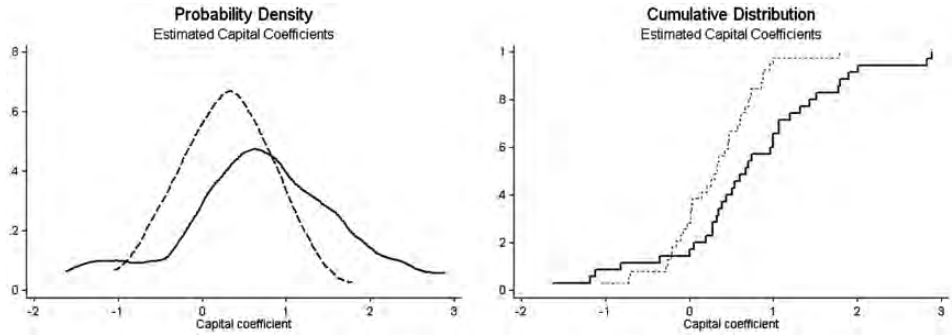
If the w_{it} are structural, operating in all time periods and countries, this would cause a systematic bias in the cross-country average estimate $\hat{\beta}^{MG}$.[33] If they are not structural but are only correlated in a particular subsample, they will lead to bias in these countries' estimates of β_i. However, averaging estimates across countries in this case yields $E(b_{iT}) = 0$, such that the biases cancel out in the average estimate $\hat{\beta}^{MG}$. The same principle applies to the CMG estimators in the presence of unobserved common factors.

We perform a basic analysis to obtain insight into the patterns of technology heterogeneity across countries. We begin by plotting the country-specific capital coefficients from the preferred agriculture and manufacturing models in table 6 against country mean aggregate income per capita (from PWT, in logs). Figure 1 presents individual country estimates and linear regression lines together with 90 percent confidence intervals for the two sectors.[34] Although the capital coefficients in agriculture appear to rise with income and those in manufacturing appear to fall, the confidence intervals indicate that neither

33. This is akin to ignoring common factors when these drive both y and x; see Eberhardt and Teal (2011: 137f).

34. We exclude the most extreme outliers from this plot using the following rule: we run a robust regression of the capital coefficients on mean income pc (in logs), reported in the note to figure 1, further computing the weights assigned to each observation by the algorithm. Countries with weights below 0.5 are then excluded (five countries in the agriculture and one country in the manufacturing sample).

FIGURE 2. Investigating Technology Heterogeneity across Sectors



Note: These graphs investigate the issue of slope heterogeneity across sectors. In the density plots on the left, we estimate separate Epanechnikov kernels (using common bandwidth .34) for the agriculture (solid line) and manufacturing (dashed line) capital coefficients from table 6, Panel (B); the right plots chart the cumulative distribution functions of the respective sector coefficients. For both sets of plots, we follow the same strategy as in figure 1 to exclude extreme outliers. *Source*: Authors' analysis based on data sources discussed in the text.

relationship is statistically precise, and (full-sample) robust regressions of the two equations yield statistically insignificant slope coefficients.[35]

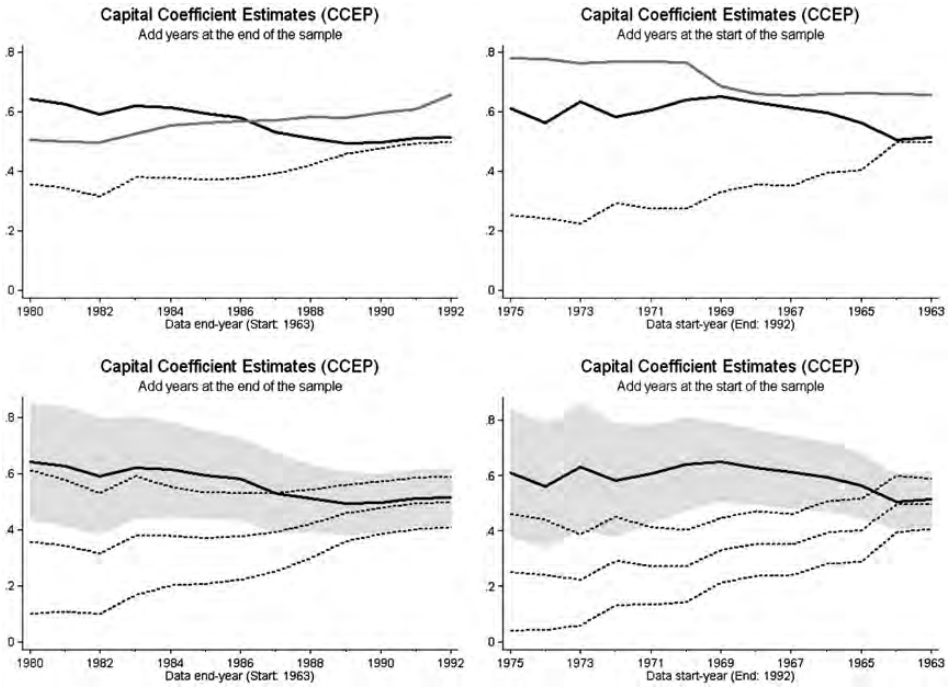
Figure 2 is somewhat less ambitious than the previous analysis. This figure provides density and distribution plots to highlight the differential distribution of capital coefficients in the agriculture and manufacturing equations. In the density plots on the left, manufacturing coefficients (dashed line) are distributed over a much narrower range than the agriculture coefficients. In other work on the cross-country production function in agriculture (Eberhardt and Teal forthcoming), we have argued that this heterogeneity[36] might be due in part to the difference in output structure (wheat vs. rice vs. livestock) and the commercialization of agriculture (subsistence vs. industrialized farming), both of which are functions of the level of development and productive specialization across countries. Manufacturing production, in comparison, represents a more homogeneous undertaking, such that the heterogeneity might be less pronounced. As the cumulative distribution plots on the right of figure 2 indicate, the robust means that we report in our regression results do not distort the underlying relative relationship, namely, that most agriculture coefficients are further to the right and thus larger than those for manufacturing.

The graphs in figures 3 and 4 address the question of slope parameter constancy over time by estimating each model with an increasing number of

35. We also replaced the mean income variable in this analysis with a number of proxies for institutions and 'social capital,' provided and investigated by Hall and Jones (1999). The patterns and significance levels for the correlations between sectoral capital coefficients and these alternative variables were very similar to those for the income correlations presented above.

36. Note that whether this refers to true technology heterogeneity or simply greater bias in the country regression for agriculture cannot be determined in this context.

FIGURE 3. Investigating Technology Constancy—Recursive Estimates (i)

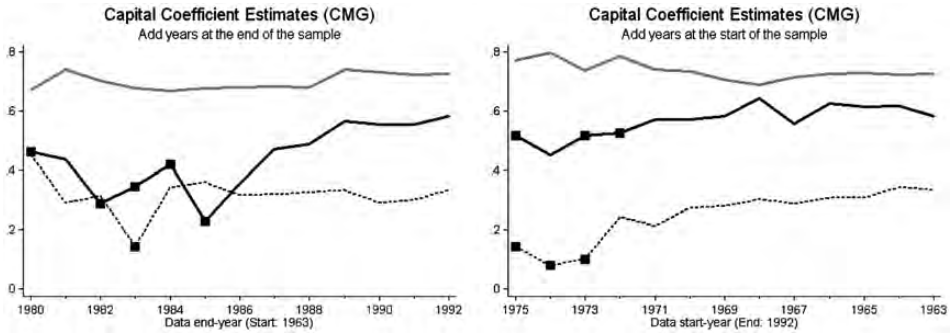


Note: These graphs investigate the issue of slope parameter constancy over time by estimating each model with an increasing number of observations and plotting the resulting estimates. We plot the robust estimates for the CCEP capital coefficients from the preferred agriculture, manufacturing, and aggregated data models, corresponding to the results presented in columns [1] to [3] of table 6, Panel (A). In each plot, the number of observations increases as we move to the right. In the left plots, all regressions include data from 1963 to 1979. The graphs then show the parameter estimates when we add one year of data at a time, at the end of the sample period, until we reach 1992. In the right plots, all regressions include data from 1976 to 1992. The graphs show the parameter estimates when we add one year at a time, at the beginning of the sample period, until we reach 1963. In each case, we begin (on the left of the plot) with a reduced sample where $T_i^{min} = 11$ and $T_i^{max} = 18$, corresponding to $n = 473$ (623 for the right plot) from $N = 34$ (38) countries. In each plot, the grey solid line represents aggregated data; black solid line, agriculture data; and black dashed line, manufacturing data. In the plots in the second row we indicate the 90 percent confidence intervals for the agriculture (grey area) and manufacturing (area between the dashed lines) estimates. Here, the estimates for the aggregated data are omitted to improve the legibility. *Source*: Authors' analysis based on data sources discussed in the text.

observations and plotting the resulting estimates.[37] We plot the estimates for the CCEP (in figure 3) and CMG (figure 4) capital coefficient $\hat{\beta}_K$ from the preferred agriculture, manufacturing and aggregated data models, corresponding to the models presented in columns [1] to [3] of table 6, Panels (A) and (B) for pooled and heterogeneous parameter models, respectively. In each plot, the

37. Following the example in our main results, we use robust means for the heterogeneous parameter models.

FIGURE 4. Investigating Technology Constancy—Recursive Estimates (ii)



Note: These graphs investigate the issue of slope parameter constancy over time by estimating each model with an increasing number of observations and plotting the resulting estimates. We plot the robust estimates for the CMG capital coefficients from the preferred agriculture, manufacturing and aggregated data models, corresponding to the results presented in columns [1] to [3] of table 6, Panel (B). See figure 3 for further details on how these plots are constructed. Squares indicate coefficients that are statistically insignificant at the 10 percent level. *Source*: Authors' analysis based on data sources discussed in the text.

number of observations increases as we move to the right. In the left plots, all regressions include data from 1963 to 1979. These graphs show the parameter estimates when we add one year of data at a time, at the end of the sample period, until we reach 1992. In the right plots, all regressions include data from 1976 to 1992. These graphs show the parameter estimates when we add one year at a time, at the beginning of the sample period, until we reach 1963. In each case, we begin (on the left of the plot) with a reduced sample, where $T_i^{min} = 11$ and $T_i^{max} = 18$, corresponding to $n = 473$ (623 for the right plot) observations from $N = 34$ (38) countries. The solid grey line indicates the results for the aggregated data, and solid and dashed black lines indicate results for agriculture and manufacturing, respectively. In the CCEP plots in the second row of figure 3, we indicate the 90 percent confidence intervals for the agriculture (grey area) and manufacturing (area between the dashed lines) estimates. The estimates for the aggregated data are omitted to improve legibility. In the CMG plots in figure 4, squares indicate that coefficients are statistically insignificant at the 10 percent level.

We use these graphs to provide insight into two specific questions: (i) From an econometric point of view, are the $\hat{\beta}_K$ coefficients on average constant over time? (ii) Following the suggestion in Mundlak, Butzer, and Larson (2012), if the β_K parameters are functions of common factors ('state variables,' in their terminology), implying that any estimated coefficient is a constant associated with the specific sample under analysis $(\hat{\beta}_K(s))$, we would expect results to vary over time given different samples. Do our recursive plots provide evidence for sample dependence in the estimated β_K coefficients? The answers to (i) and (ii) are

clearly dependent on each other because these questions seek the same information but are motivated from econometric and economic theory, respectively.

In the pooled specification where the preferred CCEP models yield relatively similar capital coefficients of approximately .5 in the full samples, the recursive regressions in figure 3 suggest that the agriculture (manufacturing) capital coefficient decreases (increases) over time as we increase our sample. Because the same pattern results whether we add years at the beginning or the end of the sample, it seems that this result is driven by small sample bias: as more observations become available in each country, the results become more precise. The associated confidence intervals included in the plots in the second row of the figure support this hypothesis. Coefficient estimates in the extreme left of each plot (the reduced sample) are contained within the 90 percent confidence interval of the coefficient estimates at the extreme right of each plot (the full sample). Turning to the heterogeneous parameter model estimates in figure 4, the robust mean coefficients marked with a square are statistically insignificant. If we eliminate these estimates from the graphs, we find remarkably stable recursive estimates for both the manufacturing and agriculture capital coefficients. Thus, the answer to question (i) on parameter constancy is a tentative 'yes.' The answer to question (ii) on sample dependence is a tentative 'no.' The former answer suggests that the assumption $\beta_{it} = \beta_i$ is valid, and the latter answer implies that we find no evidence for a systematic relationship between technology coefficients and unobserved time-varying factors (or state variables).

V. Concluding Remarks

In this paper, we employed unique panel data for agriculture and manufacturing sectors to estimate sector-level and aggregate production functions. Our empirical analysis emphasized contributions from the recent panel time series econometrics literature and, in particular, emphasized the importance of parameter heterogeneity across countries as well as sectors. In addition, we took the nonstationarity of observable and unobservable factor inputs into account and addressed concerns over cross-sectional dependence commonly found in macro panel data.

We draw the following conclusions from our attempts to highlight the importance of structural makeup and change for the empirical analysis of cross-country growth and development. First, duality matters. The empirical analysis of growth and development across countries benefits significantly from the consideration of the modern and traditional sectors that make up a developing economy. Comparing our analysis of agriculture and manufacturing with that of a stylized aggregated economy suggests that the latter analysis yields severely distorted empirical results with serious implications for estimates of TFP derived from aggregate analysis. An analysis of PWT data in parallel with the aggregated data suggests that this finding is not an artifact of our stylized empirical setup. Growth accounting exercises at the aggregate economy level thus

provide misleading results in that any technology differences across sectors within countries are assumed away, and the constructed TFP series might reflect this misspecification rather than true technological progress.

Second, focusing on technology and TFP within each sector, we find that the data rejected empirical specifications that impose common technology, common TFP evolution, and the independence of shocks across countries. Thus, the assumption of common technology in the existing work on the dual economy model using growth accounting methods is not in line with the data. If these restrictions were correct, we should be able to find pooled technology models that satisfy the most basic assumptions of stationary and cross-sectionally independent residuals. In practice, however, we find results that are much more in line with the notion of differential technology across countries, for which we have provided support from economic theory.

Third, the presence of unobserved common factors, both as latent processes driving all observables and as a conceptual framework for TFP, has been shown to have a substantial impact on empirical results. Much of the cross-country empirical literature ignores the presence of global economic shocks with heterogeneous impact and spillovers across country borders. With the experience of the recent global financial crisis, it is now more evident than ever that economic performance in a globalized world is highly interconnected and that domestic markets cannot 'de-couple' from the global financial and goods markets. In econometric terms, latent forces drive all of the observable and unobservable variables and processes that we attempt to model. An important implication is that commonly applied instruments in cross-country growth regressions are invalid, a sentiment that is echoed in recent work by Bazzi and Clemens (2009). We argue that panel time series methods allow us to develop a new type of cross-country empirics that is more informative and more flexible in the problems that it can address than its critics have allowed.

Fourth, we are aware of the serious data limitations for sectoral data from developing economies, particularly regarding the high data requirements of panel time series methods. The Crego, Larson, Butzer, and Mundlak (1998) dataset allowed us to directly compare sectoral analysis between manufacturing and agriculture. However, for alternative research questions, the use of data from one sector or the other might be sufficient. There are at least two existing data sources, FAO data for agriculture and UNIDO data for manufacturing, which are ideally suited to inform this type of analysis at the sector level for a large number of countries and over a substantial period of time.

Cross-country panel data play a crucial role in policy analysis for development. The present work represents a first step toward establishing an empirical version of a dual economy model to inform this literature. From the perspective of dual economy theory, we have only analyzed one aspect of the canon, technology heterogeneity between traditional and modern sectors of production. In future work, we will implement empirical tests to investigate the suggested

sources of growth arising from this literature, including marginal factor product differences and heterogeneous TFP levels for growth across sectors.

Appendix

Data construction and descriptive statistics

We use a total of four datasets in our empirical analysis, consisting of data for agriculture and manufacturing (Crego, Larson, Butzer, and Mundlak 1998; UNIDO 2004; FAO 2007), an 'aggregated dataset' in which the labor, output, and capital stock values for the two sectors are summed, and the PWT 6.2 dataset (Heston, Summers, and Aten 2006) for comparative purposes. The first three datasets differ significantly in their construction from the last, primarily in the choice of exchange rates and deflation: the first three datasets use international exchange rates for the year 1990, whereas the PWT dataset uses international dollars (purchasing power parity-adjusted) with the year 2000 as the comparative base. The first three datasets thus emphasize traded goods, whereas the PWT is generally perceived to better account for nontradables and service. Provided that all monetary values

Table A1. Descriptive Statistics: Sample Makeup for all Datasets

#	ISO	COUNTRY	OBS	#	ISO	COUNTRY	OBS
1	AUS	Australia	20	22	KEN	Kenya	29
2	AUT	Austria	22	23	KOR	South Korea	29
3	BEL	Belgium-Luxembourg	22	24	LKA	Sri Lanka	17
4	CAN	Canada	30	25	MDG	Madagascar	20
5	CHL	Chile	20	26	MLT	Malta	23
6	COL	Colombia	26	27	MUS	Mauritius	16
7	CYP	Cyprus	18	28	MWI	Malawi	23
8	DNK	Denmark	26	29	NLD	Netherlands	23
9	EGY	Egypt	24	30	NOR	Norway	22
10	FIN	Finland	28	31	NZL	New Zealand	19
11	FRA	France	23	32	PAK	Pakistan	24
12	GBR	United Kingdom	22	33	PHL	Philippines	24
13	GRC	Greece	28	34	PRT	Portugal	20
14	GTM	Guatemala	19	35	SWE	Sweden	23
15	IDN	Indonesia	22	36	TUN	Tunisia	17
16	IND	India	29	37	USA	United States	23
17	IRL	Ireland	23	38	VEN	Venezuela	19
18	IRN	Iran	25	39	ZAF	South Africa	26
19	ISL	Iceland	20	40	ZWE	Zimbabwe	25
20	ITA	Italy	21				
21	JPN	Japan	28			**Total**	**918**

Source: Authors' analysis based on data sources discussed in the text.

Note: ISO indicates the three-letter ISO code for each country; OBS reports the number of observations (levels regression).

TABLE A2. Descriptive Statistics

Agriculture						*Manufacturing*					
PANEL (A) VARIABLES IN UNTRANSFORMED LEVELS TERMS											
variable	**mean**	**median**	**st. dev.**	**min.**	**max.**	**variable**	**mean**	**median**	**st. dev.**	**min.**	**max.**
Output	1.8E + 10	6.0E + 09	3.0E + 10	3.5E + 07	2.2E + 11	Output	7.6E + 10	8.8E + 09	2.1E + 11	7.2E + 06	1.4E + 12
Labor	9.6E + 06	1.3E + 06	3.5E + 07	3.0E + 03	2.3E + 08	Labor	1.7E + 06	4.8E + 05	3.4E + 06	9.6E + 03	2.0E + 07
Capital	6.5E + 10	1.1E + 10	1.5E + 11	2.9E + 07	8.6E + 11	Capital	1.3E + 11	2.0E + 10	3.0E + 11	1.4E + 07	1.8E + 12
Land	1.8E + 07	3.5E + 06	4.1E + 07	6.0E + 03	1.9E + 08						
in logarithms											
Output	22.39	22.51	1.73	17.38	26.13	Output	22.84	22.89	2.29	15.79	27.99
Labor	14.00	14.04	2.02	8.01	19.27	Labor	13.10	13.08	1.65	9.17	16.79
Capital	22.96	23.07	2.28	17.18	27.48	Capital	23.64	23.74	2.27	16.46	28.22
Land	15.11	15.07	1.99	8.70	19.07						
in growth rates (percent)											
Output	1.7	1.9	10.4	−41.5	53.9	Output	4.4	3.9	10.1	−40.9	84.2
Labor	−0.6	−0.0	3.0	−28.8	13.4	Labor	1.9	1.1	6.8	−38.8	78.1
Capital	1.9	1.2	3.6	−5.1	31.4	Capital	4.8	3.6	5.0	−5.1	53.0
Land	0.1	0.0	2.2	−23.1	13.6						
PANEL (B) VARIABLES IN PER WORKER TERMS											
variable	**mean**	**median**	**st. dev.**	**min.**	**max.**	**variable**	**mean**	**median**	**st. dev.**	**min.**	**max.**
Output	12,724	6,644	13,161	44.18	57,891	Output	27,093	20,475	22,111	753	101,934
Capital	52,367	9,925	63,576	13.10	222,397	Capital	63,533	43,577	64,557	1,475	449,763
Land	9.66	3.00	20.34	0.29	110						
in logarithms											
Output	8.39	8.80	1.83	3.79	10.97	Output	9.74	9.93	1.09	6.62	11.53
Capital	8.96	9.20	2.71	2.57	12.31	Capital	10.54	10.68	1.09	7.30	13.02
Land	1.11	1.10	1.41	−1.24	4.70						
in growth rates (percent)											
Output	2.3	2.5	10.5	−43.7	56.0	Output	2.5	2.5	9.0	−67.0	73.0
Capital	2.5	2.0	4.2	−7.8	31.1	Capital	2.9	2.9	6.6	−71.7	42.4
Land	0.7	0.5	3.4	−18.4	28.8						

(Continued)

TABLE A2. Continued

Aggregated Data						Penn World Table Data					
PANEL (A) VARIABLES IN UNTRANSFORMED LEVELS TERMS											
variable	**mean**	**median**	**st. dev.**	**min.**	**max.**	**variable**	**mean**	**median**	**st. dev.**	**min.**	**max.**
Output	9.3E + 10	1.7E + 10	2.3E + 11	1.1E + 08	1.6E + 12	Output	4.3E + 11	1.3E + 11	1.0E + 12	1.3E + 09	8.0E + 12
Labor	1.1E + 07	2.4E + 06	3.6E + 07	2.2E + 04	2.4E + 08	Labor	5.1E + 07	1.3E + 07	1.2E + 08	2.1E + 05	8.5E + 08
Capital	2.0E + 11	2.9E + 10	4.3E + 11	1.0E + 08	2.3E + 12	Capital	1.2E + 12	3.3E + 11	2.9E + 12	3.3E + 09	2.3E + 13
in logarithms											
Output	23.50	23.58	2.01	18.55	28.07	Output	25.44	25.58	1.71	21.02	29.71
Labor	14.66	14.67	1.74	10.01	19.30	Labor	16.49	16.41	1.63	12.27	20.57
Capital	24.10	24.08	2.21	18.44	28.44	Capital	26.38	26.52	1.80	21.92	30.75
in growth rates (percent)											
Output	3.1	3.1	7.4	−33.9	42.1	Output	4.0	4.0	5.0	−37.1	26.6
Labor	0.2	0.4	2.6	−11.4	19.3	Labor	1.5	1.4	1.1	−1.9	4.8
Capital	3.6	2.7	3.6	−5.0	25.1	Capital	4.6	4.2	2.9	−1.3	16.4
PANEL (B) VARIABLES IN PER-WORKER TERMS											
variable	**mean**	**median**	**st. dev.**	**min.**	**max.**	**variable**	**mean**	**median**	**st. dev.**	**min.**	**max.**
Output	19,493	11,197	19,212	72	76,031	Output	11,445	10,630	8,193	594	31,074
Capital	49,634	23,140	55,541	53	236,312	Capital	37,059	32,981	31,765	661	136,891
in logarithms											
Output	8.84	9.32	1.85	4.28	11.24	Output	8.95	9.27	1.02	6.39	10.34
Capital	9.44	10.05	2.20	3.96	12.37	Capital	9.87	10.40	1.37	6.49	11.83
in growth rates (percent)											
Output	3.0	3.3	7.0	−31.0	44.5	Output	2.5	2.6	5.0	−41.2	23.2
Capital	3.4	3.2	3.8	−18.4	22.2	Capital	3.1	2.8	2.9	−4.2	14.3

Source: Authors' analysis based on data sources discussed in the text.

Note: We report the descriptive statistics for value-added (in US dollars for the year 1990 or purchasing power parity-adjusted international dollars for the year 2000), labor (headcount), capital stock (the same monetary values as VA in each respective dataset), and land (in hectares) for the regression sample (levels sample: $n = 918$; $N = 40$).

incorporated in the variables for each regression are comparable (across countries and over time) and given that the comparison of sectoral and aggregated data with the PWT is primarily intended for illustration purposes, we have no concerns about presenting results from these two conceptually different datasets.

In all cases, the results presented are for matched observations across datasets, so that the four datasets are identical in terms of country and time period coverage. We prefer this design for direct comparison even though more observations are available for individual data sources, which could improve the robustness of empirical estimates. We provide details on the sample makeup in table A1. The next two subsections describe the data construction. Descriptive statistics for all variables in the empirical analysis are presented in table A2.

Sectoral and aggregated data

Investment Data. Data for agricultural and manufacturing investment (AgSEInv, MfgSEInv) in constant year 1990 local currency units (LCU), the US$-LCU exchange rate (Ex_Rate, see comment below), and sector-specific deflators (AgDef, TotDef) were taken from Crego, Larson, Butzer, and Mundlak (1998).[38] Note that these authors also provide capital stock data, which they produced through their own calculations from the investment data. Following Martin and Mitra (2002), we believe that the use of a single year exchange rate is preferable to the use of annual rates in the construction of real output (see next paragraph) and capital stock (see below).

Output data. For manufacturing, we use data on aggregate GDP in current LCU and the share of GDP in manufacturing from the World Bank WDI (World Bank, 2008). For agriculture, we use agricultural value-added in current LCU from the same source. The two sectoral value-added series are then deflated using the Crego, Larson, Butzer, and Mundlak (1998) sectoral deflator for agriculture and the total economy deflator for manufacturing before we use the 1990 US$-LCU exchange rates to make them comparable across countries.

The currencies used in the Crego, Larson, Butzer, and Mundlak (1998) data differ from those applied in the WDI data for a number of European countries because of the adoption of the Euro. Therefore, we must use alternative 1990 US$-LCU exchange rates for these economies.[39]

38. Data are available at http://go.worldbank.org/FS3FXW7461. All data discussed in this appendix are linked at http://sites.google.com/site/medevecon/devecondata. Stata code for empirical estimators and tests is available from SSC: pescadf, xtmg, xtcd. See also Eberhardt (2012) on xtmg.

39. In detail, we apply exchange rates of 1.210246384 for AUT, 1.207133927 for BEL, 1.55504706 for FIN, 1.204635181 for FRA, 2.149653527 for GRC, 1.302645017 for IRL, 1.616114954 for ITA, 1.210203555 for NLD, and 1.406350856 for PRT. See table A1 for country codes.

Labor data. For agriculture, we adopt the variable 'economically active population in agriculture' from the FAO's (2007) PopSTAT. Manufacturing labor is taken from UNIDO's (2004) INDSTAT.

Additional data. The land variable is taken from ResourceSTAT and represents 'arable and permanent crop land' (measured in hectares) (FAO 2007). For the robustness checks (results available on request), the livestock variable is constructed from the data for the following animals in the 'live animals' section of ProdSTAT: asses (donkeys), buffalos, camels, cattle, chickens, ducks, horses, mules, pigs, sheep, goats, and turkeys. Following convention, we use the formula below to convert the numbers for individual animal species into the livestock variable:

$$\begin{aligned}\text{livestock} = {} & 1.1 \text{ camels} + \text{buffalos} + \text{horses} + \text{mules} \\ & + 0.8 \text{ cattle} + 0.8 \text{ asses} + 0.2 \text{ pigs} \\ & + 0.1 \text{ (sheep} + \text{goats)} \\ & + 0.01 \text{ (chickens} + \text{ducks} + \text{turkeys)}\end{aligned}$$

The fertilizer variable is taken from the 'fertilizers archive' of ResourceSTAT and represents 'agricultural fertilizer consumed in metric tons,' which includes 'crude' and 'manufactured' fertilizers. For human capital, we employ years of schooling attained in the population by those aged 25 years and above, from Barro and Lee (2001), interpolated to create an annual series.

Capital stock. We construct capital stock in agriculture and manufacturing by applying the perpetual inventory method described in detail in Klenow and Rodriguez-Clare (1997b), using the investment data from Crego, Larson, Butzer, and Mundlak (1998), which are transformed into US dollars by applying the 1990 US$-LCU exchange rate. For the construction of a sectoral base year capital stock in each country i, we employ average sector value-added growth rates g_{ij} (using the deflated sectoral value-added data), the average sectoral investment to value-added ratio $(I/Y)_{ij}$ and an assumed depreciation rate of 5 percent to construct

$$(K/Y)_{0ij=}(1/Y)_{ij}/(g_{ij} + 0.05)$$

for sector j (agriculture, manufacturing). This ratio is then multiplied by the sectoral value-added data for the base year to yield K_{0ij}. Note that the method deviates from that discussed in Klenow and Rodriguez-Clare (1997b) because they use per capita GDP in their computations and therefore need to account for population growth in the construction of the base year capital stock.

Aggregated data. We combine the agriculture and manufacturing data to produce a stylized 'aggregate economy.' For labor, we simply sum the headcount; for the monetary representations of output and capital stock, the same treatment is applied. Crego, Larson, Butzer, and Mundlak (1998) developed

the first large panel dataset that provides data on investment in agriculture for a long span of time, and their work affords us this ability to sum variables for the two sectors.

Penn World Table Data

As a means of comparison, we also provide production function estimates using data from PWT version 6.2. We adopt real per capita GDP in international dollars Laspeyeres (*rgdpl*) as measure for output and construct capital stock using investment data (derived from the investment share in real GDP, *ki*, and the output variable, *rgdpl*) in the perpetual inventory method described above, again adopting 5 percent depreciation (at this point, we must use the data on population from PWT, *pop*, to compute the average annual population growth rate).

References

Abramowitz, Moses. 1956. "Resource and Output Trends in the United States since 1870." *American Economic Review* 46(2): 5–23.

Arellano, Manuel, and Stephen R. Bond. 1991. "Some Tests of Specification for Panel Data." *Review of Economic Studies* 58(2): 277–297.

Azariadis, Costas, and Allan Drazen. 1990. "Threshold Externalities in Economic Development." *The Quarterly Journal of Economics* 105(2): 501–26.

Bai, Jushan, and Chihwa Kao. 2006. "On the Estimation and Inference of a Panel Cointegration Model with Cross-Sectional Dependence." In *Panel Data Econometrics: Theoretical Contributions and Empirical Applications,* ed. Badi H. Baltagi. Amsterdam: Elsevier Science.

Bai, Jushan, Chihwa Kao, and Serena Ng. 2009. "Panel Cointegration with Global Stochastic Trends." *Journal of Econometrics* 149(1): 82–99.

Bai, Jushan, and Serena Ng. 2008. "Large Dimensional Factor Analysis." *Foundations and Trends in Econometrics* 3(2): 89–163.

Baier, Scott L., Gerald P. Dwyer, and Robert Tamura. 2006. "How Important are Capital and Total Factor Productivity for Economic Growth?" *Economic Inquiry* 44(1): 23–49.

Bailey, Natalia, George Kapetanios, and M. Hashem Pesaran. 2012. *Exponent of Cross-sectional Dependence: Estimation and Inference*. Cambridge University, Faculty of Economics, January 2012.

Baltagi, Badi H., Georges Bresson, James M. Griffin, and Alain Pirotte. 2003. "Homogeneous, Heterogeneous or Shrinkage Estimators? Some Empirical Evidence from French Regional Gasoline Consumption." *Empirical Economics* 28(4): 795–811.

Banerjee, Abhijit V., and Andrew F. Newman. 1993. "Occupational Choice and the Process of Development." *Journal of Political Economy*, 101(2): 274–98.

Banerjee, Anindya, Markus Eberhardt, and J. James Reade. 2010. "Panel Econometrics for Worriers." Oxford University, Department of Economics Discussion Paper Series No. 514.

Barro, Robert J., and Jong-Wha Lee. 2001. "International Data on Educational Attainment: Updates and Implications." *Oxford Economic Papers* 53(3): 541–63.

Barro, Robert J., and Jong-Wha Lee. 2010. "A New Data Set of Educational Attainment in the World, 1950-2010." NBER Working Paper No. 15902, NBER, Washington, DC.

Basu, Susantu, and David N. Weil. 1998. "Appropriate Technology and Growth." *The Quarterly Journal of Economics* 113(4): 1025–1054.

Bazzi, Samuel, and Michael Clemens. 2009. *Blunt Instruments: On Establishing the Causes of Economic Growth*, Working Paper No. 171, Center for Global Development, Washington, DC.

Binder, Michael, and Christian J. Offermanns. 2007. "International Investment Positions and Exchange Rate Dynamics: A Dynamic Panel Analysis." Deutsche Bundesbank: Discussion Paper Series 1, Economic Studies No. 2007/23.

Blundell, Richard, and Stephen R. Bond. 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *Journal of Econometrics* 87(1): 115–43.

Bond, Stephen R. 2002. "Dynamic Panel Data Models: A Guide to Micro Data Methods and Practice." *Portuguese Economic Journal* 1(2): 141–62.

Bond, Stephen R., and Markus Eberhardt. 2009. "Cross-Section Dependence in Nonstationary Panel Models: A Novel Estimator." Paper presented at the Nordic Econometrics Meeting in Lund, Sweden, October 29–31.

Boyd, Derick A. C., and Ron P. Smith. 2002. "Some Econometric Issues in Measuring the Monetary Transmission Mechanism, with an Application to Developing Countries." In *Monetary Transmission in Diverse Economies*, ed. Lavan Mahadeva, and Peter Sinclair, Cambridge University Press.

Brock, William A., and Steven N. Durlauf. 2001. "Growth Economics and Reality." *World Bank Economic Review* 15(2): 229–72.

Butzer, Rita. (2011). *The Role of Physical Capital in Agricultural and Manufacturing Production*. PhD thesis, Department of Economics, University of Chicago, June 2011.

Cavalcanti, Tiago, Kamiar Mohaddes, and Mehdi Raissi. 2011. "Growth, Development and Natural Resources: New Evidence using a Heterogeneous Panel Analysis." *The Quarterly Review of Economics and Finance* 51(4): 305–18.

Clark, Gregory. 2007. *A Farewell to Alms: A Brief Economic History of the World*. Princeton University Press.

Coakley, Jerry, Ana-Maria Fuertes, and Ron P. Smith. 2006. "Unobserved Heterogeneity In Panel Time Series Models." *Computational Statistics & Data Analysis* 50(9): 2361–2380.

Crego, Al, Donald Larson, Rita Butzer, and Yair Mundlak. 1998. "A New Database on Investment and Capital for Agriculture and Manufacturing." Policy Research Working Paper Series No. 2013, The World Bank, Washington, DC.

Durlauf, Steven N. 1993. "Nonergodic economic growth." *Review of Economic Studies* 60(2): 349–66.

Durlauf, Steven N., Paul A. Johnson, and Jonathan R. Temple. 2005. "Growth Econometrics." In Philippe Aghion, and Steven N. Durlauf, eds., *Handbook of Economic Growth Vol. 1*, 555–677. Amsterdam: Elsevier.

Durlauf, Steven N., Andros Kourtellos, and Artur Minkin. 2001. "The Local Solow Growth Model." *European Economic Review* 45(4–6): 928–40.

Durlauf, Steven N., and Danny T. Quah. 1999. "The New Empirics of Economic Growth." In John B. TaylorMichael Woodford, eds., *Handbook of Macroeconomics Vol. 1*, 235–308. Amsterdam: Elsevier.

Easterly, William. 2002. *The Elusive Quest for Growth — Economists' Adventures and Misadventures in the Tropics*. Cambridge, MA: MIT Press.

Easterly, William, and Ross Levine. 2001. "It's Not Factor Accumulation: Stylised Facts and Growth Models." *World Bank Economic Review* 15(2): 177–219.

Eberhardt, Markus. 2012. "Estimating Panel Time Series Models with Heterogeneous Slopes." *Stata Journal* 12(1): 61–71.

Eberhardt, Markus, and Christian Helmers. 2010. "Untested Assumptions and Data Slicing: A Critical Review of Firm-Level Production Function Estimators." Oxford University, Department of Economics Discussion Paper Series No. 513.

Eberhardt, Markus, Christian Helmers, and Hubert Strauss. Forthcoming. "Do Spillovers Matter when Estimating Private Returns to R&D?" *The Review of Economics and Statistics*. DOI: 10.1162/REST_a_00272.

Eberhardt, Markus, and Francis Teal. 2011. "Econometrics for Grumblers: A New Look at the Literature on Cross-Country Growth Empirics." *Journal of Economic Surveys* 25(1): 109–55.

———. Forthcoming. "No Mangos in the Tundra: Spatial Heterogeneity in Agricultural Productivity Analysis." *Oxford Bulletin of Economics and Statistics*. DOI: 10.1111/j.1468-0084.2012.00720.x.

FAO. 2007. *FAOSTAT*. Online database, Rome: FAO, United Nations Food and Agriculture Organisation. http://faostat.fao.org/.

Granger, Clive W. J. 1997. "On Modelling the Long Run in Applied Economics." *Economic Journal* 107(440): 169–77.

Hall, Robert E., and Charles I. Jones. 1999. "Why do Some Countries Produce so Much More Output Per Worker than Others?" *The Quarterly Journal of Economics* 114(1): 83–116.

Hamilton, Lawrence C. 1992. "How Robust is Robust Regression?" *Stata Technical Bulletin* 1(2): 21–25.

Heston, Alan, Robert Summers, and Bettina Aten. 2006. *Penn World Table version 6.2*. Center for International Comparisons of Production, Income and Prices, University of Pennsylvania. http://pwt.econ.upenn.edu/php_site/pwt_index.php.

———. 2011. *Penn World Table version 7.0*. Center for International Comparisons of Production, Income and Prices, University of Pennsylvania. http://pwt.econ.upenn.edu/php_site/pwt_index.php.

Holly, Sean, M. Hashem Pesaran, and Takashi Yamagata. 2010. "A Spatio-Temporal Model Of House Prices in the US." *Journal of Econometrics* 158(1): 160–73.

Hsiao, Cheng, Yan Shen, and Hiroshi Fujiki. 2005. "Aggregate vs. Disaggregate Data Analysis — A Paradox in the Estimation of a Money Demand Function of Japan under the Low Interest Rate Policy." *Journal of Applied Econometrics* 20(5): 579–601.

Jerzmanowski, Michal. 2007. "Total Factor Productivity Differences: Appropriate Technology vs. Efficiency." *European Economic Review* 51(8): 2080–110.

Johnson, Simon, William Larson, Chris Papageorgiou, and Arvind Subramanian. 2009. "Is Newer Better? Penn World Table Revisions and Their Impact on Growth Estimates." NBER Working Papers No. 15455. NBER, Washington, DC.

Jorgensen, Dale W. 1961. "The Development of a Dual Economy." *Economic Journal* 71(282): 309–34.

Kao, Chihwa. 1999. "Spurious Regression and Residual-Based Tests for Cointegration in Panel Data." *Journal of Econometrics* 65(1): 9–15.

Kapetanios, George, M. Hashem Pesaran, and Takashi Yamagata. 2011. "Panels with Nonstationary Multifactor Error Structures." *Journal of Econometrics* 160(2): 326–48.

Klenow, Peter J., and Andres Rodriguez-Clare. 1997a. "Economic Growth: A Review Essay." *Journal of Monetary Economics* 40(3): 597–617.

———. 1997b. "The Neoclassical Revival in Growth Economics: Has it Gone too Far?" *NBER Macroeconomics Annual* 12: 73–103.

Lee, Kevin, M. Hashem Pesaran, and Ron P. Smith. 1997. "Growth and Convergence in a Multi-Country Empirical Stochastic Solow Model." *Journal of Applied Econometrics* 12(4): 357–92.

Lewis, W. Arthur. 1954. "Economic Development with Unlimited Supplies of Labour." *The Manchester School* 22: 139–91.

Lin, Justin Y. 2011. "New Structural Economics: A Framework for Rethinking Development." *World Bank Research Observer* 26(2): 193–221.

Mankiw, N. Gregory, David Romer, and David N. Weil. 1992. "A Contribution to the Empirics of Economic Growth." *The Quarterly Journal of Economics* 107(2): 407–37.

Marschak, Jacob, and William H. Andrews 1944. "Random Simultaneous Equations and the Theory of Production." *Econometrica* 12(3/4): 143–205.

Martin, Will, and Devashish Mitra. 2002. "Productivity Growth and Convergence in Agriculture versus Manufacturing." *Economic Development and Cultural Change* 49(2): 403–22.

McMillan, Margaret, and Dani Rodrik. 2011. "Globalization, Structural Change and Productivity Growth." NBER Working Papers No. 17143, NBER, Washington DC.

Moscone, Francesco, and Elisa Tosetti. 2009. "A Review and Comparison of Tests of Cross-Section Independence in Panels." *Journal of Economic Surveys* 23(3): 528–61.

———. 2010. "Health Expenditure and Income in the United States." *Health Economics* 19(12): 1385–403.

Mundlak, Yair. 1988. "Endogenous Technology and the Measurement of Productivity." In Susan M. Capalbo and John M. Antle, eds., *Agricultural Productivity: Measurement and Explanation*. Washington, DC: Resources for the Future.

Mundlak, Yair, Donald Larson, and Rita Butzer. 1999. "Rethinking within and between Regressions: The Case of Agricultural Production Functions." *Annales D'Economie et de Statistique* 55/56: 475–501.

Mundlak, Yair, Rita Butzer, and Donald Larson. 2012. "Heterogeneous Technology and Panel Data: The Case of the Agricultural Production Function." *Journal of Development Economics* 99(1): 139–49.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny. 1989. "Industrialization and the Big Push." *Journal of Political Economy* 97(5): 1003–26.

Nelson, Charles R., and Charles R. Plosser. 1982. "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications." *Journal of Monetary Economics* 10(2): 139–62.

Onatski, Alexei. 2009. "Testing Hypotheses about the Number of Factors in Large Factor Models." *Econometrica* 77(5): 1447–79.

Page, John M. 2012. *Aid, Structural Change and the Private Sector in Africa*. UNU-WIDER Working Paper No. 2012/21, Helsinki, United Nations University.

Pedroni, Peter. 2007. "Social Capital, Barriers to Production and Capital Shares: Implications for the Importance of Parameter Heterogeneity from a Nonstationary Panel Approach." *Journal of Applied Econometrics* 22(2): 429–451.

Pesaran, M. Hashem. 2004. *General diagnostic tests for cross section dependence in panels*. IZA Discussion Paper No. 1240, Bonn, Institute for the Study of Labor.

———. 2006. "Estimation and Inference in Large Heterogeneous Panels With a Multifactor Error Structure." *Econometrica* 74(4): 967–1012.

———. 2007. "A Simple Panel Unit Root Test in the Presence of Cross-Section Dependence." *Journal of Applied Econometrics* 22(2): 265–312.

Pesaran, M. Hashem, Yongcheol Shin, and Ron P. Smith. 1999. "Pooled Mean Group Estimation of Dynamic Heterogeneous Panels." *Journal of the American Statistical Association* 94(446): 621–34.

Pesaran, M. Hashem, and Ron P. Smith. 1995. "Estimating Long-Run Relationships From Dynamic Heterogeneous Panels." *Journal of Econometrics* 68(1): 79–113.

Pesaran, M. Hashem, and Elisa Tosetti. 2011. "Large Panels with Common Factors and Spatial Correlations." *Journal of Econometrics* 161(2): 182–202.

Ranis, Gustav, and John Fei. 1961. "A Theory of Economic Development." *American Economic Review* 51(4): 533–556.

Robinson, Sherman. 1971. "Sources of Growth in Less Developed Countries: A Cross-Section Study." *The Quarterly Journal of Economics* 85(3): 391–408.

Solow, Robert M. 1956. "A Contribution to the Theory of Economic Growth." *The Quarterly Journal of Economics* 70(1): 65–94.

Stock, James H., and Mark W. Watson. 2002. "Macroeconomic Forecasting Using Diffusion Indexes." *Journal of Business & Economic Statistics* 20(2): 147–62.

Stoker, Thomas M. 1993. "Empirical Approaches to the Problem of Aggregation over Individuals." *Journal of Economic Literature* 31(4): 1827–74.

Swan, Trevor W. 1956. "Economic Growth and Capital Accumulation." *Economic Record* 32(2): 334–61.

Temple, Jonathan R. 2005. "Dual Economy Models: A Primer for Growth Economists." *The Manchester School* 73(4): 435–78.

Temple, Jonathan R., and Ludger Wößmann. 2006. "Dualism and Cross-Country Growth Regressions." *Journal of Economic Growth* 11(3): 187–228.

UNIDO. 2004. *UNIDO INDSTAT Industrial Statistics Database*. Vienna: UNIDO, United Nations Industrial Development Organisation. http://www.unido.org/index.php?id=1000327.

Vollrath, Dietrich. 2009a. "The Dual Economy in Long-Run Development." *Journal of Economic Growth* 14(4): 287–312.

———. 2009b. "How Important are Dual Economy Effects for Aggregate Productivity?" *Journal of Development Economics* 88(2): 325–334.

Westerlund, Joakim, and Jean-Pierre Urbain. 2011. *Cross sectional averages or principal components?* Maastricht: METEOR, Maastricht Research School of Economics of Technology and Organization, Working Paper No. 53.

World Bank. 2008. *World Development Indicators*. Online Database, The World Bank, Washington, DC. http://data.worldbank.org/data-catalog/world-development-indicators.

Young, Alwyn. 1995. "The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience." *The Quarterly Journal of Economics* 110(3): 641–80.

Evaluating Program Impacts on Mature Self-help Groups in India

Klaus Deininger and Yanyan Liu

Despite the popularity and the unique nature of women's self-help groups in India, evidence on the economic impact of these groups is scant. On the basis of two rounds of surveys of 2,517 households, we use a strategy of double differences and propensity score matching to assess the economic effects of a program that promoted and strengthened self-help groups in Andhra Pradesh in India. Our analysis finds that longer exposure to the program has a positive impact on consumption, nutritional intake, and asset accumulation. Our investigation into the heterogeneity of these effects suggests that even the poorest households are able to benefit from the program. JEL codes: I38, O12.

India has long made efforts to expand credit availability to rural areas. Early programs, which often yielded less than satisfactory results, were gradually replaced by efforts focusing on the most disadvantaged women and organizing these women into self-help groups (SHGs) as a way to empower them socially and economically while facilitating the eventual establishment of access to bank credit. The use of SHGs for this purpose has recently undergone tremendous growth, and SHGs have emerged as one of the world's largest microfinance networks. In 2007, some 40 million households were organized into more than 2.8 million SHGs, which borrowed more than US$ 1 billion of credit from banks in 2006–2007 (Reserve Bank of India 2008). India's cumulative credit to SHGs is estimated at US$ 4.5 billion, representing approximately 10 percent of all rural credit (Garikipati 2008).

Klaus Deininger is lead economist in the Development Research Group at the World Bank; his email is kdeininger@worldbank.org. Yanyan Liu (corresponding author) is a research fellow at the International Food Policy Research Institute; her email is y.liu@cgiar.org. We thank the staff of the Center for Economics and Social Studies, particularly Prof. S. Galab, Prof. M. Dev, and Dr. P. Reddy, for their support and collaboration in making the data available, and we thank Daniel Ali, Renate Kloeppinger-Todd, Ajai Nair, Ren Mu, Tore Olsen, Yasuharu Shimamura, and three anonymous referees for insightful comments and suggestions. Funding from the Norwegian ESSD Trust Fund (poverty window), the Gender Action Plan, and the Knowledge for Change Trust Fund is acknowledged. The views expressed in this paper are those of the authors and do not necessarily represent those of the World Bank, its Board of Executive Directors, or the countries they represent. A supplementary appendix to this article is available at http://wber.oxfordjournals.org/.

THE WORLD BANK ECONOMIC REVIEW, VOL. 27, NO. 2, pp. 272–296 doi:10.1093/wber/lhs028
Advance Access Publication October 31, 2012

With features such as "saver graduation" and an inherent tendency toward membership expansion (Ahlin and Jiang 2008), India's SHGs have many desirable elements. Even within India, however, their outreach is concentrated in few states; 70 to 80 percent of SHGs are in the four southern states of Andhra Pradesh (AP), Karnataka, Tamil Nadu, and Kerala. Consequently, there is immense potential for expansion. Exploring the impact of the SHG model for microcredit on key household-level outcomes is useful to determine whether significant differences between SHG members and nonmembers (Dev, Kanbur, Galab and Alivelu 2012) can be interpreted as program effects. The results of this examination also have clear implications for the planned expansion of the SHG model under India's National Rural Livelihoods Mission (Government of India 2011).

To provide empirical evidence, this paper evaluates an SHG-based microcredit intervention in Andhra Pradesh (AP) that includes efforts to (i) foster the formation of SHGs among the poor who had previously been excluded; (ii) establish second-tier institutions at higher administrative levels, both to address widespread market imperfections in India and to diversify credit risk, build capacity, and leverage local government efforts;[1] and (iii) provide a one-time injection of equity for the seed capital needed to jump start SHGs formed by the poor. Our primary evidence comes from a two-round survey of 2,517 households in early 2004 and late 2006 that covers "treatment" mandals, where the program began in early 2001, and a set of randomly chosen control mandals, where the program became available in late 2003.[2] To derive credible estimates of impact, we must address the challenges posed by the rapid expansion of the control mandals well before the time envisaged,[3] the endogeneity of households' decisions to participate, the lack of a proper baseline, and a small sample size of randomization units (mandals). We address participants' self-selection into the program by comparing outcomes between program SHG participants in treated and control areas. To eliminate any remaining systematic differences between the treatment and the control groups due to a flawed randomization design, we use double differences and propensity score matching to eliminate any time-invariant bias while addressing bias due to time-varying observables. The unavailability of a preprogram baseline and the contamination of the control group imply that what we estimate is only the impact of

1. The SHG federation is also featured in earlier programs promoted by NGOs, such as PRADAN (Baland, Somanatha, and Vandewalle 2008).

2. Mandals are administrative units above the village and below the district and are equivalent to counties in the US. In most other states, this unit is referred to as a block.

3. In Andhra Pradesh, a first phase targeted only six districts, but a follow-up program to expand coverage to the entire state was implemented less than three years after the launch of this intervention. This period may be too short to expect large economic impacts because the target group includes the poorest households, who require considerable training and capacity building before they are in a position to successfully use and repay loans.

exposure to an additional 2.5 years of the program on (mature) groups that are three years old.

Our results point to economic gains from the program through better nutrition, higher levels of consumption, and asset accumulation by the program's SHG participants. Differentiating by participants' poverty status suggests that the effects are most pronounced for poor participants, who were able to increase their levels of consumption, nutritional intake, and asset accumulation as a result of the program. Most of these effects accrue after more than one year in the program. Furthermore, there is some evidence that program SHGs perform better than nonprogram SHGs with respect to nutritional gains, presumably because of their ability to draw on a federated network to provide access to food grains in kind. Robustness checks and evidence from the literature suggest that this is a lower bound of total program effects. Tests for the heterogeneity of the effect by initial poverty status for SHG members, length of exposure, and the type of SHG (program versus nonprogram SHGs)[4] support our main results. Additionally, our results allow us to obtain a bound for the program's overall cost-benefit ratio.

The paper is structured as follows. Section 2 describes key program features, highlights the challenges faced in the evaluation, and presents the identification strategy for both the intention to treat estimate and the average treatment effect (ATE). Section 3 presents the data characteristics and sample design as well as descriptive evidence at the group and household levels. Section 4 presents ATEs for consumption, calorie, and protein intake; asset accumulation for households in program mandals; and the ATE on members of program SHGs, together with robustness checks and efforts to discern effects on different subgroups. Section 5 concludes, drawing implications for policy and future research.

I. PROGRAM DESCRIPTION AND IDENTIFICATION STRATEGY

An innovative feature of this program is that it incorporates broad initiatives to cater to the specific needs of marginal groups in an effort to reach the poorer segments of the population, who may have been beyond the reach of pure microfinance initiatives because of a lack of opportunities. The goal is to empower the target groups socially and economically, thereby allowing them to become subjects of credit. Qualitative and descriptive accounts document the success of this approach in covering large shares of the rural population, but evidence on the extent to which target groups are reached by the program and the magnitude of the benefits they derive from it, especially in comparison

4. Nonprogram SHGs refer to SHGs formed by efforts prior to the program under study. We provide more information on earlier efforts to form SHGs in section 2.

with the program cost, remains scant. This section describes the program design, notes the challenges that it poses, discusses other evidence, and presents our identification strategy.

Program Design

The formation of women's SHGs as a means to foster female empowerment, raise awareness, and facilitate access to independent savings and financial resources was introduced in the 1980s as part of a pilot scheme for the Development of Women and Children in Rural Areas to improve the gender component of India's Integrated Rural Development Project. A typical SHG comprises 10 to 20 women who meet regularly to collect members' savings (which are deposited in a joint bank account), discuss social issues, attempt to identify critical issues or skill gaps, and improve members' skills through specific training. Once savings have been accumulated, members can apply for internal loans, drawing on accumulated savings at an interest rate set by the group. Once the group has established a record of saving and repayment, it can gain access to commercial bank loans, generally in fixed proportion (commonly four to one) to its equity capital.

Efforts to promote SHGs in the state of Andhra Pradesh (AP) have been among the most proactive and successful in advancing this concept. To build on this success, the state government, with support from a US$ 111 million World Bank loan, implemented the District Poverty Initiatives Project (DPIP) in the state's six poorest districts (Chittoor, Srikakulam, Adilabad, Vizianagaram, Mahabubnagar, and Anantapur). The goal was to expand coverage through the formation of new SHGs and to enhance the capacity of existing SHGs. To accomplish this goal, a three-pronged strategy was adopted.

First, efforts were undertaken to induce the formation of new SHGs among poor segments of the population who, in light of their limited attractiveness as subjects of credit, had failed to join earlier efforts. The tool to identify the target group was a statewide "participatory identification of the poor" that added vulnerability and social exclusion to quantitative indicators from the 2001 national census. Its main output was a set of lists, duly ratified by village assemblies, to determine the poverty status of all households in a village.[5] The program supported community organizers who explained the benefits of program participation to the poor and identified strategies that would make participation attractive to them.

Second, existing and newly formed SHGs were strengthened by creating and supporting a federated SHG structure at the village, mandal, and district (and,

5. Four categories were defined as follows. First, the *poorest of the poor* have food only when they obtain work, lack shelter, proper clothing, and respect and often cannot send children to school. The *poor* have no land, live on daily wages, and may need to send school-age children to work in times of crisis. The *not so poor* have some land and proper shelter, send children to public schools, and have access to bank credit and public services. The *nonpoor* own at least five acres of land, can hire laborers, send children to private schools, use private hospitals, lend money, and have high status.

eventually, state) levels through so-called "Village Organizations" and "mandal (later, "zilla") samakhyas." Village organizations can include 20 or more SHGs per village and are governed by an executive committee with two representatives from each member SHG. A similar pattern holds at the mandal and district levels. In addition to performing traditional functions of microfinance, such as obtaining loans from banks to on-lend resources to members, SHG federations assist with the implementation of government programs and aim to link membership to local government, possibly by forming specific committees. Other program interventions that may be implemented by group federations at the village or mandal level include agricultural marketing activities, insurance coverage, old age or disability benefits, and employment programs and job training.

An example of how such efforts can improve poverty targeting and the transparency of regular programs is the "rice credit line." Under this scheme, mandal- and village-level SHG federations essentially take over the public distribution system by acquiring subsidized rice in bulk and making it available to their SHG members as an in-kind credit and in a more regular fashion than traditional delivery channels. The savings from bulk purchasing or better control of the supply chain can be passed on to members through lower prices. Anecdotal accounts suggest that this scheme allowed federations to circumvent some well-documented problems of the public distribution system (Kochar 2005). In settings where such programs had remained out of reach, the prospect of gaining access to reliable rice supplies helped to attract new group members and establish discipline in terms of attendance at meetings, saving, and repayment.

Because weak capacity has been a key reason for the failure of previous credit and savings programs, these elements were complemented by an emphasis on capacity building. In each program community, community assistants were hired to provide technical support to entrepreneurial activities and increase credit demand. Trained "master bookkeepers" were deployed to educate villagers and to periodically assess SHG accounts to ensure proper management. A "community investment fund" was made available to jump-start lending, even among SHGs with low levels of accumulated member savings, by building sufficient group equity to provide collateral to borrow from banks and, at least in principle, to provide groups with a reliable source of income from interest payments.[6] The resources earned in this way were expected to provide revenue that could be reinvested in economic activities, such as marketing and processing.

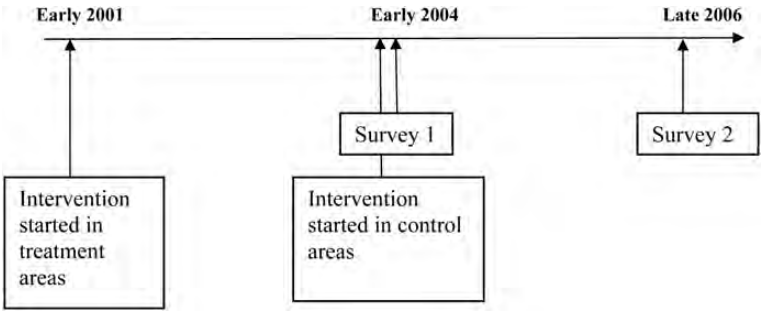
6. Community investment fund resources were initially made available to SHGs but later were offered to SHG federations at the village and mandal levels.

Previous Evidence and Evaluation Challenges

A large amount of anecdotal evidence points to the effectiveness of SHGs in organizing poor women (Center for Economic and Social Studies 2003), providing access to financial services (Basu and Srivastava 2005, Shah, Rao, and Shankar 2007, Sinha 2006), and spreading innovative practices (Nair 2005). However, even in India, the reach of SHGs is highly variable across states, and there is a strong desire on the part of the government to replicate the model at a national scale. Evidence on the economic effects of SHGs is more limited. A qualitative survey in one of the districts where SHGs were first established finds that these groups helped to reduce vulnerability to drought, encouraged entrepreneurial behavior and the diversification of livelihoods, and improved social capital (Garikipati 2008). In Andhra Pradesh, after three years of program exposure, SHG participants had better nutritional and empowerment outcomes but did not perform better economically or accumulate more assets (Deininger and Liu 2009b). Positive effects of empowerment on SHG participants are also found in a larger five-state sample (Swain and Wallentin 2008). However, robust evidence on economic impacts is more difficult to find. In fact, one study concludes that although the project had a positive impact on coping with risk, empowerment, and nonfood expenditure, lasting economic impacts are unlikely (Lastarria-Cornhiel and Shimamura 2008).

To provide more detail on the effects of SHG exposure and participation, it is important to address issues of households' self-selection into program SHGs, which is based on unobservable characteristics as well as data constraints. We use a panel survey of approximately 2,500 households in 41 mandals (10 controls and 31 treatments) from 2004 and 2006, based on a randomized design, to evaluate the project. Although 10 mandals were intended to be kept as controls, the project's perceived success meant that controls were covered much sooner than anticipated. Additionally, because all of the staff's energy was devoted to implementation, no comprehensive preintervention baseline was conducted. Figure 1 shows that the surveys provide information for 2001

FIGURE 1. Timeline of Intervention and Survey



(retrospectively), when the program was initiated in treatment areas; for 2003, when it had just begun in control areas; and for 2006, when the treatment and control areas had been exposed to the program for 5 and 2.5 years, respectively.

Identification Strategy

We are interested in two types of impact estimates: (i) the mean program impact on all households in the treatment areas, equivalent to the intention to treat and (ii) the ATE on members of program SHGs. To address the problem of a contaminated randomization design, we combine difference-in-difference (DD) estimates with propensity score (PS) matching throughout, as explained in greater detail below. We also explore the possible heterogeneity of impacts on members of program SHGs on the basis of their initial poverty status and the length of their participation. Furthermore, to the extent that sample sizes permit, we explore differences between program and nonprogram SHGs.

To evaluate the mean impact of the program, we use information from households located in treated areas compared with control areas. This selection is irrespective of households' participation in SHGs; rather, it is based on the program's availability to treatment areas in early 2001 and to control areas in late 2003. This method allows us to identify the "treatment effect" as the difference in impact between 5 years and 2.5 years of program exposure. Because households self-selected into program SHGs, credibly estimating the ATE on participants in program SHGs requires adjusting for differences in observable and unobservable characteristics. Instead of using nonparticipants in treatment mandals, we compare participants in the control mandals to participants in the treatment mandals. A combination of DD and PS matching allows us to circumvent the self-selection problem by drawing on the assumption that, conditional on observables, unobservables that affect self-selection into program SHGs and subsequent changes in outcomes are identical between participants in treatment and control areas. Because the control and treatment areas are subject to the same program (a condition that is satisfied here), this assumption is reasonable.

To assess the possible interaction of program impact with initial wealth, we use households' preprogram poverty rankings as a classification variable. On the basis of 2006 participation status, members of program SHGs from the corresponding wealth group in control mandals can serve as a control for those in treatment mandals, allowing us to estimate impacts on SHG member households by initial poverty category. Because control mandals entered the program in 2003, we estimate the impact of additional exposure to the program for approximately 2.5 years. Note that 2.5 to 3 years is close to the time required for groups to attain a level of maturity (Galab, Reddy, and Sreemaraju 2012). Our estimated impact can thus be interpreted as the impact of 2.5 years of program exposure for well-developed groups. Unless SHG members incur losses in the

first 2.5 years of their existence (e.g., because of high levels of investment in group capacity), our estimate will be a lower bound of true program impacts.[7]

To formally illustrate our approach, which combines DD and PS matching, let $D = 1$ if a household is in a treated mandal and $D = 0$ if it is in a control mandal. Let the outcome of being treated by the program and the counterfactual outcome at time be denoted by (Y_t^T, Y_t^C). The gain from treatment is $(Y_t^T - Y_t^C)$, and we are interested in the average effect of treatment on the treated (ATT), $E(Y_t^T - Y_t^C|D = 1)$. The inability to observe the counterfactual outcome for treated households prevents us from directly estimating the ATT. Because we observe outcomes from 2003/04 and 2005/06, we use DD to control for household fixed effects. With $t = 1$ denoting 2005/06 and $t = 0$ denoting 2003/04, we can write the standard DD estimator as

$$\begin{aligned} DD &= E(Y_1^T - Y_0^T|D = 1) - E(Y_1^T - Y_0^C|D = 0) \\ &= E(Y_1^T - Y_1^C|D = 1) + B_1 - B_0 - M_1 - M_0 \end{aligned}$$

where B_t is the selection bias in period t and $B_t = E(Y_t^C|D = 1) - E(Y_t^C|D = 0)$. $M_1 = E(Y_1^T - Y_1^C|D = 0)$, and $M_0 = E(Y_0^T - Y_0^C|D = 1)$. M_1 and M_0 are the program impact of the first 2.5 years of exposure in control and treatment areas, respectively. As discussed earlier, the need to include them arises from the lack of a proper baseline survey and the contamination of the control group. The signs of M_1 and M_0 are important; if they are negative (positive), DD will overestimate (underestimate) the program impacts. Although their signs cannot be determined a priori, we expect M_1 and M_0 to be nonnegative on the basis of economic reasoning (it would be difficult to expect poor people, who are the program's primary target group, to make large up-front investments) and evidence in the literature suggesting the positive impact of short-term (three-year) exposure to the program (Deininger and Liu 2009b). If M_1 and M_0 are positive and selection bias is constant over time (i.e., $B_1 = B_0$), the DD estimator will yield a lower bound of program impact, which can be interpreted as the impact of exposure to the program for 5 years compared with 2.5 years' exposure, conditional on $B_1 = B_0$.

If the initial household characteristics that affect subsequent changes to the outcome variables are distributed differently between the treatment and the control groups, the condition $B_1 = B_0$ or $E(Y_1^C - Y_0^C|D = 1) = E(Y_1^C - Y_0^C|D = 0)$ will not hold. To allow for this situation, we use PS matching to balance these variables.[8] The assumption underlying PS matching is that, conditional on observables, changes in outcome variables, if untreated, are independent of actual treatment, $[(Y_1^C - Y_0^C)\perp D|X]$. This assumption has been shown to imply

7. As pointed out by a reviewer, if SHGs initially undertake significant investments, the initial program impact can be negative. In contrast, if 2.5 years is long enough for the members to derive some initial benefits, both will be positive.

8. Clustering at the village level was used throughout to control for village-level random effects.

$(Y_1^C - Y_0^C) \perp D|P(X)$, where is the propensity score, defined as $P(X) = Prob(D = 1|X)$ (Rosenbaum and Rubin 1983). In the empirical implementation, we use a PS-weighted regression method (Hirano, Imbens, and Ridder 2003), which produces an estimate of the ATT as the parameter in a weighted least-square regression of the form

$$Y_{it} - Y_{i,t-1} = \alpha + \beta D_i + \varepsilon_i, \tag{1}$$

where i indexes households, and the weights equal one for treated observations and $\hat{P}(X)/(1 - \hat{P}(X))$ for nontreated observations. See Chen, Mu, and Ravallion (2009) for empirical application of this method.

To obtain consistent and efficient estimates, we determine the common support region with

$$A_{10} = \{X|\hat{P}(X) \leq \lambda\} \tag{2}$$

where if

$$\sup_X \frac{1}{1 - \hat{P}(X)} \leq 2E\left[\frac{1}{1 - \hat{P}(X)}|D = 1\right] \tag{3}$$

and is a solution to

$$\frac{1}{1 - \lambda} = 2E\left[\frac{1}{1 - \hat{P}(X)}|D = 1, \hat{P}(X) \leq \lambda\right] \tag{4}$$

otherwise. It has been shown that under homoscedasticity, this trimming method minimizes the variance of the estimated ATT (Crump, Hotz, Imbens, and Mitnik 2007). Our results are consistently based on trimmed PS-weighted DD, but we also report the results for the untrimmed simple DD for comparison. Exploring effects for different subsamples (e.g., by the length of program exposure) allows us to make inferences about the channels for effects to materialize, their heterogeneity, and the robustness of our results.

II. DATA AND DESCRIPTIVE EVIDENCE

This section provides a more detailed description of the sample and the content and implementation of questionnaires beyond the standard for large household surveys of this nature. Descriptive statistics from a separate SHG survey illustrate the nature of the program and the extent to which it achieved its goals of expanding group coverage and providing SHGs with access to credit and skills. The group-level data provide initial evidence of the extent to which this

program may have led to broader social and economic impacts. The data also aid in the formulation of hypotheses, which can be tested using more rigorous methods.

Sampling Framework

Our data are taken from two rounds of surveys, conducted at both SHG and household levels in early 2004 and late 2006. The sampling framework was designed at the beginning of the program by selecting three of the six program districts to represent the state's macroregions (Telangana, Coastal Andhra, and Rayalseema). Within these, a total of 41 mandals were selected, of which 10 were to serve as controls, where the program would not be available.[9] Households in each of the sample villages were to be selected randomly, with stratification aimed at oversampling the poorer groups targeted by the program.[10] The questionnaire consisted of male and female sections, intended to be answered separately by the main male and female members of the household.[11] In addition to the household sample, up to six randomly selected SHGs per village were selected to provide information on group-level activity through an SHG-level questionnaire.[12] The original sample comprised 2,639 households in 256 villages, 2,517 of which were also covered in the 2006 follow up. We illustrate the 2004 sample composition by district and by mandal treatment status in table S1 in the supplemental appendix (available at http://wber.oxfordjournals.org/).[13] In table S2, we summarize the composition of the household sample for the treatment and control areas in 2004 (columns 1–3) and 2006 (columns 4–6) and those who dropped out (columns 7–9) by type of SHG participation. The contamination of the control is evident from the fact that in 2004, approximately 36 percent of control mandals (compared with 50 percent of treated ones) participated in program SHGs. With little change in the number of nonparticipants among the treatment and control groups between 2004 and 2006, much of the subsequent expansion appears to have involved the conversion of nonprogram SHGs into program SHGs. At 5 and 4 percent, respectively, attrition rates were similar in treated and control areas.

9. As noted earlier, this procedure was not followed; in late 2003, the program was made available to control mandals as well.

10. The goal was to choose four households from the poorest of the poor, three from the poor, two from the not so poor, and one from the nonpoor.

11. For example, information on health, consumption, and female empowerment, among others, was obtained from female respondents, whereas information on agricultural production was obtained from male respondents.

12. Unfortunately, SHGs were sampled independently from households, making it impossible to link the two groups. The original and follow-up SHG surveys covered 1,473 and 1,298 SHGs, respectively (the latter including 72 SHGs formed between 2003 and 2006). A total of 510 and 953 of the SHGs in the 2004 survey were program and nonprogram SHGs, respectively (see table S1).

13. A total of 175 SHGs and 122 households in the original sample could not be found in the 2006 survey.

A Descriptive Account of SHG-level Activities

To illustrate the nature of the program activities and the extent of the implementation, in table 1, we report group-level variables for SHGs in the control and treatment areas in 2001 (which are based on retrospective information from the 2004 survey), 2003, and 2006 for all SHGs, irrespective of their type (i.e., program or nonprogram SHGs). We note clear improvements in the level of group activities and adherence to rules over time for SHGs, with some lag found in the control areas. In the treatment areas, the share of groups that met at least monthly rose from 48 percent in 2001 to 72 percent in 2003, where it remained relatively stable. By comparison, there is little change in meeting frequency for the control areas before 2003 (from 0.44 to 0.46), followed by a marked increase to 80 percent. Virtually all SHGs indicated that members contributed savings in meetings throughout the 2001 to 2006 period. The data also suggest that insurance, nutrition, marketing, and training actions by SHG federations at the village or mandal level were more pronounced in treatment mandals. Interventions to reduce vulnerability (i.e., in-kind credit, insurance, and disability programs), to provide a rice credit line, and to facilitate access to markets were implemented by 49, 40, and 10 percent of the SHGs in treated mandals, with increases to 71, 55, and, 23 percent, respectively, by 2006. Groups in the control areas lagged regarding the implementation of such activities in both 2003 and 2006, though approximately 18 percent (versus 22 percent in treated areas) had implemented job-training programs for SHG members by 2006.

Data regarding the lending portfolio indicate increased internal lending and access to loans from banks and the project-supported community investment fund. The share of groups in the treatment area practicing internal lending increased from 18 percent in 2001 to 53 percent in 2003 and to 87 percent in 2006, compared with rates of 20, 37, and 90 percent for the three respective years in the control area. Nonetheless, with exchange rates of Rs. 45 per US $ in March 2004 and September 2006, loan sizes remained modest for both control and treatment groups, with a median size between Rs. 4,500 and 10,500. The median duration of internal loans was 12 months; half of these loans were used for consumption smoothing.

In 2001, only 3 percent of groups in the treatment area accessed program funds, a figure that increased to 25 percent in 2003 and to 62 percent in 2006. The SHG survey provides information on the purpose of each loan. At four to ten times the median internal loan size, the loans from project-supported funds were larger and more likely to be used for investment than internal lending. Although groups in the control area did not have access to program funds in 2001 or 2003, they eventually caught up with the treatment areas, with 55 percent gaining access to program funds by 2006.

In line with the program's goal of linking SHGs to commercial banks, access to bank loans increased for groups in both the control and treatment

TABLE 1. Summary of SHG Activities in 2001, 2003, and 2006

	2001			2003			2006		
	Control	Treated	Sig.	Control	Treated	Sig.	Control	Treated	Sig.
SHG functioning									
Meet at least monthly	0.44	0.48		0.46	0.72	***	0.80	0.74	**
Members make savings in meetings	0.97	0.88	***	1.00	1.00	*	0.92	0.88	**
Non-microcredit activities									
Activities to reduce vulnerability	0.13	0.07	***	0.24	0.49	***	0.48	0.71	***
Access to rice credit line	0.00	0.00		0.05	0.40	***	0.23	0.55	***
Marketing activities undertaken	0.02	0.03		0.05	0.10	***	0.14	0.23	***
Employment program/job training	–	–	–	–	–	–	0.18	0.22	*
Microcredit activities									
Practice internal lending	0.20	0.18		0.37	0.53	***	0.90	0.87	
if yes, median internal loan size	6950	4500	–	6450	5850	–	10005	10500	–
if yes, median internal loan length	–	–	–	12	10	–	12	12	–
share for consumption smoothing	0.46	0.47		0.51	0.48		0.45	0.48	
share for investment	0.38	0.45		0.41	0.44		0.47	0.44	
Have access to program fund	0.00	0.03	***		0.25	***	0.55	0.62	**
if yes, median size	–	36012	–	24600	34000	–	21000	40675	–
if yes, median length (months)	–	–	–	15	20	–	12	20	–
Main purpose: cons. smoothing		0.03	–	–	0.07	–	0.07	0.05	
Main purpose: investment	–	1.00	–	–	0.96	–	0.90	0.90	
Access bank loans	0.35	0.27	***	0.36	0.44	***	0.87	0.89	
if yes, median size	15000	15000		30000	24000		56500	50000	
if yes, median length (months)	–	–	–	12	12	–	20	20	–
Main purpose: cons. smoothing	0.50	0.36	***	0.37	0.39		0.30	0.07	***
Main purpose: investment	0.66	0.76	**	0.73	0.81	*	0.64	0.88	***

Source: Authors' analysis using data from 2004 and 2006 AP DPIP SHG surveys (2001 information collected retrospectively from the 2004 survey).
Significance levels of the difference between the control and treatment found by *t* test are indicated by stars: * 10%; ** 5%; *** 1%.
SHGs include program and nonprogram SHGs, and empty cells indicate unavailable data or are not applicable.

areas from 2003 to 2006. The relatively slow increase in treatment areas from 2001 to 2003 illustrates startup problems in developing the program's implementation structure and suggests that access to bank loans is not immediate. From 2001 to 2006, the median size of bank loans (on-lend internally) increased from Rs. 15,000 to Rs. 50,000, highlighting groups' greater creditworthiness, which can be at least partly attributed to the infusion of program funds.

This summary of SHG activities points to a number of hypotheses regarding program impacts. First, improved credit access, especially the ability to use loans from banks and program funds, should increase income by encouraging investment and asset accumulation. Second, participation in groups and associated access to internal lending may reduce vulnerability, improve nutritional status, and increase income. Third, increased asset endowments and income levels could lead to higher consumption and nutritional intake levels. Fourth, interventions such as the rice credit line or efforts to provide training in dairy production and to encourage the use of credit funds to acquire cows or buffaloes may have resulted in changes in nutritional intake (calories and protein) among SHG participants in program areas by reducing the price of food independently of income effects.

Because access to program funds, bank loans, and the rice credit line were all targeted at SHGs, we expect economic benefits to be limited to SHG members. With the possible exception of nutritional benefits gained from access to a rice credit line, spillovers to non-SHG members in treatment areas are unlikely. Moreover, to the extent that SHGs offer members a menu of options, the realization of these depends, at least partially, on wealth and the nature and magnitude of effects and may be affected by initial endowments, even for SHG participants. For example, a rice credit line and internal lending will be particularly attractive to the poor, whereas marketing activities provide larger benefits to the nonpoor. At the same time, the poor and the nonpoor may differ in their use of loan funds, with the poorest potentially being averse to taking out large loans because of the lack of complementary assets or skills or because of the fear of their inability to repay.

Descriptive Evidence at the Household Level

The household survey contained male and female modules, administered separately to a key male or female person in the household, normally the household head and spouse. In table S3, we describe levels of participation in program SHGs by poverty status in three different periods: (i) 2001, the year the program was implemented in the treatment areas; (ii) 2003, the year the program began in the control areas; and (iii) 2006, the year of the final survey. Membership increased over time, both in treatment and control areas, with a lag in the latter. In 2006, approximately 59 and 47 percent of the treated and control areas, respectively, were members of program SHGs.

Outcome variables available in the survey include the value of consumption, nutritional intake, and levels of assets, which we express in per capita terms based on adult equivalents.[14] Consumption includes food and nonfood consumption over the past 30 days and lumpy items over the past year.[15] We compute the intake of calories and protein by multiplying quantities of more than 30 food items in the questionnaire's consumption section, with their caloric and protein content based on the main reference for Indian foods (Gopalan, Rama Shastri, and Balasubramanian 2004).[16] Nonfinancial assets include consumer durables, productive assets, and livestock assets.[17]

Table S4 illustrates outcome variables for control and treated mandals in the 2003/2004 and 2005/2006 periods. Comparing initial conditions between the households in the control and treated mandals in table S5 highlights some differences, such as much higher levels of scheduled tribes in treatment areas than in control areas (24 versus 6 percent), which could be due to the rather small sample of control mandals. As shown in table S6, compared with those who remained in the sample, households that dropped out were smaller in size, poorer, and more likely to be female-headed, landless, and illiterate. These findings are consistent with the notion that the probability of dropping out is higher among migrants or female-headed households.

In table S7, we present the regression results of early placement of a program for the overall sample (column 1), for the sample with participants of program SHGs only (column 2), and for the sample with nonparticipants only (column 3). In each case, the dependent variable is 1 if the household (in the subgroup of interest) is located in a treated mandal and 0 otherwise. Explanatory variables for demographics (location, caste, female headship, and literacy) and initial economic conditions (poverty status, land ownership, consumption, nutritional intake, and nonfinancial assets) are from the 2004 household survey. The low pseudo R^2 is in line with the random selection of control mandals, although the results suggest that randomization failed to eliminate some systematic differences between treatment and control areas, possibly because of small sample sizes.

14. To obtain adult equivalent measures for caloric and protein consumption, we use nutritional requirements by sex and age as weights (Gopalan, Rama Shastri and Balasubramanian 2004). Additionally, to generate adult equivalent measures of income and total consumption, we weight those above 60 and below 14 by 0.78.

15. Although the survey instrument is less disaggregated than that used by the National Sample Survey, it follows the overall structure used there.

16. For fruits or vegetables where the survey includes only aggregate spending, we use the 55th round of the National Sample Survey to derive the price and caloric content of a representative basket of those consumed in Andhra Pradesh.

17. Financial assets were excluded because of concerns about misreporting. Asset values were measured as of December 2003 in the 2004 survey and as of June 2006 in the 2006 survey.

III. ESTIMATION RESULTS

The methodological framework provides estimates of the program's overall effects and its specific effects for different groups. Positive and significant effects on SHG members' expenditure, nutritional intake, and asset accumulation appear to be driven mainly by the poorer participants. Longer exposure to the program yields a higher program impact, as expected. We do not find evidence of spillovers, either from participants in program SHGs to nonparticipants or from program SHGs to nonprogram SHGs, in the treatment areas.

Intention-to-Treat Estimate

To estimate the intention-to-treat estimator, we compare the households in the treatment mandals with those in the control mandals, irrespective of their SHG membership status. Because the above evidence points to significant differences in observables between treated and nontreated areas, we use estimated propensity scores (from the first-step logit regression, as reported in the first column of table S7) to balance variables that may influence outcomes. The first column in table S9 displays differences in the means of the matching variables between treatment and control areas for the PS-weighted and trimmed samples, illustrating that trimming and matching based on the estimated PS balanced all of the variables of interest.

Estimates for the ATEs on households in the treated mandals based on the trimmed sample are reported in the bottom panel of table 2. Simple DDs based on the total sample are shown in the top panel for comparison.[18] We find a significant impact of the intention to treat on investment in nonfinancial assets, which is estimated to be higher by Rs. 453, or 16 percent, compared with the counterfactual. This result suggests that once groups achieved a level of maturity, the objective of inducing higher levels of investment and capital formation was achieved. However, we fail to find a significant impact on total consumption or nutritional intake after trimming and reweighting. Because the estimated impact refers to the average of SHG members and nonmembers in treatment mandals, this result suggests the possibility of a program effect on members that cannot be detected in this setting. To explore this possibility, we turn to the impact on program SHG members alone.

Average Treatment Effects on Program SHG Members

To estimate program impacts on members of program SHGs, the treatment group is defined as households in the treatment areas that joined program SHGs before 2003. By comparison, the control group comprises households in

18. In tables 2 to 4, the number of observations eliminated through trimming can be obtained by comparing the number of observations between the total and trimmed samples in the relevant tables. For example, in table 2, for consumption per capita, trimming drops 181 (=1819-1638) observations in the treated group and 9 (=489-480) observations in the control group. Figure S1 in the appendix graphically illustrates the impact of trimming.

TABLE 2. Estimated Average Treatment Effects on Households Living in Treatment Areas

	Untrimmed sample, simple DD				
	Treated	Control	DD	(s.e.)	Sig.
Consumption p.c. (Rs/year)	1364	892	472	(215)	**
Food (Rs/year)	478	477	1	(109)	
Nonfood (Rs/year)	886	415	471	(155)	***
No. of obs.	1819	489			
Energy intake p.c. (Kcal/day)	156	213	−57	(60)	
Protein intake p.c. (g/day)	−0.90	0.37	−1.27	(1.30)	
No. of obs.	1917	518			
Nonfinancial assets p.c. (Rs)	750	411	338	(163)	**
No. of obs.	1926	519			
	Trimmed sample, PS weighted DD				
	Treated	Control	DD	(s.e.)	Sig.
Consumption p.c. (Rs/year)	1436	1103	333	(212)	
Food (Rs/year)	547	497	50	(118)	
Nonfood (Rs/year)	888	606	282	(152)	*
No. of obs.	1638	480			
Energy intake p.c. (Kcal/day)	214	179	36	(68)	
Protein intake p.c. (g/day)	1.05	−0.83	1.89	(1.52)	
No. of obs.	1751	507			
Nonfinancial assets p.c. (Rs)	754	301	453	(190)	**
No. of obs.	1758	508			

Source: Authors' analysis using data from 2004 and 2006 AP DPIP household surveys.

Note: "Treated" and "control" denote the change from 2004 to 2006 in an outcome for the treated group and the control group, respectively. DD denotes the double difference, and (s.e.) denotes standard errors. Significance of coefficients is shown as follows: * at 10%; ** at 5%; *** at 1%.

control areas that were registered as program SHG members in 2006. We only keep members who joined before 2003 for the treatment mandals and eliminate 342 "late joiners" who joined a program SHG at a later date in the treatment areas. We also exclude 107 households who left SHGs after 2004, noting that these departures were the result of SHGs being dissolved or becoming dysfunctional rather than individuals joining different SHGs.[19] Therefore, our treatment group includes households that were exposed to the program for between 3.5 and 6 years, whereas the control group includes households with program exposure of less than 3 years.[20]

19. Project staff indicate that the main reason for individuals to exit groups is either expulsion, in most cases because of conflict or failure to honor repayment commitments, or the dissolution of the entire group, which is often due to conflict. Incentives for functioning groups to welcome such individuals are limited, implying that mobility across groups is virtually nonexistent.

20. We exclude 107 households that had been SHG members in 2004 but that were no longer members in 2006 because their SHG had become dysfunctional.

TABLE 3. Impact on Program SHG Participants and Nonparticipants

	Untrimmed sample, simple DD					
	Participants			Nonparticipants		
Consumption p.c. (Rs/year)	768	(272)	***	80	(338)	
Food (Rs/year)	234	(135)	*	−210	(167)	
Nonfood (Rs/year)	535	(212)	**	290	(247)	
No. of obs.	510 + 239 = 749			917 + 221 = 1138		
Energy intake p.c. (Kcal/day)	85	(75)		−190	(88)	**
Protein intake p.c. (g/day)	1.96	(1.62)		−3.92	(1.84)	**
No. of obs.	535 + 246 = 781			970 + 241 = 1211		
Nonfinancial assets p.c. (Rs)	549	(243)	**	203	(243)	
No. of obs.	539 + 243 = 782			977 + 245 = 1222		
	Trimmed sample, PS weighted DD					
	Participants			Nonparticipants		
Consumption p.c. (Rs/year)	552	(309)	*	−122	(341)	
Food (Rs/year)	307	(168)	*	−257	(176)	
Nonfood (Rs/year)	245	(223)		135	(238)	
No. of obs.	414 + 228 = 642			839 + 218 = 1057		
Energy intake p.c. (Kcal/day)	161	(87)	*	−119	(92)	
Protein intake p.c. (g/day)	4.47	(1.93)	**	−1.75	(2.00)	
No. of obs.	456 + 237 = 693			887 + 235 = 1122		
Nonfinancial assets p.c. (Rs)	755	(276)	***	121	(262)	
No. of obs.	438 + 234 = 672			892 + 241 = 1133		

Source: Authors' analysis using data from 2004 and 2006 AP DPIP household surveys.

Note: Standard errors in parentheses. Significance of coefficients is shown as follows: * at 10%; ** at 5%; *** at 1%.

The column "participants" compares participants with 3.5 to 6 years in program SHGs in treatment villages to program participants with less than three years of exposure in control villages. The "nonparticipant" column compares nonparticipants in treatment villages to those in control villages.

The first column of table 3 reports the ATE on program SHG members with DD and PS matching results based on the trimmed sample in the bottom panel and simple DDs in the top panel for comparison. In contrast to the lack of significance (except for nonfinancial assets) in the intention-to-treat estimation, we find significant impacts on the consumption, nutritional intake, and accumulation of nonfinancial assets by members of program SHGs. The magnitude of these impacts is large; the increment in per capita consumption compared with the counterfactual is estimated to amount to Rs. 552/year (approximately US$ 11), or approximately 7 percent. Estimated increases in the per capita intake of energy and protein as well as investments are equivalent to increases

of 8 (161 calories/day), 10 (4.5 grams/day), and 24 (Rs. 755) percentage points, respectively. These are large effects, especially because a number of factors, such as cross-border spillovers and learning by the agencies responsible for implementation, are likely to bias estimates downward. In view of the finding of nonnegligible gains in nutritional intake, these effects may have occurred during the program's first three years (Deininger and Liu 2009b).

To assess our identification strategy, we run an identical regression comparing nonparticipants in treatment areas to nonparticipants in control mandals as a falsification test. If positive impacts for program SHG members were driven by factors that were not controlled for (for example, if the timing of program placement was based on unobservables that affected outcomes), this test should yield positive impacts for non-SHG members in treatment areas. [21] The results, as reported in column 2 of table 3, suggest that this is not the case; estimated impacts are small in magnitude, and none of the impacts is statistically significant at any conventional level. Because the standard errors are comparable to those obtained for SHG members, a positive impact is unlikely to be due to a lack of precision in the estimates, lending credence to the validity of our identification strategy.

Robustness Checks and Heterogeneity of Program Impacts

Exploring effects for subsamples (e.g., by length of program exposure or poverty status) allows us to draw inferences about the channels in which effects materialize, the robustness of our results, and their heterogeneity.

First, we follow Behrman, Cheng, and Todd (2004) in exploring the effect of varying the length of exposure by comparing program impacts for subsamples of program SHG members who joined at different times. Estimates of exposure to the program for a specific length of time can be derived by comparing pairs of groups using DD and PS matching. The results, summarized in table 4, suggest that longer exposure yields higher impacts for all statistically significant results. Although different sample sizes must be considered when interpreting the results, estimated differences in impacts are greatest between the group that joined in 2001 and the group with the least exposure (i.e., the group that joined in 2004 or thereafter). There are few significant differences for one-year exposure, suggesting that this period is too short to yield clear effects. In contrast, significant differences emerge between the 2001 and 2003 groups and the 2002 and 2004 groups.

Second, it is interesting to examine whether the effects identified here are due to the "new" model of organizing SHGs supported by the program rather than to longer exposure to the traditional SHG model. To determine the source of this effect, we compare whether impacts differ between members of program SHGs and those of nonprogram SHGs ("old" SHGs). Because the program under study differs from "old" programs in its level of organization

21. The test is valid if program benefits accrue to SHG members only.

TABLE 4. Pairwise Difference in Program Impacts Among Program SHG Members Joining at Different Points in Time

	One-year exposure									
	2001 vs. 2002			2002 vs. 2003			2003 vs. 2004			
Cons p.c. (Rs/a)	−44	(338)		537	(251)	**	167	(265)		
Food (Rs/a)	−128	(181)		138	(143)		208	(140)		
Nonfood (Rs/a)	84	(258)		399	(184)	**	−41	(210)		
No. of obs.	187 + 307 = 494			187 + 387 = 574			386 + 433 = 819			307 + 387 = 694
Energy p.c. (Kcal/d)	58	(95)		90	(78)		58	(81)		
Protein p.c. (g/day)	−0.33	(2.18)		2.42	(1.81)		1.36	(1.98)		
No. of obs.	192 + 323 = 515			192 + 402 = 594			399 + 452 = 851			323 + 402 = 725
Nonfinancial assets p.c. (Rs)	−216	(391)		516	(301)	*	37	(254)		
No. of obs.	194 + 324 = 518			193 + 403 = 596			394 + 453 = 847			324 + 403 = 727
	Multiyear exposure									
	2001 vs. 2003			**2002 vs. 2004**			2001 vs. 2004			
Cons p.c. (Rs/a)	579	(354)		547	(273)	**	821	(343)	**	
Food (Rs/a)	64	(183)		308	(140)	**	375	(163)	**	
Nonfood (Rs/a)	516	(257)	**	239	(226)		446	(277)		
No. of obs.	187 + 387 = 574			305 + 430 = 735			187 + 433 = 620			
Energy p.c. (Kcal/d)	168	(105)		109	(81)		240	(101)	**	
Protein p.c. (g/day)	2.93	(2.27)		3.02	(1.87)		4.97	(2.30)	**	
No. of observations	192 + 402 = 594			321 + 453 = 774			191 + 453 = 644			
Nonfinancial assets p.c. (Rs)	333	(330)		364	(332)		262	(370)		
No. of observations	193 + 403 = 596			322 + 455 = 777			192 + 456 = 648			

Source: Authors' analysis using data from 2004 and 2006 AP DPIP household surveys.

Note: Trimmed sample with PS-weighted DD method is used throughout.

Significance of coefficients is shown as follows: * at 10%; ** at 5%.

and outreach to the poor, we expect differences mainly in terms of access to public goods through village- or higher-level federations taking responsibility for the implementation of government programs. Unfortunately, the program's tendency to convert nonprogram SHGs into program SHGs leaves us few participants in "old" groups by 2006. Therefore, we rely on the SHG type in 2004 to define program and nonprogram SHG members and estimate the effects of the "new" model using DD and PS matching. The resulting estimates can be interpreted as the added effect of the "new" SHG model compared with the "old" SHG model. The results, as shown in table 5, point to significant effects on energy and protein intake but not on consumption or asset accumulations. This finding suggests that, in addition to the improved inclusion of the marginal groups, a key improvement of the "new" model is in non-microfinance activities, such as the rice credit line.

Third, we explore spillover effects from program to nonprogram SHGs in treatment areas. Such spillovers could arise from SHGs taking over the implementation of specific government programs (e.g., the public distribution system). We define the treatment group as the members of nonprogram SHGs in the treatment areas and the control group as the members of nonprogram SHGs in the control areas, and we use DD and PS matching. The results of this test are reported in table 6. We fail to find significant spillover effects for any of the outcome variables. However, this finding may be attributed to a small sample size (fewer than 250 observations).

Fourth, as noted earlier, the above estimates of an additional 2.5 years of exposure to the program among mature groups are only part of the total program effect. To the extent that our interest is in the latter, learning more about short-

TABLE 5. Estimated Average Treatment Effects on Program versus Nonprogram SHG Members

	Treated	Control	DD	s.e.	Sig.
Consumption p.c. (Rs/year)	1426	1308	118	389	
Food (Rs/year)	510	401	109	182	
Nonfood (Rs/year)	916	907	9	281	
No. of obs.	531	265	531	265	
Energy intake p.c. (Kcal/day)	251	49	202	94	**
Protein intake p.c. (g/day)	1.46	−2.89	4.35	2.39	*
No. of obs.	549	276	549	276	
Nonfinancial assets p.c. (Rs)	920	725	195	362	
No. of obs.	533	273	533	273	

Source: Authors' analysis using data from 2004 and 2006 AP DPIP household surveys.

Note: Trimmed sample with PS-weighted DD method used throughout.

Figures in the "treated" and "control" columns refer to the difference in outcomes between 2006 and 2004 for program SHG members and nonprogram SHG members based on the 2004 status of SHG type. DD denotes the double difference, and s.e. denotes standard errors.

Significance of coefficients is shown as follows: * at 10%; ** at 5%.

TABLE 6. Estimated Spillover Effects on Nonprogram SHG Members Residing in Program Areas

variable	Treated	Control	DD	s.e.	Sig.
Consumption p.c. (Rs/year)	1813	1154	659	774	
Food (Rs/year)	482	513	−31	362	
Nonfood (Rs/year)	1331	640	691	541	
No. of obs.	169	51	169	51	
Energy intake p.c. (Kcal/day)	133	106	27	147	
Protein intake p.c. (g/day)	−0.13	−1.66	1.53	4.25	
No. of obs.	175	52	175	52	
Nonfinancial assets p.c. (Rs)	1041	1124	−84	589	
No. of obs.	175	49	175	49	

Source: Authors' analysis using data from 2004 and 2006 AP DPIP household surveys.

Note: Trimmed sample with PS-weighted DD method used throughout. Figures in the "treated" and "control" columns refer to changes from 2004 to 2006 in outcomes for program SHG and nonprogram SHG members. DD denotes the double difference, and s.e. is the standard error. Significance of coefficients is shown as follows: * at 10%; ** at 5%; *** at 1%.

term effects (i.e., the impact of exposure to the program for 2.5 years) is of interest. To do so, we use the entire sample to regress outcome variables of interest on the interaction of treatment location and membership in program SHGs (in 2004 for treatment areas and in 2006 for control areas) and other controls.[22] The results (in table S12) indicate nonsignificant effects for all of the outcome variables considered. These results increase our confidence that the main result will not be biased upward.

Finally, because the program explicitly targets the poor, we compare estimated impacts among groups with different initial poverty statuses.[23] The results, reported in table 7, suggest positive, significant, and relatively large impacts in terms of food consumption (Rs. 748 per capita) and nutritional intake (364 kcal/day and 8.93 g/day of protein), which increase by approximately 15 percent for initially poor households. The only significant impact for the poorest groups is an increase in nonfinancial assets. Although the effects on consumption and nutritional intake are positive, they are statistically nonsignificant. No significant impacts are detected for nonpoor members, which may have resulted from limited power due to the small number of observations in the subsample.

22. This procedure is essentially a DD estimation of {(2004 outcomes of program-SHG members in treatment areas − 2004 outcomes of nonmembers in treatment areas) − (2004 outcomes of *future* members in control areas − 2004 outcomes of *future* nonmembers in control areas)}.

23. In table S4, we list means of outcome variables for our subgroups of the poorest of the poor, the poor, and the not-so-poor/nonpoor members separately. Not surprisingly, descriptive statistics suggest that richer households had higher consumption and assets than poorer ones in both periods. We observe the same trend for nutritional intake, although this finding is not as obvious as in the case of consumption and assets.

TABLE 7. Heterogeneous Impacts by Initial Poverty Status

	Untrimmed sample, simple DD								
	Poorest of the Poor			Poor			Not-So-Poor/Not Poor		
Consumption p.c. (Rs/year)	257	(330)		816	(465)	*	1357	(729)	*
Food (Rs/year)	−91	(202)		384	(224)	*	413	(332)	
Nonfood (Rs/year)	348	(247)		432	(368)		945	(531)	*
No. of obs.	228 + 84 = 312			178 + 87 = 265			106 + 68 = 174		
Energy intake p.c. (Kcal/day)	76	(120)		156	(110)		30	(172)	
Protein intake p.c. (g/day)	0.74	(2.78)		3.49	(2.46)		1.85	(3.66)	
No. of obs.	239 + 86 = 325			187 + 89 = 276			111 + 71 = 182		
Nonfinancial assets p.c. (Rs)	439	(251)	*	452	(410)		931	(749)	
No. of obs.	242 + 87 = 329			191 + 88 = 279			109 + 68 = 177		
	Untrimmed sample, simple DD								
	Poorest of the Poor			Poor			Not-So-Poor/Not Poor		
Consumption p.c. (Rs/year)	319	(385)		1394	(865)		876	(909)	
Food (Rs/year)	239	(228)		748	(280)	***	185	(469)	
Nonfood (Rs/year)	79	(281)		646	(689)		691	(599)	
No. of obs.	176 + 81 = 257			144 + 84 = 228			79 + 65 = 144		
Energy intake p.c. (Kcal/day)	156	(135)		364	(121)	***	−28	(199)	
Protein intake p.c. (g/day)	4.12	(3.40)		8.93	(2.96)	***	1.56	(5.23)	
No. of obs.	209 + 84 = 293			156 + 88 = 244			89 + 69 = 158		
Nonfinancial assets p.c. (Rs)	589	(311)	*	644.78	(457)		502	(824)	
No. of obs.	203 + 85 = 288			159 + 87 = 246			86 + 67 = 153		

Source: Authors' analysis using data from 2004 and 2006 AP DPIP household surveys.

Standard errors in parentheses.

Significance of coefficients is shown as follows: * at 10%; ** at 5%; *** at 1%.

A Simple Cost-Benefit Analysis

To assess total program benefits, we note that when our follow-up survey was conducted in 2006, the program had reached 2.29 million households, with an average size of 4.79 adult equivalents. Approximately 47.4 percent of SHG members in our sample joined program SHGs before 2003. Multiplying the estimated ATT on per capita consumption (US$ 11) with the number of participants enables us to establish a bound for project benefits. We do so by assuming two extreme scenarios. If future benefits are assumed to be maintained at current levels, applying a 0.9 discount factor puts the estimated net present value of project benefits at US$ 567.1 million, yielding a benefit-cost ratio of 3.77:1.[24] Under the more conservative assumption that consumption benefits only occurred in the period covered by our survey, the estimated benefits would amount to 56.7 million. In this case, project-supported loans, which have a delinquency rate of 23 percent (Deininger and Liu 2009a), would reduce project costs to 88 million, and the associated benefit-cost ratio would be 0.64:1.[25] The actual benefit-cost ratio will be between these extremes. Given the positive impacts on asset accumulation found by our analysis, the extreme scenario of benefits below project cost may be unlikely. The extent to which groups established under the program are sustained is obviously a key determinant of the program's longer-term impact.

IV. CONCLUSION AND POLICY IMPLICATIONS

This paper was motivated by the notion that, despite considerable interest in expanding SHG approaches to microfinance, rigorous evaluation of the impact of such interventions is scant. Even studies documenting clear social, empowerment, and nutritional impacts have been unable to ascertain economic effects. The household data collected 2.5 to 5 years after the start of the program allow us to assess the impact of longer program exposure, which, under plausible conditions, can proxy for the difference between a group that is fully functional and an immature one.

Using propensity score-weighted double differences on an appropriately trimmed sample and noting that our estimates are likely to constitute a lower bound of true effects, we find that SHG participation had significant economic impacts in the areas considered. If benefits are maintained at current levels, they significantly exceed program costs. The benefits are not confined to those who were already affluent; in fact, there is significant asset accumulation

24. The total project cost is US$ 150.60 million, of which 111 million was provided by a World Bank loan.

25. Of the total project cost, project-supported loans offered through the CIF represented 81.4 million. Therefore, the cost is computed as follows: $(150.6 - 81.4) + 81.4 \times 23\% = 88$. Considering repayment rates is not necessary under the assumption that future benefits will be maintained forever because the CIF fund will remain in the group federations.

among the poorest of the poor, who (partly as a result of participation in SHGs and partly because gains in calorie and protein intake may have been realized earlier in the program) saw their consumption increase less than that of the poor. This finding suggests that if they participate in SHGs, the poorest individuals appear to benefit not only socially but also economically.

Our results suggest that a program that not only fosters group formation but also supports more mature groups through federation and access to credit can produce significant economic benefits in the long term. To assess the overall desirability and impact of such programs, a key question concerns the extent to which the benefits will be maintained once outside support is terminated. The answer to this question will at least partly depend on whether the SHGs established by the program continue to operate (possibly adjusting the services offered to the level of member development) and, related to this, whether beneficiary households will be able to use the one-time injection of credit and capacity to place them on a permanently higher trajectory of economic activity and asset accumulation. Answering this question is beyond the scope of this paper and will require additional information based on both group and individual activity after external support has ceased. Nonetheless, our finding of nonnegligible economic impacts in the case of Andhra Pradesh implies that further investigation of the determinants and implications of the sustainability of SHGs and the benefits that they provide to their membership could be of considerable interest for researchers and policy makers alike.

REFERENCES

Ahlin, C., and N. Jiang. 2008. "Can Micro-credit Bring Development?" *Journal of Development Economics* 86 (1): 1–21.

Baland, J.-M., R. Somanathan, and L. Vandewalle (2008). "Microfinance Lifespans: A Study of Attrition and Exclusion in Self-help Groups in India." *India Policy Forum* 4.

Basu, P., and P. Srivastava. 2005. "Scaling-up Microfinance for India's Rural Poor." *World Bank Policy Research Working Paper 3646*. Washington, DC: World Bank.

Behrman, J. R., Y. Cheng, and P. E. Todd. 2004. "Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach." *Review of Economics and Statistics* 108–32.

Center for Economic and Social Studies. 2003. "Amdhra Pradesh District Initiaties Project: Mid-term report." Hyderabad: Society for the Elimination of Rural Poverty.

Chen, S., R. Mu, and M. Ravallion. 2009. "Are There Lasting Impacts of Aid to Poor Areas? Evidence from Rural China." *Journal of Public Economics* 93: 512–528.

Crump, R., V. Hotz, G. W. Imbens, and O. Mitnik. 2007. "Moving the Goalposts: Addressing Limited Overlap in Estimation of Average Treatment Effects by Changing the Estimand." *NBER Technical Working Paper 330*. Cambridge, MA: National Bureau of Economic Research.

Deininger, K., and Y. Liu. 2009a. "Determinants of Repayment Performance in Indian Micro-Credit Groups." *World Bank Policy Research Working Paper 4885*. Washington DC: World Bank.

———. 2009b. "Economic and Social Impacts Of Self-Help Groups in India." *World Bank Policy Research Working Paper 4884*. Washington, DC: World Bank.

Dev, S. M., R. Kanbur, S. Galab, and G. Alivelu. 2012. "Organization, Poverty and Women: Andhra Pradesh in Global Perspective." In S. Mahendra Dev, Ravi Kanbur, S. Galab, and G. Alivelu, eds., *Organizing Poor Women - The Andhra Pradesh Experience*. New Delhi: Academic Foundation.

Galab, S., P. Reddy, and D. Sreemaraju. 2012. "Women Empowerment and Functioning of Formal and Informal Institutions: The Case of IKP in Andha Pradesh." In S. M. Dev, R. Kanbur, S. Galab, and G. Alivelu, eds., *Organizing Poor Women - The Andhra Pradesh Experience*. New Delhi: Academic Foundation.

Garikipati, S. 2008. "The Impact of Lending to Women on Hosuehold Vulnerability and Women's Empowerment: Evidence from India." *World Development forthcoming*.

Gopalan, G., B. V. Rama Shastri, and B. C. Balasubramanian. 2004. "Nutritive Value of Indian Food." Hyderabad: National Institute of Nutrition, Indian Council of Medical Research.

Government of India. 2011. "National Rural Livelihoods Mission: Mission Document." New Delhi: Ministry of Rural Development.

Hirano, K., G. W. Imbens, and G. Ridder. 2003. "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica* 71 (4): 1161–89.

Kochar, A. 2005. "Can Targeted Food Programs Improve Nutrition? An Empirical Analysis of India's Public Distribution System." *Economic Development and Cultural Change* 54 (1): 203–35.

Lastarria-Cornhiel, S., and Y. Shimamura. 2008. "Social Well-being Outcomes of Rural Microfinance in Andhra Pradesh." *Working Paper*. Madison, WI: University of Wisconsin.

Nair, A. 2005. "Sustainability of Microfinance Self Help Groups in India: Would Federating Help?" *World Bank Policy Research Working Paper 3516*. Washington, DC: World Bank.

Reserve Bank of India. 2008. "Report on Trend and Progress of Banking in India, 2007-08." Mumbai: Reserve Bank of India.

Shah, M., R. Rao, and P. S. V. Shankar. 2007. "Rural Credit in 20th Century India: Overview of History and Perspectives." *Economic and Political Weekly* 42 (15): 1351–64.

Sinha, F. 2006. "Self Help Groups in India: A Study of the Lights and Shades." Noida and Hyderabad: APMAS and EDA Rural Systems.

Swain, R. B., and Y. Wallentin. 2008. "Does Microfinance Empower Women? Evidence from Self Help Groups in India." *Working Paper 2007-24*. Uppsala, Sweden: Department of Economics, Uppsala University.

Wage Effects of High-Skilled Migration: International Evidence

Volker Grossmann and David Stadelmann

The international migration of high-skilled workers may trigger productivity effects at the macro level such that the wage rate of skilled workers increases in host countries and decreases in source countries. We exploit data on international bilateral migration flows and provide evidence consistent with this theoretical hypothesis. We propose various instrumentation strategies to identify the causal effect of skilled migration on log differences of GDP per capita, total factor productivity, and the wages of skilled workers between pairs of source and destination countries. These strategies aim to address the endogeneity problem that arises when international wage differences affect migration decisions. JEL Codes: F22, O30

The recent surge in the international migration of high-skilled workers not only raised standard concerns about adverse brain-drain effects for developing countries but also led to worries about native high-skilled workers in advanced destination countries.[1] Domestic workers with higher education levels fear that their wages will decline in response to increased competition from similarly qualified migrants. Whereas debates on migration in the past have centered on asylum rights and low-skilled migrants, over the years, politicians and the mass media have discovered the issue of high-skilled immigration. For instance, in Switzerland and Austria, the discussion has recently become emotionally charged owing to significant inflows of tertiary-educated workers, particularly from Germany.[2] For the United States, Hanson, Scheve and Slaughter (2009)

Volker Grossman: University of Fribourg; CESifo, Munich; Institute for the Study of Labor (IZA), Bonn. Postal address: University of Fribourg, Bd. de Pérolles 90, G424, 1700 Fribourg, Switzerland. Tel.: +41 (0)26 3009383. Email: volker.grossmann@unifr.ch. David Stadelmann (corresponding author): University of Fribourg; CREMA - Center for Research in Economics, Management and the Arts, Switzerland. Postal address: University of Fribourg, Bd. de Pérolles 90, F408, 1700 Fribourg, Switzerland, +41 (0)26 3008263, david.stadelmann@unifr.ch. A supplementary appendix to this article is available at http://wber.oxfordjournals.com.

1. The number of tertiary-educated immigrants living in OECD countries increased from 12.5 million in 1990 to 20.4 million in 2000 (Docquier and Marfouk, 2006). Half of the skilled migrants resided in the United States, and approximately one-quarter resided in other Anglo-Saxon countries.

2. High-skilled immigration surged in Switzerland after the enactment of a bilateral agreement between Switzerland and the European Union on the free movement of labor in June 2007.

THE WORLD BANK ECONOMIC REVIEW, VOL. 27, NO. 2, pp. 297–319 doi:10.1093/wber/lht002
Advance Access Publication February 28, 2013

find that skilled natives tend to oppose immigration in states with a relatively skilled mix of immigrants more than in states in which the skill composition of immigrants features a high proportion of low-skilled immigrants. Similarly, a recent panel study by Müller and Tai (2010) for Europe suggests that higher-skilled workers have less favorable attitudes toward immigration when immigrants are more skilled relative to the average skill level in the destination country.

This paper examines whether domestic skilled workers have reason to oppose high-skilled immigration and, vice versa, whether nonmigrating high-skilled workers win or lose from brain drain in source countries. We argue that the international migration of high-skilled workers triggers productivity effects at the macro level such that the wage rate of skilled workers may rise in host countries and decline in source countries. By exploiting data on international bilateral migration flows from Docquier, Marfouk and Lowell (2007), we empirically examine the impact of an increase in high-skilled emigration rates on log differences in GDP per capita, total factor productivity (TFP), and the wage income of skilled workers between pairs of source and destination countries. We propose a range of instrumental variables to address the potential reverse causality problem that arises when international wage differences affect individual migration decisions (e.g., Lucas, 2005; Egger and Radulescu, 2009; Grogger and Hanson, 2011).

Our theoretical model suggests that even when considering adjustments in educational decisions, an increase in high-skilled emigration (immigration) lowers (raises) the domestic skill intensity in production.[3] This relationship has two effects on the relative wages of high-skilled workers between destination and source economies. First, for a given TFP and as a consequence of the declining marginal productivity of a certain type of labor, high-skilled workers lose in the destination economy and win in the source economy. However, external effects of migration on TFP (positive in destination, adverse in source) may reverse this result. The net effect of high-skilled migration on international wage differences is thus theoretically ambiguous. This theoretical approach makes the relationship between high-skilled migration and wages an empirical question. Our analysis suggests that, if anything, the external productivity effect is likely to dominate. Moreover, because of complementarity between high-skilled and low-skilled labor, an increase in low-skilled migration unambiguously benefits high-skilled workers in the receiving country.

Our findings are consistent with the recent literature on the wage effects of high-skilled immigration in single countries. Borjas (2003) and Dustmann, Fabbri and Preston (2005) provide evidence for a small but positive impact of

3. Grossmann and Stadelmann (2011) develop an overlapping-generations model with endogenous education choice that shows how migration is triggered by a decrease in the mobility costs of high-skilled workers and how it may evolve over time. In the present paper, we focus empirically on the effect of higher international migration.

an inflow of immigrants with a college degree on wages for college-educated natives in the United States and United Kingdom, respectively. Similarly, Friedberg (2001) suggests that native wages may rise after immigrants enter high-skilled occupations in the Israeli labor market. Our empirical contribution is to provide international evidence for the theoretical possibility of positive wage effects in destination countries relative to source countries. We exploit data on bilateral migration between country pairs, thereby complementing single-country studies on labor market effects of immigration.

Another strand of literature has emphasized the positive effects of brain drain on market income in the source economy (e.g., Mountford, 1997; Stark, Helmenstein and Prskawetz, 1997; Beine, Docquier and Rapoport, 2001, 2008). This possibility arises from the idea that an increase in immigration quotas in advanced countries improves immigration prospects for skilled workers in developing countries and thereby raises incentives to acquire education. However, empirically, the net effect on the size of the skilled labor force appears to be positive, except for very poor countries and/or countries with low levels of human capital (Beine et al., 2001, 2008). In our theoretical framework, brain drain reduces the skill intensity in the source country, even when educational decisions are adjusted. Because our empirical framework investigates the effect of skilled migration on relative outcomes between destination and source, we do not test the alternative hypothesis advanced in the "brain gain" literature. We can conclude, however, that the destination country tends to gain more from skilled migration than the source country.

The remainder of this paper is organized as follows. Section I presents a simple theoretical model. The model provides the basis for the empirical analysis in section II on the effects of higher emigration on relative GDP per capita, relative TFP, and the relative wage income of skilled workers between the source and the destination. The last section provides concluding remarks.

I. Theoretical Considerations

Our theoretical analysis shows that the presence of the external productivity effects of skilled labor implies that in response to an increase in high-skilled migration, the wage level of educated workers may increase in the host country relative to the source country.

Set Up

Consider two economies, home and foreign. There is a homogenous consumption good, which is chosen as the numeraire. Output Y is produced under perfect competition according to the technology

$$Y = AF(H, L) \equiv ALf(k) \tag{1}$$

where H and L denote the high-skilled and low-skilled labor inputs, respectively, A is TFP, the function F is linearly homogenous, $k \equiv H/L$ denotes the skill

intensity of production, and $f(k) \equiv F(k, 1)$. Furthermore, f is increasing, strictly concave, and fulfills the standard boundary conditions.

Before migration, there is (for simplicity) the same number N of individuals/workers in both countries. There is a positive external effect of a higher concentration of skilled labor, $h \equiv H/N$, on TFP,

$$A = a(h) \tag{2}$$

where a is an increasing function. This assumption captures human capital externalities as formalized, for instance, by Lucas (1988) in the context of endogenous growth. These human capital externalities may arise from learning spillover effects across workers, increased innovation activity in firms, and better institutional quality in a country, which may be associated with a more highly skilled domestic population. The empirical literature on human capital externalities is somewhat inconclusive but is mostly supportive. For instance, Acemoglu and Angrist (2000) find modest evidence in favor of human capital externalities from secondary schooling, whereas Ciccone and Peri (2006) find no evidence. Iranzo and Peri (2009) argue in favor of strong human capital externalities from college graduates in the United States but not from an increased share of high school graduates. In a recent study, Gennaioli et al. (2011) find strong empirical evidence of human capital externalities. They employ a new data set with 1569 subnational regions from 110 countries and argue that human capital is the primary driver of regional development. Moreover, they complement their finding with firm-level evidence on regional education levels for productivity and find large effects. The authors conclude that the previous empirical literature has underestimated the magnitude of human capital externalities. Similarly, Hunt (2011) employs a U.S. state panel data set for the period from 1940 to 2000 to show that an increase in the share of the immigrant college graduate population of one percentage point increases the number of patents per capita by 9–18 percent. This is strong evidence in favor of the hypothesis that skilled immigration increases TFP.

Each individual decides whether to become skilled and whether to migrate. Both skilled and unskilled individuals are internationally mobile, but they may differ in migration costs. Formally, let c_i denote the consumption level of individual i. Utility level u_i is given by

$$u_i = \begin{cases} c_i \text{ if } i \text{ stays at home} \\ c_i/\theta_i \text{ if } i \text{ migrates} \end{cases} \tag{3}$$

where $\theta_i = \theta^H > 1$ if i is skilled and $\theta_i = \theta^L > 1$ if i is unskilled. The modeling of migration costs as discounted consumption follows Stark et al. (1997), among others. Education comes at time cost $e_i \geq 0$, which may be interpreted to be a learning cost. Whereas an unskilled individual supplies one unit of time

to a perfect labor market, a skilled individual i supplies only $1 - e_i$ units of time. The wage rate per unit of time of high-skilled and low-skilled individuals at home is denoted by w_H and w_L, respectively. Moreover, denote all foreign variables and functions with superscript *. Therefore, the consumption of individual i born at home is given by

$$c_i = \begin{cases} (1 - e_i)w_H \text{ if } i \text{ is skilled and stays at home} \\ w_L \text{ if } i \text{ is unskilled and stays at home} \\ (1 - e_i)w_H* \text{ if } i \text{ is skilled and emigrates} \\ w_L* \text{ if } i \text{ is unskilled and emigrates.} \end{cases} \tag{4}$$

Denote by $G(e)$ the cumulative distribution function of the learning cost e in the population at home. For convenience, suppose that G is continuously differentiable. We allow functions G^*, F^*, and a^* (characterizing the foreign country) to be different to functions G, F, and a, respectively.

As will become apparent, the equilibrium outcome is the same regardless of whether we assume that migration possibilities are already considered in the education decisions of individuals. This condition is an implication of the simplifying assumptions that (i) learning abilities and migration costs are uncorrelated and (ii) individual migration costs are the same for all workers within a skill group.

Derivation of Testable Hypotheses

We will now derive the testable hypotheses. For this purpose, we treat migration as exogenous. According to equations (1) and (2), competitive factor prices are as follows:

$$w_H = a(h)f'(k), \tag{5}$$

$$w_L = a(h)[f(k) - kf'(k)]. \tag{6}$$

According to equations (3) and (4), an individual of skill type $j \in \{H, L\}$ chooses to migrate if $w_j^*/\theta^j \geq w_j$; therefore, in an interior equilibrium,

$$\frac{w_H^*}{\theta^H} = w_H, \ \frac{w_L^*}{\theta^L} = w_L. \tag{7}$$

A nonmigrating individual i chooses education whenever $(1 - e_i)w_H \geq w_L$. Moreover, staying at home and being educated yields higher utility than migrating and remaining unskilled if $(1 - e_i)w_H \geq w_L^*/\theta^L = w_L$, which is the same condition. Similarly, we find that a migrating individual chooses education if $(1 - e_i)w_H^*/\theta^H \geq w_L^*/\theta^L$, which, in view of equation (7), again gives us

condition $(1 - e_i)w_H \geq w_L$. Moreover, migrating and being educated yields higher utility than not migrating and remaining unskilled if $(1 - e_i)w_H^*/\theta^H = (1 - e_i)w_H \geq w_L$.

Therefore, all individuals with learning costs below some endogenous threshold level, $\bar{e}$, which depends on domestic wages only, become skilled:

$$e_i \leq 1 - \frac{w_L}{w_H} = 1 - \frac{f(k) - kf'(k)}{f'(k)} \equiv \bar{e}(k). \tag{8}$$

Because $f'' < 0$, we have $\bar{e}' < 0$. As the skill intensity, k, increases, the wage rate of unskilled individuals relative to skilled individuals, w_L/w_H, increases; consequently, more individuals remain unskilled, indicating that the threshold learning cost $\bar{e}$ is lower.

The fraction of domestically born unskilled workers, U, is given by

$$U = 1 - G(\bar{e}(k)) \equiv \tilde{U}(k) \tag{9}$$

where $\tilde{U}' > 0$. The effective units of skilled labor in the home country per native, before migration, are given by[4]

$$S = \int_0^{\bar{e}(k)} (1 - e)dG(e) \equiv \tilde{S}(k). \tag{10}$$

Therefore, $\tilde{S}' < 0$.

Denote by m_S and m_U the fraction of skilled and unskilled labor units emigrating to the foreign country ("emigration rates"), respectively. After migration, we have $h := H/N = S - m_S$ and $l : L/N = U - m_U$, respectively. Therefore, using equations (9) and (10), the skill intensity at home, $k = H/L$, is implicitly given by

$$k = \frac{\tilde{S}(k) - m_S}{\tilde{U}(k) - m_U}. \tag{11}$$

Using $\tilde{U}' > 0$ and $\tilde{S}' < 0$, we see that the right-hand side of equation (11) is decreasing in k. Therefore, in an interior labor market equilibrium, the skill intensity given by equation (11), denoted by $k \equiv \tilde{k}(m_S, m_U)$, is unique. The function $\tilde{k}$ is decreasing in the emigration rate of skilled labor, m_S, and increasing in the emigration rate of unskilled labor, m_U.

4. Recall that individual i provides $1 - e_i$ units of skilled labor when $e_i \leq \bar{e}(k)$.

In a two-country world, emigrants of one country are immigrants of the other country. Therefore, the foreign skill intensity k^* is uniquely given by[5]

$$k^* = \frac{\tilde{S}^*(k^*) + m_S}{\tilde{U}^*(k^*) + m_U}. \tag{12}$$

We write $k^* \equiv \tilde{k}^*(m_S, m_U)$. The function $\tilde{k}^*$ is increasing in m_S and decreasing in m_U.

Using $h = S - m_S$ and $h^* = S^* + m_S$, TFP in the foreign (host) country relative to the home (source) country can be written as[6]

$$\alpha \equiv \frac{A^*}{A} = \frac{a^*(\tilde{S}^*\left(\tilde{k}^*(m_S, m_U)\right) + m_S)}{a(\tilde{S}\left(\tilde{k}(m_S, m_U)\right) + m_S)} \equiv \tilde{\alpha}(m_S, m_U) \tag{13}$$

according to equation (2). Moreover, according to equation (5), the relative wage rate for skilled workers is

$$\omega_H \equiv \frac{w_H^*}{w_H} = \frac{a^*(\tilde{S}^*\left(\tilde{k}^*(m_S, m_U)\right) + m_S)(f^*)'(\tilde{k}^*(m_S, m_U))}{a(\tilde{S}\left(\tilde{k}(m_S, m_U)\right) + m_S)f'(\tilde{k}(m_S, m_U))} \equiv \tilde{\omega}_H(m_S, m_U). \tag{14}$$

Define the elasticities of the skill intensity at home and in the foreign country with respect to the migration of skilled and unskilled labor from the home country to the foreign country:

$$\kappa_S \equiv -\frac{m_S}{\tilde{k}}\frac{\partial \tilde{k}}{\partial m_S}, \ \kappa_U \equiv \frac{m_U}{\tilde{k}}\frac{\partial \tilde{k}}{\partial m_U}, \tag{15}$$

$$\kappa_S^* \equiv \frac{m_S}{\tilde{k}^*}\frac{\partial \tilde{k}^*}{\partial m_S}, \ \kappa_U^* \equiv -\frac{m_U}{\tilde{k}^*}\frac{\partial \tilde{k}^*}{\partial m_U}. \tag{16}$$

Note that the elasticities are defined such that they are positive: κ_S, κ_U, κ_S^*, $\kappa_U^* > 0$. Moreover, define by

$$\varepsilon(h) \equiv \frac{ha'(h)}{a(h)} \tag{17}$$

5. Functions $\tilde{U}^*$ and $\tilde{S}^*$ are defined analogously to equations (9) and (10), respectively.
6. Without a loss of generality, we label the foreign country the host country.

$$\eta(k) \equiv -\frac{kf''(k)}{f'(k)} \tag{18}$$

the elasticity of TFP with respect to skilled labor per native h and the elasticity of f with respect to skill intensity k. (We define ε^* and η^* analogously.)

It is simple to show the following results. First, the elasticity of relative destination-to-source TFP ($\alpha = A^*/A$) with respect to the emigration rate of the skilled (m_S) and unskilled (m_U) is given by

$$\frac{m_S}{\tilde{\alpha}}\frac{\partial\tilde{\alpha}}{\partial m_S} = \varepsilon(h)\left(\frac{\tilde{S}'(k)}{l}\kappa_S + \frac{m_S}{h}\right) + \varepsilon^*(h^*)\left(\frac{(\tilde{S}^*)'(k^*)}{l^*}\kappa_S^* + \frac{m_S^*}{h^*}\right), \tag{19}$$

$$\frac{m_U}{\tilde{\alpha}}\frac{\partial\tilde{\alpha}}{\partial m_U} = -\,\varepsilon(h)\frac{\tilde{S}'(k)}{l}\kappa_U - \varepsilon^*(h^*)\frac{(\tilde{S}^*)'(k^*)}{l^*}\kappa_U^* \tag{20}$$

respectively. Therefore, if the effect of a change in the skill intensity (triggered by migration) on the education decision is small (i.e., the magnitude of derivatives $\tilde{S}'$, $(\tilde{S}^*)' < 0$ are small), the model predicts that an increase in the migration rate of skilled labor (m_S) has a positive effect on relative destination-to-source TFP (α). Moreover, an increase in the migration rate of unskilled labor (m_U) has a positive but small effect on α because the migration of unskilled labor only has an indirect TFP effect by lowering education incentives in the source country (and vice versa in the destination country). By contrast, as a result of human capital externalities (ε, $\varepsilon^* > 0$), the emigration of skilled labor also has a direct TFP effect on skilled labor input per native (h) in the source country (and, again, vice versa in the destination country). The effect is mitigated because an increase in m_S fosters education incentives in the source country (and provides disincentives in the destination country).

Second, the elasticity of the destination-to-source relative wage income of skilled labor ($\omega_H = w_H^*/w_H$) with respect to the emigration rate of skilled and unskilled labor is given by

$$\frac{m_S}{\tilde{\omega}_H}\frac{\partial\tilde{\omega}_H}{\partial m_S} = \frac{m_S}{\tilde{\alpha}}\frac{\partial\tilde{\alpha}}{\partial m_S} - \eta(k)\kappa_S - \eta^*(k^*)\kappa_S^*, \tag{21}$$

$$\frac{m_U}{\tilde{\omega}_H}\frac{\partial\tilde{\omega}_H}{\partial m_U} = \frac{m_U}{\tilde{\alpha}}\frac{\partial\tilde{\alpha}}{\partial m_U} + \eta(k)\kappa_U + \eta^*(k^*)\kappa_U^* \tag{22}$$

respectively. Therefore, the impact of the migration of unskilled labor (increase in m_U) on the relative destination-to-source wage income of skilled labor is

unambiguously positive. The relative TFP increases as a result of education effects, and the resulting increase in skill intensity k reduces the wages of skilled labor in the source country (and vice versa in the destination country, where the skill intensity decreases). By contrast, for a given TFP, the wage rate of skilled labor decreases with the skill intensity; therefore, the impact of the migration of skilled labor (increase in m_S) on the relative destination-to-source wage income of skilled labor (ω_H) is ambiguous, even if the relative destination-to-source TFP (α) increases. Only if the TFP effects are sufficiently large owing to human capital externalities does an increase in m_S increase ω_H.

In sum, we predict that an increase in the emigration rate of high-skilled labor (m_S) increases relative TFP $\alpha = A^*/A$, whereas the impact of the emigration of unskilled labor (m_U) on α may be small. Moreover, an increase in m_U has a positive and possibly large effect on the relative wages of the skilled, $\omega_H = w_H^*/w_H$. Finally, an increase in m_S may lead to an increase in ω_H if TFP effects are sufficiently large. These are important theoretical results for political debate in some destination countries of skilled workers.

We have focused the theoretical analysis on the predictions of the effects of migration, although we allowed individuals to consider the migration decision when choosing education. Because migration is endogenous according to the model and depends (inter alia) on international wage differences, the model also indicates an empirical endogeneity issue, which we try to address by using instrumentation strategies.

II. Empirical Analysis

Our theoretical analysis has highlighted the effect of the emigration of high-skilled and low-skilled labor on TFP differences and the wage income gap of skilled labor to potential host economies of expatriates. We have seen that there may be counteracting channels by which skilled migration affects the wages of skilled workers: the external TFP effects of migration and the effect on the marginal productivity of skilled labor when TFP is held constant.

The direction from (wage) income differences to migration flows has been examined empirically elsewhere. Two recent papers are notable. First, Grogger and Hanson (2011) provide convincing evidence for the critical role of wage differences between country pairs on the emigration patterns of tertiary educated workers.[7] Second, Beine et al. (2011) show that in addition to wage differences, network effects are important for the migration decisions of both

7. In Grossmann and Stadelmann (2008), we presented evidence for the interaction between emigration flows and income changes using a structural equation model. However, we examined the impact of a higher aggregate emigration stock of a country on its per capita income. That is, we did not consider bilateral relationships.

high-skilled and low-skilled workers. The authors show that emigrants already living in the destination country positively affect migration flows.[8]

Our analysis complements research on the interaction between wage differences and skilled migration by focusing on the opposite direction, the impact of migration on both international (wage) income differences for skilled workers and TFP differences between country pairs. Inter alia, we instrument skilled migration with past migration stocks, as suggested by Beine et al. (2011).

Data and Estimation Strategy

The emigration rate of highly skilled individuals is our main explanatory variable. Docquier and Marfouk (2006) established a dataset of emigration stocks and rates by educational attainment for the years 1990 and 2000. The authors count as emigrants all foreign-born individuals at least 25 years old who live in an OECD country and class them by educational attainment and country of origin. Thereby, emigration into OECD countries is captured, representing approximately 90 percent of educated migrants in the world.[9] Because we are interested in bilateral migration patterns, we employ an extended dataset by Docquier et al. (2007). We construct the high-skilled emigration rate from country i to j in year t, denoted by $SMig_{ij,t}$, as the stock of skilled emigrants from country i living in (OECD) country j divided by the stock of skilled residents in (source) country i. In some regressions, we also control for the (lagged) low-skilled emigration rate, $UMig_{ij,t\text{-}1}$, which is constructed analogously.

Denote by $y_{i,t}$ the outcome measure in country i in year t. We consider GDP per capita, TFP, and the wage income of skilled workers. For a country pair (i, j), we estimate

$$\log\left(\frac{y_{j,t}}{y_{i,t}}\right) = \beta_0 + \beta_1 SMig_{ij,t} + \beta_2 UMig_{ij,t-1} + x'_{ij,t-1}\beta_x + u_{ij}. \tag{23}$$

Equation (23) is theoretically motivated by relationships $w_H^*/w_H = \tilde{\omega}_H(m_S, m_U)$ and $A^*/A = \tilde{\alpha}\ (m_S,\ m_U)$; see equations (14) and (13) derived in section I, respectively. According to equation (19), the theoretical model suggests that $\beta_1 > 0$ when the log difference in TFP, log (A^*/A), is the dependent variable. When the log difference of wages for skilled workers, log (w_H^*/w_H), is the dependent variable, then we predict $\beta_1 > 0$ if and only if the TFP effects of migration are sufficiently high, according to equation (21). Moreover, we predict $\beta_2 > 0$ when log (w_H^*/w_H) is the dependent variable. $x_{ij,t-1}$ is a vector of other (lagged) controls that potentially affect log income differences between i and j,

8. This finding suggests that there exist mobility-cost reducing network effects from communities of people from the same nation and from friends and relatives already living abroad (see also Massey et al., 1993).

9. See Docquier and Marfouk (2006) for a detailed discussion concerning data collection and construction issues.

such as relative school enrollment rates, relative investment rates, relative urban population shares, and fixed effects for the source country to capture institutional differences to OECD destination countries. With respect to the dependent outcome measures, we focus on the year 2000 and usually measure controls other than skilled migration in lagged form (for 1990) to reduce endogeneity bias. u_{ij} is an error term.

To construct a measure of log (w_H^*/w_H), we would like to use (log) wages differences for high-skilled individuals. However, because wage income by education category is not available, we construct several empirical proxy measures. Freeman and Oostendorp (2000) analyze information on earnings by occupation and industry from the ILO October Inquiry Survey from 1983 to 1998 for a number of countries.[10] For each country, we use Freeman and Oostendorp's earnings measures to calculate the 80th and the 90th percentiles as two measures for wages of high-skilled workers. For most countries, data are available for only a few years. Therefore, for each country, we take the mean across the period from 1995 to 2003 to obtain wage data for the year 2000.[11] The two constructed (log) relative wage variables for the 80th and the 90th percentiles are denoted by $RelWage80_{ij,t}$ and $RelWage90_{ij,t}$.

One may argue that migrating skilled workers do not receive wage income in the same percentile as they do at home. In particular, high-skilled workers from developing countries may not be considered highly skilled in the destination country. Therefore, as a robustness check, we assume that someone working in the 80th percentile at home earns only the median wage income abroad. The corresponding relative wage measure is denoted by $RelWage80to50_{ij,t}$.

For relative GDP and relative TFP between destination and source countries, denoted by $RelGDP_{ij,t}$ and $RelTFP_{ij,t}$, respectively, we use Penn World Tables and the UNIDO World productivity database. In particular, GDP data have better availability than wage data such that the number of observations increases. Details on variable definitions, data sources, and the summary statistics of the employed variables are presented in the appendix (table A1).

As indicated, although recent empirical literature has focused on the impact of income differences on migration patterns, we aim to examine the opposite channel. In a first attempt to address endogeneity, we replace the high-skilled emigration rate in 2000 by the lagged one in 1990, denoted by $SMig_{ij,t-1}$, in OLS regressions. Doing so allows for the possibility that the TFP effects of the migration flows of skilled workers (for instance, through innovation activity) take time to come into effect.

10. To correct for differences in how countries report earnings, Freeman and Oostendorp (2000) use a standardization procedure to make the data comparable across countries and time. In 2005, they provided an update for their earnings measures for the 1983-2003 ILO October Inquiry data using an improved version of the standardization procedure and the application of country-specific data type correction factors. A detailed technical documentation of the standardization procedure for the 1983-2003 ILO October Inquiry data is available online at http://www.nber.org/oww/.

11. We also included data for Turkey for the year 1994.

Second, we explore potential instruments for the high-skilled emigration rate for 2000. We use the lagged rate of expatriates in 1990 emigrating from country i to j, denoted by $TotalMig_{ij,t-1}$, as an instrument for $SMig_{ij,t}$, thereby predicting the rate of high-skilled emigrants by the lagged rate of all emigrants. This approach is motivated by the notion that a larger percentage of emigrants from a certain source country already living abroad act as a signal to potential high-skilled migrants regarding the destination country's openness and its administrative bodies' treatment of foreigners. The presence of more emigrants to a certain destination creates mobility cost-reducing network effects for potential emigrants (e.g., Massey et al., 1993; Beine, Docquier and Ozden, 2011).[12] Past migration also measures other intangible factors unrelated to income, such as trust, cultural proximity, and social openness to migrants of the destination as perceived by emigrants of the source country. Moreover, we employ indicators for geographical factors ($Dist_{ij}$, $Contig_{ij}$) and linguistic proximity ($ComLang_{ij}$), which are typically used in the literature on migration as additional instruments.

To further address potential endogeneity bias, we use the total emigration rate in 1960 instead of $TotalMig_{ij,t-1}$ as an instrument, which, however, cannot be readily observed at the time of analysis. We thus construct a proxy for the total emigration rate. Denote by $NetMig_{i,1960}$ the total net emigration rate (the number of emigrants minus the number of immigrants divided by population size) in country i in 1960, provided by the United Nations Population Division.[13] Our measure of bilateral total emigration rates in 1960 is defined by

$$TotalMig_{ij,1960} = \frac{NetMig_{i,1960}}{100} \times \frac{Pop_{j,1960}}{Pop_{i,1960}} \tag{24}$$

where $Pop_{i,1960}$ is population size in the source i and $Pop_{j,1960}$ is the population size in the destination j in the year 1960.[14] As suggested by Beine, Docquier and Rapoport (2001), one may use countries' population sizes to reflect immigration quotas. $NetMig_{i,\ 1960} \times Pop_{j,\ 1960}$ is thus a proxy for the net stock of emigrants from country i received in country j in 1960. Because our empirical strategy focuses on emigration rates rather than stocks, we divide this measure by (100 times) the population size of source country i to obtain an estimate for the past bilateral emigration rate.[15] The fraction of high-skilled migrants before 1960 was comparatively low; therefore, potential effects of past migration should only

12. Another way to capture the effect of mobility cost-reducing network effects is to use the past total number of migrants instead of the past emigration rate as the instrument for contemporaneous migration. We confirm that the results do not change.

13. Countries with negative net emigration are coded to have an emigration rate equal to zero.

14. The measure is inspired by Beine, Docquier and Ozden (2011). They use a similarly constructed proxy as an instrument for the total diaspora of migrants in 1990 (rather than the high-skilled emigration rate).

15. Calculating partial correlations confirms that the past total emigration rate is well correlated with the high-skilled emigration rate in 2000.

work through induced high-skilled emigration. In other words, the instrument should be uncorrelated with the dependent variable, which is supported by J-tests.

Results

Reported standard errors from all estimates account for destination clusters, following Grogger and Hanson (2011), among others.[16]

Table 1 presents OLS estimates of equation (23). We first omit the low-skilled migration rate. We observe that the estimated effects of an increase in the high-skilled migration rate on relative GDP ($RelGDP_{ij,t}$), relative TFP ($RelTFP_{ij,t}$), and relative wages ($RelWage80_{ij,t}$ and $RelWage90_{ij,t}$) between destination and source countries are positive and significant. Using the lagged high-skilled migration rate ($SMig_{ij,t-1}$) rather than the contemporaneous one ($SMig_{ij,t}$) only slightly decreases the coefficient. Therefore, an increase in the high-skilled emigration rate increases (log) income differences between countries. The control variables of all estimates include the lagged relative school enrolment (primary and tertiary), the relative capital investment, and the relative urban population share as well as source fixed effects. Except for (lagged) primary school enrollment, which is never significant, the controls have the expected signs. The (lagged) relative investment rate and the (lagged) relative urban population share are typically significantly different from zero.

To consider the effect quantitatively, we use a coefficient β_1 of about 0.2 in the wage regressions presented in columns (5)–(8). Doubling the high-skilled emigration rate ($SMig_{ij,t}$) from its mean level of 0.025 thus implies that the relative wage for high-skilled workers between the destination and the source increases by approximately 0.5 percent ($=0.2 \times 0.025$).[17] This effect is small, which is consistent with the microeconomic estimates of the effect of high-skilled immigration on wages for highly skilled individuals inside the United States by Borjas (2003) and for the United Kingdom by Dustmann et al. (2005).

Tables 2–4 address the potential problem of reverse causality by providing instrumental variable estimations of (23). The upper panels report second-stage results, whereas the lower panels in tables 2 and 3 report the partial correlations of the instruments in the first stage.

We start with the results for relative GDP as a dependent variable in table 2. In columns (1) and (2), we use the total emigration rate from country i to j in 1990 ($TotalMig_{ij,t-1}$) as a single instrument. In columns (3)–(6), the bilateral geographical distance between i and j ($Dist_{ij}$), an indicator for a common border ($Contig_{ij}$), and an indicator for the common language of the source and

16. We use the Huber-White method implemented in the R Design package to adjust the variance-covariance matrix from our least squares results.

17. In fact, between 1990 and 2000, the number of tertiary-educated immigrants living in OECD countries almost doubled (Docquier and Marfouk, 2006).

TABLE 1. Effect of High-Skilled Emigration Rates on Wage, GDP, and TFP Differences between Countries

	Dependent variable: **RelGDP$_{ij,t}$**		Dependent variable: **RelTFP$_{ij,t}$**		Dependent variable: **RelWage80$_{ij,t}$**		Dependent variable: **RelWage90$_{ij,t}$**	
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
SMig$_{ij,t}$	0.1630***		0.0830***		0.2168***		0.2290***	
	(0.0276)		(0.0140)		(0.0490)		(0.0483)	
SMig$_{ij,t-1}$		0.1386***		0.0796***		0.1645**		0.1738**
		(0.0418)		(0.0198)		(0.0678)		(0.0699)
RelInvest$_{ij,t-1}$	0.2331*	0.2317*	0.0333	0.0327	0.4989**	0.4975**	0.4356*	0.4341*
	(0.1216)	(0.1215)	(0.0618)	(0.0617)	(0.2533)	(0.2533)	(0.2430)	(0.2430)
RelUrban$_{ij,t-1}$	0.2113***	0.2109***	0.0617	0.0615	0.6594**	0.6587**	0.5761*	0.5754*
	(0.0805)	(0.0806)	(0.0432)	(0.0433)	(0.3052)	(0.3054)	(0.3015)	(0.3017)
RelPrimSchool$_{ij,t-1}$	−0.3658	−0.3683	−0.4618	−0.4634	−1.0022	−1.0057	−0.5458	−0.5495
	(0.7655)	(0.7668)	(0.3875)	(0.3882)	(2.2117)	(2.2127)	(2.0325)	(2.0336)
RelTertSchool$_{ij,t-1}$	0.0046	0.0047*	0.0022*	0.0022*	0.0105	0.0106	0.0104	0.0105
	(0.0028)	(0.0028)	(0.0013)	(0.0013)	(0.0102)	(0.0101)	(0.0099)	(0.0099)
(Intercept)	3.6064	3.6211	0.6013	0.6045	0.6731	0.6802	0.3170	0.3245
	(3.0786)	(3.0845)	(0.4408)	(0.4415)	(2.7047)	(2.7058)	(2.5903)	(2.5916)
Origin FE	YES	YES	YES	YES	YES	YES	YES	YES
Adj. R^2	0.9429	0.9428	0.9541	0.9541	0.8584	0.8582	0.8555	0.8553
N	2275	2275	1550	1550	1010	1010	1010	1010

Note: All dependent variables are expressed in logs and represent relative differences between countries j and i. SMig$_{ij,t}$ denotes the stock of high-skilled emigrants from country i living in country j divided by the stock of high-skilled residents in i. RelInvest$_{ij,t-1}$, RelUrban$_{ij,t-1}$, RelPrimSchool$_{ij,t-1}$ and RelTertSchool$_{ij,t-1}$ denote the lagged relative investment share, relative urbanization share, relative primary school enrollment, and relative tertiary school enrollment between j and i. Robust clustered standard errors are in parentheses. *** indicates a significance level below 1 percent; ** indicates a significance level between 1 and 5 percent; * indicates a significance level between 5 and 10 percent.

TABLE 2. Effect of High-Skilled Emigration Rates on GDP and TFP Differences between Countries (Instrumental Variables Estimations)

	Dependent variable: $\mathbf{RelGDP_{ij,t}}$						Dependent variable: $\mathbf{RelTPF_{ij,t}}$					
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*	*(11)*	*(12)*
$SMig_{ij,t}$	0.3036*	0.3269***	0.3017**	0.3015***	0.3883* (0.2371)	0.5138**	0.1771**	0.1452***	0.1863**	0.1437***	0.3569***	0.4021***
	(0.1601)	(0.0882)	(0.1532)	(0.0875)		(0.2064)	(0.0784)	(0.0235)	(0.0734)	(0.0256)	(0.0587)	(0.0703)
$UMig_{ij,t-1}$		−0.1677		−0.0672		−0.9417		0.3707		0.3789 (0.4349)		−1.0854
		(0.3579)		(0.4101)		(0.8753)		(0.4117)				(0.8486)
Other controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Origin FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adj. R^2	0.9430	0.9434	0.9429	0.9431	0.9420	0.9422	0.9547	0.9549	0.9548	0.9549	0.9547	0.9549
N	2275	2275	2266	2266	2250	2250	1550	1550	1550	1550	1536	1536
F-Test (first stage)	-	-	12.67	22.69	14.40	16.89	14.46	30.53	14.69	30.00	14.90	16.81
J-Test	-	-	0.4611	0.4654	0.1397	0.3187	-	-	0.5060	0.3858	0.8406	0.9022
Instruments used	$TotalMig_{ij,t-1}$	$TotalMig_{ij,t-1}$	$TotalMig_{ij,t-1}$ + $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$	$TotalMig_{ij,t-1}$ + $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$	$TotalMig_{ij,1960}$ + $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$	$TotalMig_{ij,1960}$ + $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$	$TotalMig_{ij,t-1}$	$TotalMig_{ij,t-1}$	$TotalMig_{ij,t-1}$+ $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$	$TotalMig_{ij,t-1}$+ $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$	$TotalMig_{ij,1960}$ + $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$	$TotalMig_{ij,1960}$ + $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$
					First stage (partial correlations)							
$TotalMig_{ij,t-1}$	0.0124***	0.0322***	0.0123***	0.0323***			0.0184***	0.0437***	0.0187***	0.0438***		
	($3.7e^{-04}$)	($7.9e^{-04}$)	($3.8e^{-04}$)	($8.0e^{-04}$)			($5.8e^{-04}$)	(0.0010)	($6.0e^{-04}$)	(0.0010)		
$TotalMig_{ij,1960}$					$1.2e^{-04}$***	$1.0e^{-04}$***					$3.9e^{-04}$***	$3.2e^{-04}$***
					($1.1e^{-05}$)	($1.0e^{-05}$)					($3.1e^{-05}$)	($3.0e^{-05}$)
$Dist_{ij}$			−0.0166***	−0.0217***	−0.0265***	−0.0197***			−0.0198***	−0.0184***	−0.0365***	−0.0299***
			(0.0053)	(0.0045)	(0.0063)	(0.0059)			(0.0074)	(0.0060)	(0.0091)	(0.0087)
$ComLang_{ij}$			0.0227**	−0.0026	0.0943***	0.0615***			0.0054 (0.0126)	−0.0169	0.0836***	0.0545***
			(0.0108)	(0.0093)	(0.0126)	(0.0120)				(0.0102)	(0.0153)	(0.0148)
$Contig_{ij}$			−0.1009***	−0.0537***	−0.0606**	−0.0951***			−0.1992***	−0.0621**	−0.0736*	−0.1652***
			(0.0219)	(0.0189)	(0.0260)	(0.0246)			(0.0339)	(0.0278)	(0.0416)	(0.0403)

Note: All dependent variables are expressed in logs and represent relative differences between countries j and i. $SMig_{ij,t}$ ($UMig_{ij,t-1}$) denotes the stock of high- (low-) skilled emigrants from country i living in country j divided by the stock of high- (low-) skilled residents in i. All estimations include $RelInvest_{ij,t-1}$, $RelUrban_{ij,t-1}$, $RelPrimSchool_{ij,t-1}$ and $RelTertSchool_{ij,t-1}$ as additional control variables. $TotalMig_{ij,t-1}$, $Dist_{ij}$, $ComLang_{ij}$, and $Contig_{ij}$ represent the share of the emigrant population from country i living in country j, the distance between i and j, whether i and j share a common language, and whether i and j have a common border, respectively. Robust clustered standard errors are in parentheses. *** indicates a significance level below 1 percent; ** indicates a significance level between 1 and 5 percent; * indicates a significance level between 5 and 10 percent.

TABLE 3. Effect of High-Skilled Emigration Rates on Wage Differences between Countries (Instrumental Variables Estimations)

	Dependent variable: **RelWage80$_{ij,t}$**						Dependent variable: **RelWage90$_{ij,t}$**					
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*	*(11)*	*(12)*
$SMig_{ij,t}$	0.6026*** (0.1457)	0.2490*** (0.0673)	0.5948*** (0.1406)	0.2125*** (0.0653)	0.6676*** (0.2123)	0.5447 (0.4028)	0.5888*** (0.1443)	0.2795*** (0.0588)	0.5788*** (0.1374)	0.2360*** (0.0557)	0.6875*** (0.2193)	0.6382 (0.4059)
$UMig_{ij,t-1}$		5.2286*** (1.9307)		5.5204*** (2.0235)		2.7071 (4.5031)		4.5736*** (1.6763)		4.9178*** (1.8263)		1.5121 (4.2180)
Other controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Origin FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adj. R^2	0.8609	0.8611	0.8607	0.8610	0.8590	0.8608	0.8582	0.8585	0.8583	0.8589	0.8563	0.8581
N	1010	1010	1010	1010	1010	1010	1010	1010	1010	1010	1010	1010
F-Test (first stage)	25.09	68.44	24.74	69.07	15.91	18.73	25.09	68.44	24.74	69.07	15.91	18.73
J-Test	-	-	0.8055	0.7491	0.7022	0.6947	-	-	0.8055	0.7491	0.7022	0.6947
Instruments used	$TotalMig_{ij,t-1}$	$TotalMig_{ij,t-1}$	$TotalMig_{ij,t-1}$ + $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$	$TotalMig_{ij,t-1}$ + $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$	$TotalMig_{ij,1960}$ + $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$	$TotalMig_{ij,1960}$ + $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$	$TotalMig_{ij,t-1}$	$TotalMig_{ij,t-1}$	$TotalMig_{ij,t-1}$ + $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$	$TotalMig_{ij,t-1}$ + $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$	$TotalMig_{ij,1960}$ + $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$	$TotalMig_{ij,1960}$ + $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$
						First stage (partial correlations)						
$TotalMig_{ij,t-1}$	0.0198*** ($5.5e^{-04}$)	0.0196*** ($5.7e^{-04}$)	0.0196*** ($5.7e^{-04}$)	0.0463*** ($8.8e^{-04}$)			0.0459*** ($8.8e^{-04}$)	0.0198*** ($5.5e^{-04}$)	0.0196*** ($5.7e^{-04}$)	0.0463*** ($8.8e^{-04}$)		
$TotalMig_{ij,1960}$					$1.8e^{-04}$*** ($1.6e^{-05}$)	$9.5e^{-05}$*** ($1.7e^{-05}$)					$1.8e^{-04}$*** ($1.6e^{-05}$)	$9.5e^{-05}$*** ($1.7e^{-05}$)
$Dist_{ij}$			−0.0099 (0.0068)	−0.0205*** (0.0046)	−0.0145 (0.0096)	−0.0088 (0.0089)			−0.0099 (0.0068)	−0.0205*** (0.0046)	−0.0145 (0.0096)	−0.0088 (0.0089)
$ComLang_{ij}$			0.0277* (0.0151)	0.0050 (0.0102)	0.1294*** (0.0209)	0.0904*** (0.0197)			0.0277* (0.0151)	0.0050 (0.0102)	0.1294*** (0.0209)	0.0904*** (0.0197)
$Contig_{ij}$			−0.0772*** (0.0236)	−0.0207 (0.0159)	−0.0516 (0.0334)	−0.0830*** (0.0312)			−0.0772*** (0.0236)	−0.0207 (0.0159)	−0.0516 (0.0334)	−0.0830*** (0.0312)

Note: All dependent variables are expressed in logs and represent relative differences between countries j and i. $SMig_{ij,t}$ ($UMig_{ij,t-1}$) denotes the stock of high- (low-) skilled emigrants from country i living in country j divided by the stock of high- (low-) skilled residents in i. All estimations include $RelInvest_{ij,t-1}$, $RelUrban_{ij,t-1}$, $RelPrimSchool_{ij,t-1}$ and $RelTertSchool_{ij,t-1}$ as additional control variables. $TotalMig_{ij,t-1}$, $Dist_{ij}$, $ComLang_{ij}$, and $Contig_{ij}$ represent the share of the emigrant population from country i living in country j, the distance between i and j, whether i and j share a common language, and whether i and j have a common border, respectively. Robust clustered standard errors are in parentheses. *** indicates a significance level below 1 percent; ** indicates a significance level between 1 and 5 percent; * indicates a significance level between 5 and 10 percent.

TABLE 4. Effect of High-Skilled Emigration Rates on Wage Differences between Countries when Migrants Change from the 80th Percentile to the 50th Percentile

	Dependent variable: **RelWage80to50$_{ij,t}$**							
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
$SMig_{ij,t}$	0.1963*** (0.0511)		0.5774*** (0.1464)	0.2006** (0.0822)	0.5722*** (0.1421)	0.1707** (0.0751)	0.6249*** (0.2097)	0.4626 (0.3889)
$SMig_{ij,t-1}$		0.1461** (0.0679)						
$UMig_{ij,t-1}$				5.5716*** (2.0616)		5.8135*** (2.1127)		3.3417 (4.4442)
Other controls	YES	YES	YES	YES	YES	YES	YES	YES
Origin FE	YES	YES	YES	YES	YES	YES	YES	YES
Adj. R^2	0.8355	0.8353	0.8361	0.8381	0.8364	0.8384	0.8363	0.8382
N	1010	1010	1010	1010	1010	1010	1010	1010
F-Test (first stage)	-	-	25.09	68.44	24.74	69.07	15.91	18.73
J-Test	-	-	-	-	0.8617	0.8157	0.8015	0.7719
Instruments used	-	-	$TotalMig_{ij,t-1}$	$TotalMig_{ij,t-1}$	$TotalMig_{ij,t-1}$ + $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$	$TotalMig_{ij,t-1}$ + $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$	$TotalMig_{ij,1960}$ + $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$	$TotalMig_{ij,1960}$ + $Dist_{ij}$ + $ComLang_{ij}$ + $Contig_{ij}$

Note: All dependent variables are expressed in logs and represent relative differences between countries j and i. $SMig_{ij,t}$ ($UMig_{ij,t-1}$) denotes the stock of high- (low-) skilled emigrants from country i living in country j divided by the stock of high- (low-) skilled residents in i. All estimations include $RelInvest_{ij,t-1}$, $RelUrban_{ij,t-1}$, $RelPrimSchool_{ij,t-1}$ and $RelTertSchool_{ij,t-1}$ as additional control variables. $TotalMig_{ij,t-1}$, $Dist_{ij}$, $ComLang_{ij}$, and $Contig_{ij}$ represent the share of the emigrant population from country i living in country j, the distance between i and j, whether i and j share a common language, and whether i and j have a common border, respectively. Robust clustered standard errors are in parentheses. *** indicates a significance level below 1 percent; ** indicates a significance level between 1 and 5 percent; * indicates significance level between 5 and 10 percent.

destination countries ($ComLang_{ij}$) are used as instruments in addition to the total emigration rate. We use $TotalMig_{ij,t-1}$ in columns (3) and (4) and our proxy for the total emigration rate for 1960, $TotalMig_{ij,1960}$, in columns (5) and (6). As in table 1, we control for the lagged relative values of school enrollment, private investment, and urbanization and include source country fixed effects (results not shown). The effect of high-skilled migration on log GDP differences between the destination and the source country is positive, as in the OLS estimations. All estimates suggest a significant and higher effect of skilled migration on relative GDP compared to the OLS estimates in table 1. Columns (2), (4), and (6) also control for the (lagged) low-skilled migration rate in 1990, $UMig_{ij,t-1}$. We observe that the coefficient on $UMig_{ij,t-1}$, β_2 in equation (23), is not significantly different from zero and does not alter the coefficient of the instrumented variable $SMig_{ij,t}$ in an important way.

Columns (7)–(12) in table 2 present the results for relative TFP analogously to columns (1)–(6). The results are similar to those for relative GDP: The estimated effect of high-skilled migration is always positive and increases compared with OLS estimates, whereas low-skilled migration is not significant. In particular, the estimates of β_1 in columns (7)–(12) of table 2 confirm our theoretical prediction that $\alpha = A^*/A$ is increasing in m_S as a result of human capital externalities. Again, the coefficient of the (lagged) low-skilled migration rate, β_2, is not significantly different from zero and is sometimes positive, in line with the theoretical model.

An F-test for the first-stage results shows that the instruments are significantly related to the emigration rate. In particular, past migration appears to be an important determinant of high-skilled migration.[18] None of the J-statistics suggest problems with the instruments.

In table 3, we present the results analogous to table 2 for relative wages in the 80th and 90th percentiles instead of relative GDP and relative TFP, respectively. Again, columns (1)–(2) and (7)–(8) use the total emigration rate in 1990 as a single instrument for the high-skilled emigration rate. The first-stage results indicate that the total emigration rate in 1990 is well correlated with $SMig_{ij,t}$. β_1 is again positive and significantly different from zero. According to the other estimations in table 3, the results are similar when using the measure for the total migration rate in 1960, geographical variables, and linguistic proximity as instruments. According to the theoretical prediction in equation (21), β_1 should be higher when relative TFP (α) rather than the relative wages of skilled labor (ω_H) is the dependent variable. In the estimates presented in tables 2 and 3, this is not the case. It is important to note, however, that sample sizes are very different because wage data are available for less (and, on average, richer) countries than TFP.

18. That contiguity (variable $Contig_{ij}$) has a negative effect on high-skilled emigration in our first-stage estimate parallels a finding similar to Grogger and Hanson (2011). They explain the result by selection and sorting effects.

Estimated coefficients of the instrumented high-skilled migration rate in 2000, $SMig_{ij,t}$, become smaller when we also control for the lagged low-skilled migration rate in 1990, $UMig_{ij,t-1}$. Moreover, coefficient β_2 of $UMig_{ij,t-1}$ is positive and typically significant (and is higher than β_1). This finding is in line with the theoretical prediction and is due to the complementarity between skilled and unskilled labor. Only in columns (6) and (12) does β_1 become insignificant; in these cases, it remains positive and quantitatively sizable.

In sum, we may conclude that the effect of skilled migration on international wage differences, albeit limited in magnitude, is positive and also often significant. The results of relative TFP in table 2 and those in table 3 in connection with our theoretical considerations appear to suggest that the possible positive effects of skilled immigration on the wages of skilled workers are derived from the positive TFP effects of skilled immigration. Moreover, low-skilled migration tends to benefit the skilled labor force in the receiving country.

The first-stage results in table 3 suggest that factors that are potentially unrelated to income, such as network effects, language, and geography, drive the high-skilled emigration rate. Interestingly, the coefficients of the instrumented variable $SMig_{ij,t}$ in table 3 are often more than twice as high as in OLS regressions (table 1). This finding suggests that migrants who arrive through social networks have a particularly high impact on international differences in (log) wages of skilled workers. Migrants who arrive through social networks appear to find it easier to integrate into the host country and thus have a larger effect on TFP (possibly being employed in jobs that are more suitable to their qualifications) than workers without social networks.

In fact, we cannot rule out that skilled immigrants work in different jobs than they do in the source country, often earning wages that are within a lower percentile of the wage distribution than at home. For instance, a university degree in a developing source country may reflect a lower acquired skill level than a university degree in an OECD destination country. Moreover, a skilled immigrant may occupy a low-skilled job briefly after arrival owing to language problems in the destination country. We account for these possibilities by using as the dependent variable the log difference between the wage of the median in the destination country and the 80th percentile in the source country, $RelWage80to50_{ij,t}$.

The results are reported in table 4. Columns (1) and (2) are analogous to the OLS estimations in table 1 and show similar results as the wage regressions (5)–(8) in table 1. Columns (3)–(8) are instrumental variable estimations, which are analogous, for instance, to columns (1)–(6) of table 3 with respect to the use of instruments. The instrumental variable estimates are similar in significance and magnitude to the results of the wage regressions in table 3.

We conduct a further sensitivity analysis (see tables S1 to S4 at http://wber.oxfordjournals.org/). This analysis suggests that our conclusions are fairly robust overall. First, we include destination fixed effects rather than source fixed effects. The results with destination fixed effects are similar to those with

source fixed effects.[19] We also examineed whether results are sensitive to a specific destination country. We run "rolling" regressions, in which we omit one destination country each time, to confirm that the results are basically unchanged. Second, we include regional dummies and a dummy variable that indicates whether the source country also belongs to the OECD[20] instead of fixed effects as controls to consider institutional differences, which may affect income differences, in an alternative way. Third, we employ an emigration data set by Defoort (2006) to construct an alternative proxy for the total emigration rate. The data set contains emigration to six important destination countries in the year 1975. The proxy is constructed analogously to equation (24) and is used as an instrument for the skilled migration rate. Finally, we use the stock of high-skilled and low-skilled migrants rather than migration rates as regressors. Our main conclusions remain qualitatively unchanged and overall robust.

III. Concluding Remarks

In this paper, we analyzed the impact of an increase in the international bilateral migration of high-skilled and low-skilled workers on relative income and relative TFP between pairs of source and destination countries of expatriates. Our theoretical model suggested that an increase in the number of skilled migrants increases international wage inequality by adversely affecting TFP in the source economy and increasing it in the host economy. Our empirical analysis provided evidence which is consistent with this hypothesis. Using a data set on the bilateral emigration of skilled workers, our results suggested that an increase in high-skilled emigration rates tends to slightly increase TFP differences and therefore (albeit also slightly) wage income for skilled workers in destination countries relative to source countries in a causal way. None of our estimations suggested that skilled workers in the destination country lose from skilled migration relative to the source country. Finally, skilled workers in the receiving countries unambiguously gain from low-skilled migration.

Acknowledgments

The authors are grateful to three anonymous referees for valuable comments and suggestions. Moreover, we thank Michel Beine, Bruno S. Frey, Mark Gradstein, Hillel Rapoport, Avi Simhon, and John Wilson for comments on an earlier draft. We also benefited from discussion with seminar participants at the Ben-Gurion University, University of Zurich, University of Geneva, University of Siegen, the Annual Meeting of the European Economic

19. We cannot include both simultaneously because they would, by construction, fully explain the different relative income variables due to multicollinearity.

20. Recall that all destination countries are OECD countries.

Association in Milan, and the conference "Globalization and the Brain Drain: Theory, Evidence and Policy" in Jerusalem and Ramat Gan. A supplementary appendix to this article is available at http://wber.oxfordjournals.org/.

Appendix Table A.1. Data Description and Sources

Variable	Description & Source	N	Mean	SD
$SMig_{ij,t}$	Stock of emigrants of educational category "high" aged 25+ born in country *i* and living in OECD country *j* in *t* (2000 or 1990) divided by stock of residents of educational category "high" in country *i* in year *t*. Stock of emigration and stock of residents of educational category "high" from Docquier, Marfouk and Lowell (2007).	3052	0.0246	0.1909
$RelGDP_{ij,t}$	Log of GDP per capita of country *j* minus log of GDP per capita of country *i* in year 2000. GDP data from Penn World Table Version 6.2.	3052	1.4360	1.2890
$RelTFP_{ij,t}$	Log of TFP (measure TPF_K06) per capita of country *j* minus log of TFP of country *i* in year 2000. UNIDO World Productivity Database, Isaksson (2007).	1983	0.7860	0.7628
$RelWage80_{ij,t}$	Log of wage in 80th percentile of country *j* minus log of wage in 80th percentile of country *i*. Wage data from Occupational Wages around the World Database.	1247	1.2650	1.4945
$RelWage90_{ij,t}$	Log of wage in 90th percentile of country *j* minus log of wage in 90th percentile of country *i*. Wage data from Occupational Wages around the World Database.	1247	1.1810	1.3953
$RelWage80to50_{ij,t}$	Log of wage in 80th percentile of country *j* minus log of wage in 50th percentile of country *i*. Wage data from Occupational Wages around the World Database.	1247	0.9409	1.4348
$UMig_{ij,t-1}$	Stock of emigrants of educational category "low" aged 25+ born in country *i* and living in OECD country *j* in *t-1* (1990) divided by stock of residents of educational category "low" in country *i* in *t-1*. Stock of emigration and stock of residents of educational category "low" from Docquier, Marfouk and Lowell (2007).	3052	0.0026	0.0197
$RelPrimSchool_{ij,t-1}$	Primary school enrollment in country *j* divided by primary school enrollment in country *i* in *t-1* (1990). Primary school enrollment rate from *GDF* and *WDI*.	2403	1.2040	0.5211

(Continued)

Appendix Table A.1. Continued

Variable	Description & Source	N	Mean	SD
$RelTertSchool_{ij,t-1}$	Tertiary school enrollment in country *j* divided by tertiary school enrollment in country *i* in *t-1* (1990). Tertiary school enrollment rate from *GDF* and *WDI*.	2477	10.2700	22.2216
$RelInvest_{ij,t-1}$	Investment share in country *j* divided by investment share in country *i* in *t-1* (1990). Investment share from Penn World Table Version 6.2.	3052	2.3350	1.9566
$RelUrban_{ij,t-1}$	Urban population share in country *j* divided by urban population share in country *i* in *t-1* (1990). Urban population share from *GDF* and *WDI*.	3013	2.0500	1.8872
$TotalMig_{ij,t-1}$	Emigrant population from country *i* living in country *j* divided by population (in thousands) of country *i* in *t-1* (1990). Docquier, Marfouk and Lowell (2007).	3052	1.6870	11.1509
$TotalMig_{ij,1960}$	Proxy of emigrant population from country *i* living in country *j* in year 1960. Constructed as described in text, based on data from the United Nations Population Division.	3052	1.6120	20.5570
$Dist_{ij}$	Log geodesic distance in km between countries *i* and *j*. Mayer and Soledad (2006).	3042	8.5170	0.9313
$ComLang_{ij}$	Dummy variable capturing if same language is spoken by at least 9 percent of the population in country *i* and *j*. Mayer and Soledad (2006).	3052	0.1311	0.3375
$Contig_{ij}$	Dummy variable capturing whether countries *i* and *j* are contiguous. Mayer and Soledad (2006).	3052	0.0269	0.1617

Note: The ranges, means, and standard deviations are not weighted and are based on the respective number of observations. Destination countries are the 30 OECD members. The total number of observations depends on the data availability for destination and source countries. Observations are excluded if bilateral data are not available or the source country does not have any emigrants in the destination country.

References

Acemoglu, D., and J. Angrist. 2000. "How Large are Human-Capital Externalities? Evidence from Compulsory Schooling Laws." *NBER Macroeconomics Annual* 15: 9–59.

Beine, M., F. Docquier, and H. Rapoport. 2001. "Brain Drain and Economic Growth: Theory And Evidence." *Journal of Development Economics* 64: 275–89.

———. 2008. "Brain Drain and Human Capital Formation in Developing Countries: Winners and Losers." *Economic Journal* 118: 631–52.

Beine, M., F. Docquier, and C. Ozden. 2011. "Diasporas." *Journal of Development Economics* 95: 30–41.

Borjas, G. J. 2003. "The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market." *Quarterly Journal of Economics* 118: 1335–74.

Defoort, C. 2006. "Tendances de Long Terme en Migrations Internationales: Analyse à Partir de 6 Pays Receveurs." Université Catholique de Louvain (mimeo), Louvain-la-Neuve, Belgium.

Docquier, F., and A. Marfouk. 2006. "International Migration by Educational Attainment (1990–2000) - Release 1.1." In C. Ozden and M. Schiff, eds., *International Migration, Remittances and Development*, 151–99. New York, NY: Palgrave Macmillan.

Docquier, F., B. L. Lowell, and A. Marfouk. 2007. "A Gendered Assessment of the Brain Drain." IZA Discussion Papers No. 3235. Institute for the Study of Labor, Bonn, Germany.

Dustmann, C., F. Fabbri, and I. Preston. 2005. "The Impact of Immigration on the British Labour Market." *Economic Journal* 115: F324–41.

Egger, P., and D. M. Radulescu. 2009. "The Influence of Labour Taxes on the Migration of Skilled Workers." *The World Economy* 32: 1365–79.

Friedberg, R. M. 2001. "The Impact Of Mass Migration On The Israeli Labor Market." *Quarterly Journal of Economics* 116: 1373–1408.

Freeman, R. B., and R. H. Oostendorp. 2000. "Wages Around the World: Pay Across Occupations and Countries." NBER Working Paper No. 8058. National Bureau of Economic Research, Cambridge, MA.

Gennaioli, N., R. La Porta, F. Lopez-de-Silanes, and A. Shleifer. 2011. "Human Capital and Regional Development." NBER Working Paper No. 17158. National Bureau of Economic Research, Cambridge, MA.

Grogger, J., and G. H. Hanson. 2011. "Income Maximization and the Selection and Sorting of International Migrants." *Journal of Development Economics* 95: 42–57.

Grossmann, V., and D. Stadelmann. 2008. "International Mobility of the Highly Skilled, Endogenous R&D, and Public Infrastructure Investment." IZA Discussion Paper No. 3366. Institute for the Study of Labor, Bonn, Germany.

———. 2011. "Does International Mobility of High-Skilled Workers Aggravate Between-Country Inequality?" *Journal of Development Economics* 95: 88–94.

Hanson, G. H., K. Scheve, and M. J. Slaughter. 2009. "Individual Preferences over High-Skilled Immigration in the United States." In J. Bhagwati, and G. Hanson, eds., *Skilled Immigration Today: Problems, Prospects, and Policies* 207–46. Oxford, UK: Oxford University Press.

Hunt, J., and M. Gauthier-Loiselle. 2011. "How Much Does Immigration Boost Innovation?" *American Economic Journal: Macroeconomics* 2:31–56.

Iranzo, S., and G. Peri. 2009. "Schooling Externalities, Technology, and Productivity: Theory and Evidence from U.S." *States, Review of Economics and Statistics* 91: 420–31.

Lucas, R. E. (1988). "On the Mechanics of Economic Development," *Journal of Monetary Economics* 22: 3–42.

Lucas, R. E.B. 2005. *International Migration and Economic Development: Lessons from Low-income Countries*. Cheltenham, UK: Edward Elgar Publishing.

Massey, D. S., J. Arango, G. Hugo, A. Kouaouci, A. Pellegrino, and J. E. Taylor. 1993. "Theories of International Migration: A Review and Appraisal." *Population and Development Review* 19: 431–66.

Mayer, T., and S. Zignago. 2006. "A Note on CEPII's Distances Measures." Explanatory note. Centre d'Etudes Prospectives et d'Informations Internationales, Paris, France.

Mountford, A. 1997. "Can a Brain Drain be Good for Growth in the Source Economy?" *Journal of Development Economics* 53: 287–303.

Müller, T., and S.H.T. Tai. 2010. "Individual Attitudes Towards Migration: A Reexamination of the Evidence." University of Geneva, mimeo, Geneva, Switzerland.

Stark, O., C. Helmenstein, and A. Prskawetz. 1997. "A Brain Gain with a Brain Drain." *Economics Letters* 55: 227–34.

Is Foreign Aid Fungible? Evidence from the Education and Health Sectors

Nicolas Van de Sijpe

This paper adopts a new approach to the issue of foreign aid fungibility. Unlike most existing empirical studies, I employ panel data that contain information on the specific purposes for which aid is given. This approach enables me to link aid that is provided for education and health purposes to recipient public spending in these sectors. In addition, I distinguish between aid flows that are recorded on a recipient's budget and those that are not recorded, and I illustrate how the previous failure to differentiate between on- and off-budget aid produces biased estimates of fungibility. Sector program aid is the measure of on-budget aid, whereas technical cooperation serves as a proxy for off-budget aid. I show that the appropriate treatment of off-budget aid leads to lower fungibility estimates than those reported in many previous studies. Specifically, I find that in both sectors and across a range of specifications, technical cooperation, which is the largest component of total education and health aid, leads to, at most, a small displacement of recipient public expenditures. JEL: E62, F35, H50, O23

The effect of foreign aid on economic growth, poverty, and developmental outcomes may depend heavily on the fiscal response of recipient governments. One

Departmental Lecturer in Development Economics, Department of International Development, and Research Associate, Centre for the Study of African Economies (CSAE), University of Oxford. Address: Queen Elizabeth House, 3 Mansfield Road, Oxford, OX1 3TB, United Kingdom. E-mail: nicolas.vandesijpe@qeh.ox.ac.uk. I am grateful to Sudhir Anand, Channing Arndt, Steve Bond, Maarten Bosker, Paolo de Renzio, Markus Eberhardt, Bernard Gauthier, Benedikt Goderis, Clare Leaver, Anita Ratcliffe, Måns Söderbom, Francis Teal, Dirk Van de Gaer, David Vines, John Wilson, the participants at various conferences and seminars, and especially Chris Adam, Paul Collier, Shanta Devarajan, Jon Temple, Frank Windmeijer, Adrian Wood and three anonymous referees for useful comments and discussions. I am also grateful to Ali Abbas, Nicolas Depetris Chauvin, Ibrahim Levent and Gerd Schwartz for assistance in obtaining data. I would like to thank the people at the OECD Development Assistance Committee (DAC), especially Cécile Sangaré, for patiently answering my questions about their development assistance data and Brian Hammond for comments on the aid data construction. I gratefully acknowledge funding from the Department of Economics, the Oxford Centre for the Analysis of Resource Rich Economies (OxCarre), the Oxford Institute for Global Economic Development (OxIGED) and St. Catherine's College, University of Oxford. This paper presents the research results of the Belgian Programme on Interuniversity Poles of Attraction (contract no. P6/07) initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. A supplemental appendix to this article is available at http://wber.oxfordjournals.org/.

THE WORLD BANK ECONOMIC REVIEW, VOL. 27, NO. 2, pp. 320–356 doi:10.1093/wber/lhs023
Advance Access Publication October 31, 2012

aspect of this fiscal response is the possibility that aid may be fungible (i.e., the net effect of earmarked aid differs from the intended effect).

This paper endeavors to determine the extent to which earmarked education and health aid are fungible. Many studies of foreign aid fungibility are hampered by a lack of comprehensive data pertaining to the intended purpose of aid. I use the OECD's Creditor Reporting System (CRS), which disaggregates aid by sector, to overcome this problem. To cope with the incompleteness of the CRS data, I propose a novel data construction method that begins with the CRS and adds information from other OECD aid databases to provide more complete measures of education and health aid disbursements.

These data also enable me to divide education and health aid into on- and off-budget components. I demonstrate how a failure to adequately deal with off-budget aid (aid that is not recorded in a recipient government's budget) may have biased previous estimates of fungibility. When donor-based measures of aid are employed, a potentially large fraction of this aid is off-budget aid. Hence, even if aid is used in the targeted sector, some of it may not be recorded as the sectoral expenditures of a recipient government. This failure to record some aid reduces the estimated marginal effect of total sectoral aid on government sectoral expenditures and thus leads to an overestimation of the extent of fungibility. Other papers employ aid data that are reported by recipient governments. In this case, the effect of on-budget aid on government expenditures is estimated, and off-budget aid acts as an omitted variable. Hence, the first problem is that we cannot estimate the degree of fungibility of off-budget aid. Moreover, because off- and on-budget aid are likely correlated, the estimated effect of on-budget aid is biased unless the marginal effect of off-budget aid on government spending is zero.

I use sector program (SP) aid as a measure of on-budget aid and technical cooperation (TC) as a proxy for off-budget aid. Fixed effects (FE) results illustrate the need to consider on- and off-budget aid separately. In both sectors, SP aid has an approximately one-to-one correlation with the public sectoral expenditures of recipient countries. For TC, the proxy for off-budget aid, the same result of limited fungibility is found: its coefficient is close to and typically not significantly smaller than zero, indicating that TC does not displace recipients' own public spending in either sector. The result of limited fungibility for TC, which constitutes the bulk of total education and health aid, is robust across a range of specifications. In contrast, although the effect of SP aid is robust in the context of a static panel data model that is estimated with FE, the coefficient of SP aid becomes imprecise and volatile in a dynamic model that is estimated with system GMM because of the lack of variation in SP aid.

This paper follows the example of Feyzioglu, Swaroop, and Zhu (1998) and Devarajan, Rajkumar, and Swaroop (2007), among others, in estimating the degree of fungibility from a panel consisting of a large number of countries. For each country, the maximum time span for which data on both government education/health expenditure and education/health aid disbursements are

available is 14 years. Therefore, I avoid estimating country-specific degrees of fungibility, an approach followed by some researchers in this body of literature (e.g., Pack and Pack, 1990, 1993, 1999). In addition, this paper does not examine the potential consequences of fungibility (for examples of papers that do so, see McGillivray and Morrissey, 2000; Pettersson, 2007a, 2007b; Wagstaff, 2011). Rather, the paper draws attention to a significant weakness of previous studies that do not adequately address the presence of off-budget aid.

The next section illustrates how the inappropriate treatment of off-budget aid may yield biased estimates of the degree of fungibility. Section II briefly explains why aid may not be fungible. Section III discusses the data and the empirical model, and section IV presents the results. Section V concludes the paper.

I. Fungibility and Off-budget Aid

Fungibility occurs when aid is not used for the purpose that is intended by donors (McGillivray and Morrissey, 2004). More precisely, targeted aid is fungible if it is transformed into a pure revenue- or income-augmenting resource that can be spent in any manner in which a recipient government chooses (Khilji and Zampelli, 1994). For instance, earmarked health aid would be fungible if, rather than leading to a one-to-one increase in government health expenditures, this aid were used to finance other types of spending, lower taxes, or reduce the deficit.[1] In this section, I discuss how the presence of off-budget aid may lead to an inaccurate assessment of the degree of fungibility; throughout this section, for the sake of concreteness, I focus on the fungibility of health aid.

First, consider a simple regression of government health spending (*HSP*) on on- and off-budget health aid (*HAIDON* and *HAIDOFF*, respectively):

$$HSP = \beta_0 + \beta_{ON} HAIDON + \beta_{OFF} HAIDOFF + u_1. \tag{1}$$

Off-budget health aid is aid that is not recorded on a recipient government's budget and that arises from the direct provision of goods and services by donors that does not involve channeling resources through the recipient government's budget (e.g., donors building hospitals, training medical personnel, or hiring consultants). In equation (1), we assess the degree of fungibility of health aid via our estimates of β_{ON} and β_{OFF}. On-budget health aid is not fungible if $\widehat{\beta}_{ON}$ is greater than or equal to 1, in which case every dollar of health aid that is channeled through a recipient government's budget increases government health expenditures by at least one dollar. On-budget health aid is

1. Even if every dollar of health aid is spent in the health sector, health aid may still be fungible if the recipient government reduces health expenditures from its own resources. I discuss this situation in greater detail below with respect to the fungibility of off-budget aid.

fungible if $\widehat{\beta}_{ON}$ is smaller than 1, and full fungibility entails that $\widehat{\beta}_{ON}$ is not greater than the marginal effect of unconditional resources R (resources that are not earmarked for any of the expenditure categories: the sum of domestic revenue and net borrowing). A coefficient $\widehat{\beta}_{ON}$ that is significantly larger than 1 would suggest that a recipient government matches on-budget health aid by increasing its own health expenditures.

To determine the degree of fungibility of off-budget health aid, however, we must compare $\widehat{\beta}_{OFF}$ to a different benchmark. Because off-budget health aid is not considered part of a government's health expenditure *HSP* even if there is no fungibility, a lack of fungibility for off-budget health aid occurs when $\widehat{\beta}_{OFF}$ is greater than or equal to 0, not 1. Off-budget health aid is fungible if $\widehat{\beta}_{OFF}$ is negative. For instance, if a donor finances the building of new hospitals with off-budget health aid, then fungibility would occur if the recipient government reacted by building fewer hospitals and reallocating some of its health spending to other sectors. In that case, the off-budget health aid of the donor is at least partly fungible because the total amount of resources devoted to the health sector (the sum of government health spending and off-budget health aid) increases by less than the amount of off-budget health aid.[2] Full fungibility occurs if $\widehat{\beta}_{OFF}$ is not greater than the marginal effect of unconditional resources R minus 1, whereas a significantly positive coefficient for *HAIDOFF* constitutes evidence of matching behavior by recipient governments.

We are now in a position to discuss how previous studies may have produced biased fungibility estimates. Some studies have relied on aid data reported by donors. These data are either collected directly from donors or obtained from databases managed by the OECD's Development Assistance Committee (DAC) (e.g., McGuire 1982; Khilji and Zampelli, 1994; Pettersson, 2007a, 2007b). In this case, an equation of the following form is estimated:

$$HSP = \beta_0 + \beta HAID + u_2 \tag{2}$$

where $HAID = HAIDON + HAIDOFF$ is total health aid, the sum of on- and off-budget health aid. The estimated marginal effect of health aid on recipient government health expenditures, $\widehat{\beta}$, is used to evaluate whether aid is fungible; a $\widehat{\beta}$ value that is close to 1 is evidence of low fungibility, whereas an estimate

2. Implicitly, this test assumes that off-budget aid resources cannot be directly diverted to other purposes because this direct diversion of off-budget aid would not reduce *HSP*. For example, if medicines are supplied by donors as off-budget health aid, then this assumption implies that a recipient government cannot sell these medicines and spend the proceeds in another sector. As a result, the only way for a recipient government to render off-budget health aid fungible is to reduce its own health expenditure, which is tested in equation (1). The exclusion of off-budget aid from budgetary records reflects a lack of exclusive control of the government over these resources; thus, according to its nature, most off-budget aid should fall into this category of aid that cannot directly be diverted to other sectors. Even if this categorization does not apply to all types of off-budget aid, in the empirical application below, I focus on a specific type of off-budget aid, technical cooperation, for which this assumption is plausible.

that is close to 0 leads to the conclusion that health aid is mostly fungible. The OLS estimate of β can be written as a weighted average of the OLS estimates of β_{ON} and β_{OFF} in equation (1) (see, e.g., Lichtenberg, 1990):

$$\widehat{\beta} = \widehat{\beta}_{ON}\frac{\sigma^2_{ON} + \sigma_{ON,OFF}}{\sigma^2} + \widehat{\beta}_{OFF}\frac{\sigma^2_{OFF} + \sigma_{ON,OFF}}{\sigma^2}. \tag{3}$$

The weights depend on the sample variances of on- and off-budget health aid (σ^2_{ON} and σ^2_{OFF}; σ^2 is the variance of total health aid) and the sample covariance between on- and off-budget health aid ($\sigma_{ON,OFF}$).[3] Because off-budget health aid is not counted as part of government health spending even when it is used within the health sector, $\widehat{\beta}_{OFF}$ will be close to zero even if there is no fungibility. More generally, if on- and off-budget health aid are equally fungible, then we observe that $\widehat{\beta}_{OFF} = \widehat{\beta}_{ON} - 1$. As a result, the presence of off-budget aid in the donor-based aid measure lowers the estimated marginal effect of total health aid on health spending and leads to an overestimation of the degree of fungibility. A marginal effect that is smaller than 1 does not necessarily indicate that aid is fungible; such a value could simply indicate that some aid is not recorded on a recipient government's budget. This bias in the assessment of the degree of fungibility is larger if the variance of off-budget health aid is larger than the variance of on-budget health aid.[4]

Other studies have estimated fungibility for a single country using a time series of recipient-based aid data (e.g., Pack and Pack, 1990, 1993; Franco-Rodriguez, Morrissey, and McGillivray, 1998; Feeny, 2007). In this case, because a recipient government's reports of aid, by definition, exclude off-budget aid, only the effect of on-budget aid on government expenditures is estimated:

$$HSP = \beta_0 + \beta_{ON}HAIDON + u_{3.} \tag{4}$$

Hence, the first problem is that we cannot estimate the degree of fungibility of off-budget health aid. Moreover, because off-budget health aid acts as an omitted variable and off- and on-budget health aid are most likely correlated, $\widehat{\beta}_{ON}$ is biased unless the marginal effect of off-budget health aid on health spending is zero. The sign of the bias is ambiguous because it depends on the

3. For simplicity, the exposition focuses on a cross-sectional case without control variables. Later in the paper, I will primarily examine panel data models that include control variables and that use a fixed effects estimator. In these models, the variables in equation (1) and (2) can be understood as the residuals of the variables after the fixed effects and control variables have been partialled out. In that case, in (3), σ^2_{ON}, σ^2_{OFF}, σ^2 and $\sigma_{ON,OFF}$ refer to the variances and covariance of the partialled-out versions of the relevant variables.

4. I am grateful to an anonymous referee for suggesting this framework to discuss the bias that may be caused by off-budget aid.

partial correlation between on- and off-budget health aid, which could be positive or negative.

This section has clarified the criticism of McGillivray and Morrissey (2000, p. 422) who claim that because a large portion of the aid that is reported by donors is not reflected in the public sector accounts of recipients, such aid measures "... are inappropriate for analyzing fungibility." In addition, this section has shown that the use of recipient-reported aid data is also problematic unless separate data exist that can measure off-budget aid such that equation (1) can be estimated rather than equation (4). Off-budget aid is likely to be sizable in many countries and to vary both between and within countries. Thus, the effects of its inappropriate treatment may be important. With regard to aggregate aid, Fagernas and Roberts (2004a) show that OECD DAC figures for Uganda exceed the external financing recorded by the government by substantial margins (in some years, in excess of 10% of GDP). In Zambia, the gap is as wide as 20–40% of GDP in some years (Fagernäs and Roberts, 2004b). In both countries, the amount of off-budget aid varies substantially over time. Thus, for aggregate aid, σ^2_{OFF} in (3) is unlikely to be small relative to σ^2_{ON}. For Senegal, Ouattara (2006) finds that OECD DAC aid during the 1990s was, on average, twice as high as the aid reported by the local Ministry of Finance (12% vs. 6% of GDP, respectively), although his plots appear to suggest that the variation in aggregate aid over time is predominantly driven by on-budget aid.[5]

The correct method of assessing whether earmarked aid is fungible involves separating on- and off-budget sectoral aid and comparing the marginal effect of on-budget aid on recipient sectoral spending to 1 and the marginal effect of off-budget aid to 0. The aim of this paper is to apply this method in the education and health sectors using a newly constructed dataset of disaggregated aid disbursements. Before presenting the empirical analysis, the next section of this article discusses some of the reasons that earmarked aid may not be fungible.

II. Why Aid May Not Be Fungible

As illustrated in a number of papers (e.g., Pack and Pack, 1993; Feyzioglu et al., 1998; McGillivray and Morrissey, 2000), standard microeconomic theory predicts that fungibility arises as the natural response of a rational government to an inflow of earmarked aid. However, several reasons may explain why aid may not be fully fungible. The most compelling reason may be donor conditionality. The earmarking of aid is automatically accompanied by a

5. Other studies report similarly large shares of off-budget aid out of total aid but do not allow us to assess the extent of variation in off-budget aid over time. In Fiji and Vanuatu, 70% of all aid is off-budget aid (Feeny, 2007). In Malawi, approximately 40% is off-budget aid (Fagernäs and Schurich, 2004), and in Liberia, approximately 75% is off-budget aid (Republic of Liberia Ministry of Finance, 2009).

certain type of conditionality: that aid leads to a full increase in expenditures in the targeted sector. If a donor is able to monitor the fiscal policy choices of a recipient government and to enforce conditionality in a credible manner, then fungibility can be reduced (Adam, Andersson, Bigsten, Collier, and O'Connell, 1994).

A lack of information on the part of a recipient government may also reduce the degree of fungibility. McGillivray and Morrissey (2001) argue that even if policymakers in a recipient country intend for earmarked aid to be fully fungible, fungibility may be reduced as a result of errors in the perception of the implementing officials ("aid illusion"). Incomplete information may contribute particularly to a reduction in the fungibility of off-budget aid. If governments in aid-receiving countries are not aware of the extent to which donors directly provide goods and services in a sector via off-budget aid, then they may not realize that the amount of resources spent in the sector is higher than what they consider optimal. As a result, they may neglect to reduce their own expenditures in the sector when they encounter an inflow of off-budget aid.

There is a final reason to expect less than full fungibility for off-budget aid. The presence of off-budget health aid that cannot directly be diverted to other sectors determines a lower bound for the total amount of resources spent in the health sector (the sum of government health expenditures and off-budget health aid). If the government's desired amount of total resources spent in the health sector is exceeded by the amount of non-divertible off-budget health aid, then fungibility is necessarily reduced.[6] This reason becomes more relevant if we think of the government as separately targeting optimal amounts of various types of health goods that cannot easily substitute for one another rather than one aggregate health good. In that context, the non-divertible off-budget health aid that is directed toward one or several of these specific health goods (e.g., hospitals, syringes, health technical cooperation) would be more likely to exceed the government's preferred expenditure for that good, such that the fungibility of earmarked health aid as a whole is decreased (Gramlich, 1977, makes exactly this point in the context of intergovernmental grants).

Thus, the extent to which earmarked aid is fungible must ultimately be determined empirically. The remainder of this paper is devoted to this task.

6. For example, suppose that in the absence of any health aid, a recipient government spends 100 million dollars in the health sector. If a donor provides 200 million dollars of off-budget health aid, then full fungibility would entail that the recipient government reduces its own health expenditures at an approximately one-to-one rate (i.e., the recipient government reduces its health expenditure by 200 million dollars). However, the government cannot implement such a reduction because health expenditure would need to decrease below zero. The most that this government can do is to reduce its health expenditure by 100 million dollars; in this situation, health aid is only partially fungible.

III. Data and Empirical Model

Sectoral Aid Data

Knowledge of the intended purpose of aid is crucial to obtain an accurate estimate of the degree of fungibility. Therefore, the use of sectorally disaggregated aid in this paper constitutes a marked improvement over previous studies that lack complete information on the purposes for which aid is given. Fiscal response models (FRMs) typically focus on the effect of aggregate aid on a recipient's budget and evaluate aid as being fungible if it is diverted away from public investments or developmental expenditures (e.g., Heller, 1975; Franco-Rodriguez et al., 1998; Feeny, 2007).[7] Early fungibility studies (McGuire, 1982, 1987; Khilji and Zampelli, 1991, 1994) distinguish between military and economic aid and evaluate how these types of aid affect public military and non-military expenditures. Other studies (Feyzioglu et al., 1998; Swaroop, Jha, and Rajkumar, 2000; Devarajan et al., 2007) attempt to investigate aid at the sectoral level but are only able to disaggregate concessionary loans; thus, the omission of sectoral grants may influence their results. In this body of literature, Pack and Pack (1990, 1993, 1999) are the only studies that employ a comprehensive sectoral disaggregation of foreign aid by focusing on countries whose recipient governments report both public expenditures and aid received in a disaggregated form.[8]

In addition, several recent studies (Chatterjee, Giuliano, and Kaya, 2007; Pettersson, 2007a, 2007b) have used sectorally disaggregated aid data from the OECD's Creditor Reporting System (CRS), as described in OECD (2002), to study fungibility.[9] The CRS database disaggregates foreign aid according to a number of dimensions, most importantly the sector or purpose of aid, but has two main disadvantages. First, the CRS data are incomplete. Only some of the total disbursements that flow from each donor to each recipient in any given year are reported. Coverage becomes weaker as one examines earlier periods in time. Second, although information pertaining to commitments is available beginning from 1973, disbursement information is available only for the period after 1990. As a result, many existing papers utilize sectoral commitments even when disbursements are the more relevant quantity.

7. Many of the papers in this body of literature disaggregate aid into grants and loans, multilateral and bilateral aid, or by aid modality, but not by sector.

8. The studies that are referenced in this paragraph estimate the degree of fungibility using panel data for either a large (Feyzioglu et al., 1998; Devarajan et al., 2007) or small (Heller, 1975; Feeny, 2007) number of countries, or they report country-specific estimates of fungibility (all other studies referenced in this paragraph).

9. I describe the OECD's aid databases as they were when I began to construct the sectoral aid data (December 2006). Since then, the CRS and DAC Directives have been updated, and the databases have undergone minor changes (see OECD, 2007a, 2007b).

Several studies (e.g., Mavrotas, 2002; Pettersson, 2007a, 2007b) attempt to avoid these problems with the assistance of data from OECD DAC table 2a, as described in OECD (2000a). DAC2a contains *complete* aggregate aid disbursements but does not include sectoral disaggregation. These studies estimate sectoral disbursements for each recipient and each year ($\widehat{d}^{s}_{RY}$) by calculating the share of each sector s in total CRS commitments and then multiplying these shares by aggregate disbursements from DAC2a ($DAC2a^{agg}_{RY}$):[10]

$$\widehat{d}^{s}_{RY} = DAC2a^{agg}_{RY}\left(\frac{CRS^{s,comm}_{RY}}{CRS^{agg,comm}_{RY}}\right) \tag{5}$$

for $s = 1, \ldots, S$. This strategy yields sectoral aid disbursements even for those years in which only commitment information is available in CRS. Moreover, because $DAC2a^{agg}_{RY}$ is complete, it corrects for the incomplete nature of the CRS data in a simple manner.

This method assumes that the sectoral distribution of incomplete CRS commitments is a good guide to the actual distribution of total disbursements across sectors. This assumption may not hold if, for instance, a donor's propensity to report disaggregated aid to the CRS database varies by sector, or if donors that report a good deal of their aid to CRS have different sectoral preferences than donors that largely fail to report disaggregated aid. As a result, equation (5) may yield highly imperfect measures of sectoral disbursements, especially if CRS coverage is low, such that the sectoral distribution of CRS commitments that is used to allocate aggregate DAC2a disbursements across sectors is based on only a small subset of the total aid committed to a recipient.

To address these problems, I first restrict the analysis to the 1990-2004 period, for which CRS disbursement information is available. More importantly, I construct more complete data on earmarked education and health aid disbursements by accounting for additional information available in DAC table 2a and DAC table 5. Because the method is described in detail in the supplemental appendix, available at http://wber.oxfordjournals.org/, I provide only a brief summary here.

I begin with aggregate and sectoral gross CRS disbursements in a recipient-donor-year (RDY) format, labeled CRS^{agg}_{RDY} and CRS^{s}_{RDY} (for $s = 1, \ldots, S$), respectively. For each RDY observation, the amount of aid that is absent from CRS is calculated as the difference between DAC2a and CRS disbursements:

$$RES^{agg}_{RDY} = DAC2a^{agg}_{RDY} - CRS^{agg}_{RDY.} \tag{6}$$

10. *RY* denotes recipient-year, *agg* denotes aggregate aid, and *comm* denotes commitments. No superscript is used for disbursements.

The aim is to allocate this total residual (RES^{agg}_{RDY}) across sectors, thereby generating sectoral residuals that can be added to the CRS sectoral disbursements to compensate for the incomplete nature of the latter.

To achieve this goal, I use data from DAC table 5. DAC5 comprises aggregate aid and its sectoral distribution but organizes information only by donor and not by recipient ($DAC5^{agg}_{DY}$ and $DAC5^{s}_{DY}$, respectively). However, DAC5 has an advantage in that these data contain more complete information than CRS.[11] By converting the CRS data into the same donor-year (DY) format, I can calculate the amount of sectoral aid that is absent from CRS in each DY (RES^{s}_{DY}) for each sector. As a result, for each DY and sector, I can compute the share of the sectoral residual in the total residual:

$$SHRES^{s}_{DY} = \frac{RES^{s}_{DY}}{\sum_{s=1}^{S} RES^{s}_{DY}}. \tag{7}$$

This donor- and year-specific allocation of the total residual across sectors is then applied to the total residual in the original recipient-donor-year format:

$$\widehat{RES}^{s}_{RDY} = SHRES^{s}_{DY} RES^{agg}_{RDY}. \tag{8}$$

This procedure yields sectoral residual variables ($\widehat{RES}^{s}_{RDY}$) that are added to CRS sectoral disbursements to create more complete measures of sectoral aid (labeled $\widetilde{CRS}^{s}_{RDY}$). Summing across donors arranges the sectoral disbursements in the required recipient-year format. For some donors, insufficient information is available in DAC5 to allocate the total residual across sectors; therefore, for some observations, the constructed sectoral aid variables still do not reflect the total amount of aid received. Therefore, as a final step, I scale the sectoral disbursements to ensure that their sum matches aggregate disbursements ($DISB_{RY}$):

$$\widehat{CRS}^{s}_{RY} = DISB_{RY} \left(\frac{\widetilde{CRS}^{s}_{RY}}{\sum_{s=1}^{S} \widetilde{CRS}^{s}_{RY}} \right). \tag{9}$$

Aid disbursements are constructed for the following sectors: education (DAC5 sector code 110), health (120), commodity aid/general program assistance (500), action relating to debt (600), donor administrative costs (910), support to NGOs (920) and other sectors (the sum of all remaining sector codes). In addition, data that partition education and health disbursements into four prefix codes or aid types are constructed: investment projects (IP), sector program (SP) aid, technical cooperation (TC), and other (no mark) (ONM). As I explain below, the prefix

11. The data in DAC5 are a mix of disbursements and commitments. To account for this, I scale the DAC5 data to ensure that the sum of the sectoral aid variables matches the aggregate disbursements from DAC2a for every donor-year.

codes are useful because, to some extent, they allow for the separation of on- and off-budget aid flows and thus enable a test of fungibility that is consistent with the framework that is discussed in section I.

This data construction method takes into account that donors that report only a small portion of their aid to CRS might allocate aid across sectors differently than donors that report a larger portion of their aid. Similarly, this method considers that, for a given donor, the sectoral allocation of unreported aid may differ from that of the reported portion. The method ensures that the distribution of aggregate aid across sectors for each donor-year closely follows the sectoral allocation in DAC5, which contains complete disaggregated aid data. Subsequently, the main assumption is that the donor-year-specific sectoral allocation of the total residual applies equally to each recipient that receives aid from the donor in that year that is not accounted for in CRS.

In the final step of the data construction, I scale the sectoral aid variables such that their sum matches aggregate aid received, similar to the scaling performed in previous studies (recall equation (5)). However, because the sectoral disbursements prior to scaling are based on more extensive information than in previous studies, these disbursements are more likely to provide a useful guide to the true sectoral allocation of total disbursements. Therefore, the scaling should be less problematic. On average, the constructed disbursements before scaling constitute more than 76% of the complete aggregate disbursements, whereas this value for CRS disbursements is only 31.9% (see table S1.1 and the surrounding text in the supplemental appendix). For the majority of observations, the scaling that is performed in the final step is limited in magnitude and is substantially smaller than if the CRS sectoral disbursements were scaled without any adjustment. For instance, for more than three-quarters of the observations, the CRS disbursements constitute less than half of the aggregate aid. The constructed sectoral disbursements constitute less than half of the aggregate aid for fewer than 10% of observations. Thus, the sectoral allocation of the aid data before scaling is more likely to provide a reasonable reflection of the actual sectoral allocation that one would find if the data were complete. The failure to scale the sectoral disbursements would increase the risk of underestimating the amount of aid received.[12]

12. Since the construction of the data for this paper, two new disaggregated aid datasets have become available. Ravishankar, Gubbins, Cooley, Leach-Kemon, Michaud, Jamison, and Murray (2009) construct data on health aid by estimating disbursements on the basis of the less incomplete CRS commitments and by adding data from separate reports for a number of NGOs and multilateral and private donors. These data are used by Lu et al. (2010) to estimate the fungibility of health aid. One disadvantage is that a large portion of the data cannot be allocated by recipient country. Lu et al. (2010, p. 1379) state that only 21% of all health aid in 1995 can be traced to recipient countries, and 30% of this aid can be traced to recipient countries in 2006. In addition, it is not immediately clear how one would further divide health aid into on- and off-budget components in these data. A second recent dataset, AidData (http://www.aiddata.org), attempts to construct a more complete disaggregation of aggregate aid into all of its constituent parts according to a number of dimensions but focuses almost exclusively on commitments.

Empirical Model and Other Data

First, I consider models that do not distinguish between on- and off-budget sectoral aid:

$$SSP_{it} = \beta SAID_{it} + \gamma \boldsymbol{A}_{it} + \boldsymbol{\delta X}_{it} + \lambda_t + \eta_i + \varepsilon_{it} \tag{10}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. SSP_{it} denotes recipient government spending on education or health, whereas $SAID_{it}$ are disbursements that are earmarked for the same sector. $\boldsymbol{A}_{it}$ and $\boldsymbol{X}_{it}$ contain other aid variables and control variables that are described below. λ_t is a set of year dummies, η_i captures country-specific time-invariant effects, and ε_{it} is the transient error. Aid and spending variables are expressed as percentages of GDP.[13] High-income countries (2005 GNI per capita of 10726 US$ or more, following World Bank, 2006c) are eliminated from the sample. I begin with a static panel data model similar to that employed by cross-country fungibility studies that utilize information on the intended purpose of aid, particularly Feyzioglu et al. (1998) and Devarajan et al. (2007). This allows for an easier comparison of the results. Later in the paper, I briefly discuss the results from more general models that allow for some dynamics.

I focus on education and health for a number of reasons. First, education and health play a prominent role in the Millennium Development Goals (MDGs). In addition to their importance in the first goal, which involves eradicating extreme poverty and hunger, several other goals explicitly establish targets related to education and health. This suggests that donors have preferences for education and health spending and should be concerned about the extent of fungibility in these sectors. Second, as partially evidenced by their prominent role in the MDGs, there is a widespread belief that better education and health have immediate consequences for human welfare and play important roles in spurring development and alleviating poverty. This belief suggests that the fungibility of aid that is directed toward these sectors may be relevant for the welfare of the population in recipient countries and may influence the overall effectiveness of aid. Third, these areas are rather clearly defined areas of spending, which should increase the definitional overlap between sectoral aid and sectoral spending.

Public education and health expenditure are staff estimates from the IMF's Fiscal Affairs Department (FAD) and are available for the period prior to 2003.[14] The data are obtained from IMF country documents and have been verified and reconciled by country economists (Baqir, 2002). The main

13. Current US$ GDP from World Bank (2006c) is used to express sectoral aid disbursements as a percentage of GDP.

14. These data are not publicly available, although they have been used in a variety of publications (e.g., Gupta, Clements, and Tiongson, 1998; Baqir, 2002). I am grateful to Gerd Schwartz for sharing these data and to Ali Abbas for assistance in obtaining them.

advantage over other datasets (International Monetary Fund, 2006; World Bank, 2006a, 2006c) is the significantly improved coverage. Moreover, although the level of government (central or general, in which the latter also includes state and local government) spending differs across countries, it is fixed over time. Thus, average differences in government expenditure shares in GDP between countries that result from differences in the government level on which reporting is based can be absorbed by fixed effects (Baqir, 2002).[15]

A_{it} includes commodity aid/general program assistance (henceforth called general aid) and support to NGOs. If targeted toward education and health, support to NGOs may have an effect on a recipient government's spending in these sectors (Lu, Schneider, Gubbins, Leach-Kemon, Jamison, and Murray, 2010, find that health aid to NGOs increases the health spending of recipient governments from their own resources). General aid may partially finance education and health spending or, if linked to structural adjustment programs, may be conditional on lowering public spending. The final variable in A_{it} is other non-education or non-health aid. In the equation for public education spending, other non-education aid includes health aid, and vice versa.

Another aid variable, action relating to debt, is not included in the regression model. Debt relief may be important, but it is not adequately captured by actions relating to debt, including debt forgiveness, debt rescheduling, and other actions (such as service payments to third parties, debt conversions, and debt buybacks) (OECD, 2000b). The debt forgiveness component measures the face value of total debt that is forgiven in a year rather than its present value (PV). Because the average concessionality of debt varies strongly across countries, this may be misleading (Depetris Chauvin and Kraay, 2005). For most types of debt rescheduling, the reduction in debt service in a given year as a result of present and past rescheduling is recorded. Again, this fails to capture the PV of current and future reductions in debt service as a result of debt rescheduling in the current year.[16] For these reasons, I omit action relating to debt as a regressor and instead control for the PV of public and publicly guaranteed long-term external debt as well as public and publicly guaranteed long-term external debt service. These variables should capture most of the effects of debt relief on social spending. Less debt service means that more resources are available to spend on other purposes, whereas a lower stock of debt means that the intertemporal budget constraint is loosened, which may increase the government's appetite for spending. The PV of debt is obtained from Dikhanov

15. For Fiji, the observation in 1998 for both sectors is approximately ten times smaller than that in the surrounding years, most likely due to a typographical error. For instance, public education expenditures account for 0.572% of the GDP in 1998, whereas these expenditures range from 5.19% to 6.37% of the GDP in all other years from 1993 to 2002. Hence, I change this value to 5.72. Similarly, I adjust the public health expenditure value for 1998 from 0.253% to 2.53% of GDP.

16. Only for Paris Club concessional debt reorganizations is the net present value reduction in debt achieved by the current rescheduling recorded (OECD, 2000b, p. 17).

(2004), which is updated through 2004.[17] The source for debt service is the Global Development Finance database (World Bank, 2006b). Again, I use current US$ GDP from (World Bank, 2006c) to express both variables as percentages of GDP.

Other control variables that are included in $\boldsymbol{X}_{it}$ are real GDP per capita (thousands of constant 2000 international dollars) and its growth rate, urbanization (urban population, % of total) and trade (% of GDP) (all from World Bank, 2006c). Because aid that is expressed as a % of GDP is likely to be correlated with GDP (per capita), excluding the latter may induce a spurious relationship between aid and expenditure. Growth is included to capture the reaction of expenditure to short-term shocks in GDP per capita. If government education and health expenditure do not immediately adjust to a higher (lower) level in the event of a positive (negative) growth shock, then a negative coefficient is expected. The effect of trade is a priori ambiguous (e.g., Rodrik, 1998). Greater openness may erode a government's capacity to finance expenditure as tax bases become more mobile. Moreover, tariff reductions may increase trade openness while starving the government of revenue, which again suggests a negative association between trade and public education or health expenditure. However, openness to trade may also increase the demand for social spending to insure against increased external risk and to redistribute gains from trade, and public education and health expenditure may play a role in these effects. Urbanization may also have a positive or negative effect. Some services should be easier to administer in a more urbanized society (Hepp, 2005), and urbanization may create more opportunities for economies of scale. However, lower transportation costs and easier lobbying for government services in urbanized societies may increase the demand for education and health services (Hepp, 2005; Baqir, 2002). For health spending, the risk of contagion and pollution may be higher in cities (Gerdtham and Jönsson, 2000).

Table 1 shows summary statistics for the education and health regression samples. Education aid constitutes approximately 28% of public spending in the education sector, whereas health aid accounts for approximately 22% of public health spending. Slightly less than one-fifth of aid (excluding actions relating to debt and donor administrative costs) is targeted toward education or health.

Hypothesis Tests for No Fungibility and Full Fungibility

As discussed in section I, the presence of off-budget aid in the donor-based measure of sectoral aid ($SAID_{it}$) decreases the estimate of β, thereby overstating the true degree of fungibility. For a correct assessment of fungibility, it is necessary to distinguish between on- and off-budget sectoral aid. Consequently, I also estimate models that partition education and health disbursements into

17. I am grateful to Ibrahim Levent for sending me the updated data (received December 2006) and the Dikhanov paper.

TABLE 1. Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.
Education sector: 1082 observations (108 countries, annual data for 1990–2003)				
Public education expenditure	4.02	1.92	0.38	13.61
Education aid	1.13	1.45	0.01	14.19
Education IP	0.13	0.23	0	3.6
Education SP	0.04	0.09	0	0.95
Education TC	0.81	1.1	0	10.85
Education ONM	0.16	0.34	0	5.83
General aid	1.2	1.92	0	22.78
Support to NGOs	0.13	0.24	0	3.02
Other non-education aid	5.84	6.78	0.01	62.84
Real GDP per capita	3.63	2.98	0.47	17.96
Real GDP per capita growth	1.6	5.46	−30.28	49.86
Urbanization	42.4	20.36	6.3	91.56
Trade	78.11	41.06	10.83	280.36
PV debt	52.15	60.07	0.09	892.12
Public debt service	4.02	3.47	0	35.24
Health sector: 1087 observations (108 countries, annual data for 1990–2003)				
Public health expenditure	1.96	1.25	0.17	7.44
Health aid	0.44	0.54	0	3.63
Health IP	0.11	0.18	0	1.69
Health SP	0.05	0.1	0	1.75
Health TC	0.18	0.23	0	1.91
Health ONM	0.1	0.18	0	1.46
General aid	1.21	1.97	0	22.78
Support to NGOs	0.13	0.24	0	3.02
Other non-health aid	6.56	7.5	0.02	66.11
Real GDP per capita	3.64	2.98	0.47	17.96
Real GDP per capita growth	1.58	5.4	−30.28	28.5
Urbanization	42.24	20.4	6.3	91.56
Trade	77.8	41.2	10.83	280.36
PV debt	51.12	59.14	0.09	892.12
Public debt service	3.91	3.24	0	35.24

Note: All variables as % of GDP except real GDP per capita (thousands of constant 2000 international dollars) and its growth rate and urbanization (urban population, % of total).

Source: Author's analysis based on data described in the text.

the four prefix codes:

$$\begin{aligned} SSP_{it} = \; & \beta_{IP} SAIDIP_{it} + \beta_{SP} SAIDSP_{it} + \beta_{TC} SAIDTC_{it} \\ & + \beta_{ONM} SAIDONM_{it} + \gamma \boldsymbol{A}_{it} + \boldsymbol{\delta X}_{it} + \lambda_t + \eta_i + \varepsilon_{it} \end{aligned} \tag{11}$$

where *IP* represents investment projects, *SP* denotes sector program aid, *TC* represents technical cooperation, and *ONM* denotes other (no mark) aid.

SP aid should primarily be on-budget aid because, by definition, program aid involves a government-to-government transfer of resources. In contrast, TC is a good proxy for off-budget aid. The costs of providing training and

scholarships in donor countries, remunerating experts and consultants, and financing equipment and administrative costs associated with TC primarily involve direct payments from donor governments rather than transfers of money to recipient governments. In fact, Sundberg and Gelb (2006) argue that many aspects of TC, such as finance for training programs, analytical reports and expert advice, involve resources that never even leave donor countries. For the seven countries that they study, IDD and Associates (2006, p. 23 in annex B) indicate that off-budget aid is explained, among other things, by "aid in kind e.g. TA [technical assistance] and other aid where expenditure is undertaken directly by the donor." Similarly, Fagernäs and Roberts (2004b) argue that technical assistance involves donors making direct payments that are not reflected in budget documents, and Feeny (2007, p. 442) states that "the salaries of external consultants will not enter public sector accounts." Feeny argues that a larger share of aid is off-budget in Fiji and Vanuatu compared with Papua New Guinea and the Solomon Islands because the former two countries receive a large proportion of their aid in the form of technical assistance. In addition, Fagernäs and Roberts (2004a) attribute discrepancies between donor and recipient reports of aid in Uganda at least partially to the omission of TC from the budgets of recipient governments. Johnson and Martin (2005, p. 6) conclude that "HIPCs see direct payments by donors to foreign suppliers as highly problematic, as they are often not informed of the actual disbursements. This is especially true for technical assistance provided by expatriate experts, who are hired and paid by the donor." Baser and Morgan (2001) find that TC is off-budget in the six African countries that they investigate. Drawing from the experiences of a much larger group of countries, OECD (2008, p. 59) notes that "technical co-operation expenditures are described as a particular problem in recording aid on budget." Mokoro (2008), which is a detailed study of the role of aid in the budget process based on both an extensive literature review and case studies of ten Sub-Saharan African countries, identifies a clear hierarchy in the extent to which different aid modalities are disbursed via the treasuries of recipient governments and are captured in their accounts: most likely for general budget support and program aid, much less likely for project aid, and even less likely for technical assistance.

The summary statistics in table 1 suggest that education aid is more than 70% TC, whereas approximately 40% of aid is TC in the health sector. This dominant role of TC in health aid and, especially, education aid is confirmed in the CRS directives (OECD, 2002, p. 26). Average SP aid is small and reflects that for many country-years, education and health SP aid are nearly zero. Particularly in the education sector, the variance in TC is large compared with that of the other sectoral aid modalities, which further reinforces the notion that the bias created by the failure to adequately address off-budget aid may be substantial (recall equation (3)).

The extent to which IP and ONM aid are reported in government budgets is more uncertain. Thus, the estimates of β_{IP} and β_{ONM} are less informative for

TABLE 2. Null Hypotheses for No and Full Fungibility with On- and Off-budget Aid

	No fungibility	Full fungibility
Theoretical null hypothesis:		
Aid on-budget (SP)	$\beta_{SP} \geq 1$	$\beta_{SP} \leq \frac{\partial SSP_{it}}{\partial R_{it}}$
Aid off-budget (TC)	$\beta_{TC} \geq 0$	$\beta_{TC} \leq \frac{\partial SSP_{it}}{\partial R_{it}} - 1$
Implemented null hypothesis:		
Aid on-budget (SP)	$\beta_{SP} \geq 1$	$\beta_{SP} \leq 0$
Aid off-budget (TC)	$\beta_{TC} \geq 0$	$\beta_{TC} \leq -1$

gauging the degree of fungibility.[18] However, using $SAIDSP_{it}$ and $SAIDTC_{it}$ as measures of on- and off-budget sectoral aid, respectively, it is possible to test the null hypothesis of no fungibility and the null of full fungibility in a manner consistent with the analysis in section I, as shown in table 2.

The full fungibility tests require knowledge of the marginal effect of unconditional resources R (typically measured as government expenditure net of aid), which may be obtained by following the two-stage procedure outlined in Devarajan et al. (2007).[19] Nevertheless, the data that I received from the IMF's FAD do not contain total expenditures, revenue or borrowing. Because data availability for these variables in other databases is significantly more limited, a large fraction of the sample would be lost by following this procedure. Instead, I set $\partial SSP_{it}/\partial R_{it} = 0$, such that the implemented tests become those shown in the bottom half of table 2.

In practice, $\partial SSP_{it}/\partial R_{it}$ should be close to zero in both sectors. Unless there is a substantial break in policy, the marginal effect of R should be close to the average share of unconditional resources that are spent in the education and

18. Mokoro (2008) expressly warns against the assumption that aid projects are always off-budget (p. 7) but suggests that the degree to which these projects are captured in budgets is low. (See, e.g., p. 23: "levels of aid on budget are strongly driven by budget support aid (which, by definition, is on budget). In many cases, off budget proportions for other aid modalities still remain very high" and p. 52: "however, budget support has limits, and project aid has been growing. The problems associated with poorly integrated project aid still loom large. The bigger challenge, therefore, is to bring project aid on budget.") In section IV, I discuss how we can gain insight into the degree of fungibility of IP and ONM aid despite the greater uncertainty regarding the extent to which these types of aid are on- or off-budget.

19. As explained in Devarajan et al. (2007), unconditional resources R (or their component parts, domestic revenue and net borrowing) should be excluded from the estimated equation to ensure that the full effect of earmarked aid on sectoral spending is captured. For instance, if sectoral aid reduces tax revenue but the latter is held fixed, then the effect of aid on spending may be overestimated. This two-step procedure entails the inclusion of the residual from a regression of R on the right-side variables in equation (11) as an explanatory variable in the model. Because this residual is, by construction, orthogonal to the other right-side variables, its inclusion does not alter the sectoral aid coefficients, which capture the full effect of earmarked aid. However, its inclusion facilitates the estimation of $\partial SSP_{it}/\partial R_{it}$.

health sectors. As an approximation, if I proxy this share by the share of public education and health expenditure in total government expenditure, then for government expenditure in the range of 20% to 30% of GDP, the figures in table 1 suggest a marginal effect of unconditional resources of approximately 0.13–0.2 for education expenditure and 0.07-0.1 for health expenditure. Devarajan et al. (2007) estimate the effect of unconditional resources on public education (health) spending to be 0.12 (0.04). Feyzioglu et al. (1998) find even smaller effects of 0.08 (0.02) for education (health) expenditure. Therefore, setting $\partial SSP_{it}/\partial R_{it} = 0$ is unlikely to have a significant influence on the conclusions that are drawn from the estimated coefficients and the full fungibility tests, although the probability of rejecting the null hypothesis of full fungibility may be slightly increased.

IV. Results

Table 3 presents the results of the OLS and fixed effects (FE) estimations of equation (11), with total donor-reported education or health aid as the main regressor of interest. Therefore, the hypothesis tests for no fungibility and full fungibility in this table are based on the assumption that education and health aid are completely on-budget. All reported standard errors are robust to heteroskedasticity and are clustered at the country level, thereby allowing for serial correlation in the error term (Arellano, 1987; Bertrand, Duflo, and Mullainathan, 2004).

In both the OLS and FE estimations, public education expenditure has no discernible correlation with education aid, and the null hypothesis of no fungibility is strongly rejected. By contrast, public health expenditure is positively correlated with health aid, and this effect is estimated precisely enough to reject the null hypothesis of full fungibility and the null hypothesis of no fungibility. However, the size of the FE coefficient of health aid is small: an increase in health aid of 1% of GDP is associated with an increase in public health expenditure of only 0.26% of GDP. On the basis of this result, one would still conclude that health aid is mostly fungible.

The results in table 3 are likely to overestimate the extent of fungibility because the presence of off-budget aid decreases the estimated effect of sectoral aid on public sectoral expenditure. Table 4 presents the results from the estimation of equation (11), in which sectoral aid is further partitioned into four prefix codes. This partitioning enables the implementation of the more appropriate fungibility tests described in table 2, using SP aid as a measure of on-budget aid and TC as a proxy for off-budget aid.

The further disaggregation of sectoral aid markedly changes the results. In both sectors, the marginal effect of SP aid in the FE model is close to 1; this result suggests that the bulk of SP aid is used in the intended sector. Full fungibility can be rejected, but the null hypothesis of no fungibility cannot be rejected. The effect of TC is close to zero in both sectors, and the null hypothesis of

TABLE 3. Total Education and Health Aid

	Public education exp.		Public health exp.	
	OLS	FE	OLS	FE
Education aid	0.047	0.0042		
	(0.082)	(0.068)		
Health aid			0.47***	0.26**
			(0.18)	(0.12)
General aid	−0.0032	0.032	0.016	0.0037
	(0.053)	(0.029)	(0.030)	(0.019)
Support to NGOs	−0.41	−0.38*	−0.13	−0.18***
	(0.33)	(0.21)	(0.17)	(0.091)
Other non-education aid	0.0026	−0.0041		
	(0.022)	(0.018)		
Other non-health aid			0.0084	−0.012
			(0.017)	(0.012)
GDP per capita	0.085	0.26*	0.17***	0.14*
	(0.059)	(0.14)	(0.048)	(0.085)
GDP per capita growth	−0.049***	−0.028***	−0.025***	−0.020***
	(0.016)	(0.0093)	(0.012)	(0.0074)
Urbanisation	−0.010	0.080	0.0026	0.056*
	(0.0083)	(0.056)	(0.0053)	(0.033)
Trade	0.015***	−0.014***	0.010***	−0.0075*
	(0.0038)	(0.0068)	(0.0031)	(0.0041)
PV debt	−0.0038	−0.0025	0.00025	0.000032
	(0.0035)	(0.0017)	(0.0022)	(0.00056)
Public debt service	0.050	−0.063***	−0.040**	−0.024**
	(0.062)	(0.022)	(0.019)	(0.012)
R^2	0.178	0.207	0.294	0.171
Hausman		0.000		0.000
$\beta \leq 0$	0.285	0.475	0.005	0.019
$\beta \geq 1$	0.000	0.000	0.002	0.000
Countries	108	108	108	108
Observations	1082	1082	1087	1087

Note: OLS and fixed effects (FE) results, annual data, 1990−2003. All regressions include time dummies, coefficients not reported. Heteroskedasticity-robust standard errors, clustered by country, in brackets. *, **, and *** denote significance at 10, 5 and 1%, respectively. In the case of FE estimation, R^2 refers to the within R^2. Hausman shows the p-value of a generalized Hausman test of the null hypothesis that η_i is uncorrelated with the regressors. $\beta \leq 0$ ($\beta \geq 1$) is the p-value for the test of full (no) fungibility for total sectoral aid.

Source: Author's analysis based on data described in the text.

full fungibility is strongly rejected. The hypothesis of no fungibility cannot be rejected; thus, there is no evidence that sectoral TC displaces a recipient government's own expenditure in either sector. The TC effect is similar in OLS, whereas the coefficients of SP aid become larger but are also estimated less precisely. The larger SP aid coefficients in OLS may indicate that time-invariant unobservables are positively correlated with both SP aid and sectoral public expenditures. In the FE estimation, the coefficients are identified from the within-country variation in the data, which reduces the problem of omitted variables

TABLE 4. Disaggregated Education and Health Aid

	Public education exp.		Public health exp.	
	OLS	FE	OLS	FE
Education IP	0.091	0.12		
	(0.25)	(0.12)		
Education SP	2.53*	1.21**		
	(1.35)	(0.55)		
Education TC	0.032	−0.0070		
	(0.10)	(0.082)		
Education ONM	0.14	0.021		
	(0.21)	(0.19)		
Health IP			0.40	0.20
			(0.34)	(0.21)
Health SP			1.19*	0.84***
			(0.60)	(0.31)
Health TC			−0.12	0.0067
			(0.35)	(0.32)
Health ONM			0.74**	0.41*
			(0.36)	(0.23)
General aid	−0.0012	0.031	0.023	0.0055
	(0.051)	(0.029)	(0.031)	(0.019)
Support to NGOs	−0.56*	−0.39**	−0.15	−0.16
	(0.30)	(0.19)	(0.16)	(0.11)
Other non−education aid	−0.0081	−0.0055		
	(0.022)	(0.018)		
Other non−health aid			0.014	−0.013
			(0.017)	(0.011)
GDP per capita	0.084	0.29*	0.17***	0.15*
	(0.060)	(0.15)	(0.048)	(0.085)
GDP per capita growth	−0.051***	−0.029***	−0.028**	−0.021***
	(0.015)	(0.0091)	(0.011)	(0.0072)
Urbanisation	−0.0089	0.085	0.0026	0.055*
	(0.0081)	(0.055)	(0.0053)	(0.031)
Trade	0.016***	−0.013**	0.011***	−0.0071*
	(0.0039)	(0.0067)	(0.0032)	(0.0040)
PV debt	−0.0040	−0.0027*	−0.000074	−0.000092
	(0.0034)	(0.0016)	(0.0021)	(0.00059)
Public debt service	0.052	−0.065***	−0.039**	−0.022*
	(0.062)	(0.021)	(0.019)	(0.011)
R^2	0.187	0.215	0.302	0.183
Hausman		0.000		0.000
$\beta_{SP} \leq 0$	0.032	0.015	0.026	0.004
$\beta_{SP} \geq 1$	0.870	0.645	0.621	0.307
$\beta_{TC} \leq -1$	0.000	0.000	0.006	0.001
$\beta_{TC} \geq 0$	0.621	0.466	0.363	0.508
Countries	108	108	108	108

(Continued)

TABLE 4. Continued

	Public education exp.		Public health exp.	
	OLS	FE	OLS	FE
Observations	1082	1082	1087	1087

Note: OLS and fixed effects (FE) results, annual data, 1990-2003. All regressions include time dummies, coefficients not reported. Heteroskedasticity-robust standard errors, clustered by country, in brackets. *, **, and *** denote significance at 10, 5 and 1%, respectively. In the case of FE estimation, R^2 refers to the within R^2. Hausman shows the p-value of a generalized Hausman test of the null hypothesis that η_i is uncorrelated with the regressors. $\beta_{SP} \leq 0$ ($\beta_{SP} \geq 1$) and $\beta_{TC} \leq -1$ ($\beta_{TC} \geq 0$) are p-values for the test of full (no) fungibility for sector program aid and technical cooperation, respectively.

Source: Author's analysis based on data described in the text.

in instances in which such variables do not change substantially over time. For the FE results in tables 3 and 4, a generalized Hausman test that allows for heteroskedasticity and serial correlation is reported (Arellano, 1993; Wooldridge, 2002, pp. 290–291).[20] The null hypothesis that η_i is uncorrelated with the regressors is always rejected; this result suggests that FE should be preferred over random effects. Growth consistently has a negative effect, which suggests that education and health expenditures do not immediately adjust to a higher (lower) level in the event of a positive (negative) short-term shock to GDP per capita (Dreher, 2006, obtains a similar result for total and social expenditures in OECD countries).

As a robustness test, I obtain qualitatively similar FE results with aid variables that are constructed by scaling up sectoral CRS disbursements to ensure that their sum matches the aggregate DAC2a disbursements (equation (5) but applied to CRS disbursements rather than commitments). The main change is that for some aid variables, the estimated coefficients are closer to zero and/or estimated less precisely, which is consistent with greater measurement error in the aid data that are constructed using this short-cut method.[21]

Table 4 illustrates that a failure to properly address the presence of off-budget aid may yield misleading conclusions. After on- and off-budget aid are separated and their effects are assessed against appropriate benchmarks, the FE results suggest that there is little if any fungibility. This conclusion is robust to a large number of specification changes. I replace the PV of debt with a non-PV measure of long-term external public and publicly guaranteed debt expressed as a percentage of GDP (from World Bank, 2006b). I also add to the

20. This test is performed in Stata using the xtoverid command (Schaffer and Stillman, 2006).

21. For instance, the coefficient on education SP aid is almost halved, to 0.64, and full fungibility is therefore rejected less strongly. The coefficient on health aid is reduced to 0.07, whereas in the disaggregated model, the coefficients on health IP and health ONM are much closer to zero. In both sectors, the coefficient on support to NGOs is estimated less precisely and/or substantially reduced in magnitude. The only exception is that the coefficient on health SP aid nearly doubles with the short-cut method (from 0.84 to 1.61), but its standard error rises commensurably.

model, in turn, two different measures of the PV of debt relief constructed by Depetris Chauvin and Kraay (2005).[22] Because debt relief is often linked to higher social expenditure, one might expect it to have a larger positive effect on public education and health expenditure than the effect achieved by a reduction in debt or debt service that arises through means other than debt relief. If this effect is indeed larger, then we would expect a positive effect of debt relief even after controlling for the level of debt and debt service. However. I do not find evidence of this effect. Even without controlling for debt and debt service, I find no effect of the PV of debt relief. I further include GDP per capita in log form rather than in thousands of dollars. I add (one at a time) control variables for female labor force participation or the birth rate (both from World Bank, 2006c), measures of corruption, the rule of law and bureaucratic quality from the International Country Risk Guide (ICRG) (The Political Risk Services Group, 2008), the sum of these three ICRG variables (as a general measure of institutional quality), and measures of democracy obtained from Polity IV (Marshall and Jaggers, 2007). Feyzioglu et al. (1998) control for the share of agriculture in GDP rather than urbanization. Therefore, I replace urbanization with the share of agriculture in GDP (from WDI) or add the share of agriculture in GDP alongside urbanization. Many papers also control for the size and composition of the population when explaining variation in public expenditures (e.g., Baqir, 2002; Rodrik, 1998). As a result, I consider models that add the percentage of the population under 15 and/or the percentage of the population over 65 to the model, the age dependency ratio (dependents to working-age population) or the log of population (all from WDI). Finally, Feyzioglu et al. (1998) control for lagged infant mortality, whereas Devarajan et al. (2007) control for lagged secondary and primary school enrollment in the education expenditure equation and for lagged infant mortality in the health expenditure equation. A possible concern is that such variables may be more fruitfully viewed as outcomes than as determinants of public education and health expenditures. Nonetheless, I include either the current value or the once-lagged value of primary gross enrollment, secondary gross enrollment, infant mortality or under-five mortality (mortality data from WDI and enrollment data from Edstats). In all cases, the results are qualitatively unchanged. The only exception is that when the ICRG measures are added, the coefficient of health TC decreases to approximately -0.25, and I can reject the null hypothesis of no fungibility, implying partial (but low) fungibility of health TC.[23]

Influential Observations

Especially given the limited variation in education and health SP aid and, to a lesser extent, TC, one concern may be that the effects of these variables are

22. I am grateful to Nicolas Depetris Chauvin for sharing these data.

23. In addition, no clear evidence is found to suggest that the degree of fungibility depends on the quality of institutions.

TABLE 5. Disaggregated Education and Health Aid, Marginal Effects of the Log-linear Model

	Public education exp.	Public health exp.
$\widehat{\beta}_{SP}$	1.342	1.092
$\beta_{SP} \leq 0$	0.005	0.006
$\beta_{SP} \geq 1$	0.750	0.585
$\widehat{\beta}_{TC}$	0.0522	0.0602
$\beta_{TC} \leq -1$	0.000	0.000
$\beta_{TC} \geq 0$	0.632	0.591

Note: $\widehat{\beta}_{SP}$ and $\widehat{\beta}_{TC}$ are marginal effects, calculated at the sample means, based on the fixed effects estimation of equation (11) in log-linear form. Annual data, 1990–2003. All regressions include time dummies and the standard set of control variables (coefficients not reported) and are estimated with heteroskedasticity-robust standard errors, clustered by country. $\beta_{SP} \leq 0$ ($\beta_{SP} \geq 1$) and $\beta_{TC} \leq -1$ ($\beta_{TC} \geq 0$) are p-values for the test of full (no) fungibility for sector program aid and technical cooperation, respectively.

Source: Author's analysis based on data described in the text.

driven by a small number of observations. Although the inclusion of additional control variables generally does not change the conclusions, the point estimates on the variables of interest shift by a relatively large amount in several instances, especially when the inclusion of an additional variable leads to a large decrease in sample size. Such a shift always results from a change in the sample composition and not because the additional control variable eliminates some of the explanatory power of sectoral SP aid or TC.[24]

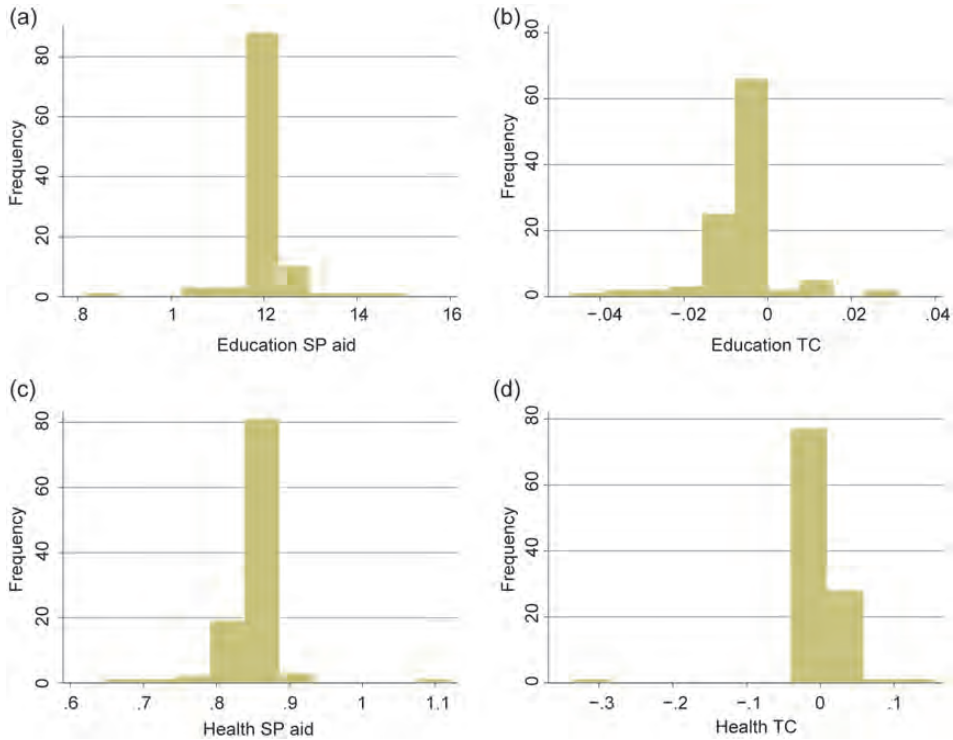
As a first attempt to evaluate the sensitivity of the results to outliers, I re-estimate equation (11) in log-linear form. Taking the natural logarithm of all variables compresses the upper tail and is thus likely to reduce the influence of observations with larger values of education and health SP aid or TC on the estimated regression line.[25] Table 5 displays the marginal effects for SP aid and TC calculated at the sample means (the full results are available upon request). The results are similar to those obtained in the linear model. In both sectors, the effect of TC is close to zero, and the effect of SP aid on public expenditure is close to 1. Full fungibility is rejected across the board, but the null hypothesis of no fungibility cannot be rejected in any of the cases.

As a more direct and arguably superior approach to determine the effects of influential observations, I re-estimate equation (11) by eliminating one country at a time. Figure 1 shows the resulting distribution of the estimated SP aid and TC coefficients. The marginal effect of TC is more stable than that of SP aid in both sectors, which is consistent with the more limited variation in SP aid. A small number of countries induce fairly large changes in the effect of SP aid.

24. The most extreme deviation occurs when the birth rate is added: the sample size in the health model decreases to 612, and the effect of health SP aid in the FE model rises to 1.34.

25. To address zero values in the public expenditure, aid and debt variables, I add 1 before taking the log. Because GDP per capita growth can be negative, I include this variable without taking its log.

FIGURE 1. Distribution of Coefficients When Dropping One Country at a Time



Source: Author's analysis based on data described in the text.

For instance, when Lesotho is eliminated, the effect of education SP aid decreases to 0.82. When Tonga is excluded, this effect increases to 1.51. In contrast, the distribution of the estimated coefficient of education TC has a substantially smaller range. For health TC, two countries have a sizable influence on the estimated coefficient when they are omitted from the sample, but the remainder of the distribution is substantially narrower.[26]

To examine how sensitive the results are to the removal of countries that appear to exert an undue influence on the coefficients of interest, I omit countries for which the absolute value of the $DFBETA_i$ influence statistic for SP aid or TC exceeds the size-adjusted cut-off value of $2/\sqrt{N}$ (in this case, N is the number of countries) proposed by Belsley, Kuh, and Welsch (1980).[27] This

26. Without Eritrea, the estimated effect of education TC becomes -0.33. Without Guinea-Bissau, the effect is 0.16.

27. Using SP aid as an example, I calculate $DFBETA_i$ as $DFBETA^i_{SP} = (\widehat{\beta}^i_{SP} - \widehat{\beta}_{SP})/(\widehat{SE}_{\widehat{\beta}^i_{SP}})$, where $\widehat{\beta}_{SP}$ is the estimated coefficient in the full sample, $\widehat{\beta}^i_{SP}$ is the estimate when country i is eliminated and $\widehat{SE}_{\widehat{\beta}^i_{SP}}$ is the estimated standard error of the coefficient in the model without country i (see, e.g., Bollen and Jackman, 1990).

procedure removes 14 countries in the education sector and 5 countries in the health sector.[28] Table 6 presents the results of estimating equation (11) for this reduced sample, and figure 2 shows partial scatter plots of the key relationships in the FE regressions for both the full and reduced samples. The FE results in the reduced sample are similar to those in the full sample. The effect of TC in both sectors remains close to zero, and full fungibility is easily rejected. The effect of education SP aid decreases sharply to 0.83, which is also the size of the nearly unchanged coefficient of health SP aid. However, full fungibility is rejected in both cases. This result suggests that the conclusions from table 4, namely that the fungibility of education and health SP aid and TC is limited, are not solely driven by the particular experience of a small number of aid recipients.[29] In what follows, I continue to work with this reduced sample.

To interpret the FE coefficients in a causal way requires a potentially strong assumption of strict exogeneity. This assumption would be violated if, for instance, the allocation of education (health) SP aid and TC were partially determined on the basis of past or current values of public education (health) expenditure. In fact, table 6 contains some evidence indicating that strict exogeneity for SP aid is unlikely to hold. If a first-differenced version of equation (11) is estimated with OLS (columns 2 and 4, labeled FD), then the effect of SP aid differs markedly from its FE estimate and even becomes negative. This stark difference between the FE and FD estimates of the SP aid coefficients suggests a violation of the strict exogeneity assumption because such a violation causes both FE and FD to be inconsistent and to have different probability limits (Wooldridge, 2002, pp. 284-285). However, the effect of TC is similar in the first-differenced model. There is some evidence of a negative effect of TC, especially in the education sector, in which the hypothesis of no fungibility can be rejected at a 10% significance level, but any displacement of sectoral public expenditure is minimal. Hence, the conclusion that the fungibility of TC is limited is confirmed in the FD model. A second indication that the FE model may be misspecified emerges from a serial correlation test of the idiosyncratic errors.[30] For both sectors, I reject the null hypothesis of no serial correlation at a significance level of less than 1%. Although clustering standard errors on the recipient country should ensure that inferences are valid, the presence of a

28. These countries are Burkina Faso, Côte d'Ivoire, Eritrea, Guinea-Bissau, Guyana, Lesotho, Mozambique, Nicaragua, Papua New Guinea, Samoa, Seychelles, Sierra Leone, Tajikistan and Tonga for education and Eritrea, Guinea-Bissau, Sierra Leone, Tajikistan, and Zambia for health.

29. The elimination of outliers that are identified by either the method proposed by Hadi (1992, 1994) (hadimvo in Stata) or the method proposed by Billor, Hadi and Velleman (2000) (bacon in Stata, see Weber, 2010) yields similar results. I follow Roodman (2007) in applying these methods to the partialled-out versions of public sectoral expenditure and sectoral SP aid and TC (i.e., the residuals that are obtained from the FE regressions of these variables on the other variables).

30. Under the null hypothesis of no serial correlation, the residuals in the first-differenced model should have an autocorrelation of -0.5. Thus, a Wald test of this hypothesis can be performed to test for the presence of serial correlation in ε_{it} (Wooldridge, 2002, p. 283; Drukker, 2003). I conduct this test in Stata using the xtserial command.

TABLE 6. Disaggregated Education and Health Aid, Reduced Sample

	Public education exp.		Public health exp.	
	FE	FD	FE	FD
Education IP	0.22	0.34***		
	(0.15)	(0.12)		
Education SP	0.83**	−0.34		
	(0.34)	(0.55)		
Education TC	0.024	−0.070		
	(0.059)	(0.046)		
Education ONM	−0.25	−0.044		
	(0.24)	(0.11)		
Health IP			0.17	−0.19*
			(0.19)	(0.11)
Health SP			0.83**	−0.19
			(0.36)	(0.24)
Health TC			−0.15	−0.040
			(0.20)	(0.10)
Health ONM			0.31*	0.095
			(0.17)	(0.12)
General aid	0.027	0.00092	0.0082	−0.0074
	(0.020)	(0.015)	(0.018)	(0.011)
Support to NGOs	−0.48	−0.17	−0.055	−0.025
	(0.31)	(0.23)	(0.14)	(0.073)
Other non-education aid	0.00046	0.0049		
	(0.016)	(0.010)		
Other non-health aid			−0.019*	0.0039
			(0.011)	(0.0047)
GDP per capita	0.22*	−0.058	0.093	0.13*
	(0.12)	(0.088)	(0.071)	(0.063)
GDP per capita growth	−0.019***	−0.0079**	−0.015***	−0.011***
	(0.0045)	(0.0031)	(0.0043)	(0.0031)
Urbanisation	0.039	0.0033	0.019	0.017
	(0.045)	(0.064)	(0.025)	(0.026)
Trade	−0.0035	−0.0025	−0.0013	0.00046
	(0.0041)	(0.0037)	(0.0023)	(0.0021)
PV debt	−0.0055**	−0.0038	−0.00026	−0.00017
	(0.0025)	(0.0038)	(0.00056)	(0.00072)
Public debt service	−0.059***	−0.024*	−0.019	−0.0031
	(0.019)	(0.012)	(0.012)	(0.0057)
R^2	0.183	0.062	0.135	0.051
Hausman	0.000		0.000	
$\beta_{SP} \leq 0$	0.008	0.731	0.012	0.781
$\beta_{SP} \geq 1$	0.307	0.008	0.313	0.000
$\beta_{TC} \leq -1$	0.000	0.000	0.000	0.000
$\beta_{TC} \geq 0$	0.658	0.066	0.239	0.347
Countries	94	94	103	102

(Continued)

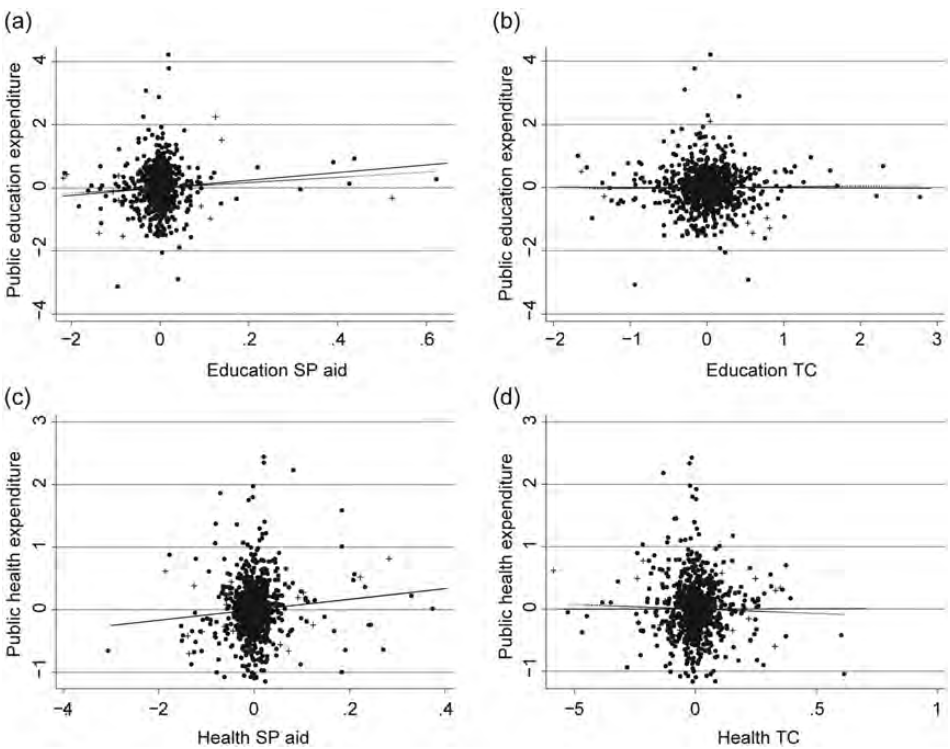
TABLE 6. Continued

	Public education exp.		Public health exp.	
	FE	FD	FE	FD
Observations	921	819	1024	912

Note: Fixed effects (FE) and first−differenced OLS (FD) results, annual data, 1990−2003, reduced sample. All regressions include time dummies, coefficients not reported. Heteroskedasticity-robust standard errors, clustered by country, in brackets. *, **, and *** denote significance at 10, 5 and 1%, respectively. In the case of FE estimation, R^2 refers to the within R^2. Hausman shows the p−value of a generalized Hausman test of the null hypothesis that η_i is uncorrelated with the regressors. $\beta_{SP} \leq 0$ ($\beta_{SP} \geq 1$) and $\beta_{TC} \leq -1$ ($\beta_{TC} \geq 0$) are p-values for the test of full (no) fungibility for sector program aid and technical cooperation, respectively.

Source: Author's analysis based on data described in the text.

FIGURE 2. Partial Scatter Plots



Note: Partial scatter plots in the full (solid line) and reduced (dotted line) samples correspond to the FE results in tables 4 and 6, respectively. + denotes observations that are excluded from the reduced sample.

Source: Author's analysis based on data described in the text.

serial correlation in ε_{it} may indicate that the model is dynamically misspecified, which would again render the FE estimates inconsistent.

Therefore, I also examine the results that are obtained when the strict exogeneity assumption is relaxed by employing a system GMM estimator (Arellano and Bover, 1995; Blundell and Bond, 1998). This estimator further enables the consistent estimation of a more general model that includes a lagged dependent variable (which removes the serial correlation in ε_{it}):[31]

$$\begin{aligned} SSP_{it} = \alpha SSP_{i,t-1} + \beta_{IP} SAIDIP_{it} + \beta_{SP} SAIDSP_{it} + \beta_{TC} SAIDTC_{it} \\ + \beta_{ONM} SAIDONM_{it} + \gamma \boldsymbol{A}_{it} + \boldsymbol{\delta X}_{it} + \lambda_t + \eta_i + \varepsilon_{it}. \end{aligned} \tag{12}$$

Equation (12) is estimated using a two-step system GMM estimator applying Windmeijer's (2005) correction for the downward bias in the two-step standard errors. All education (health) aid prefix code variables, support to NGOs, and trade are treated as endogenous, whereas all other variables are treated as predetermined. Time dummies are treated as strictly exogenous and are thus added to the instrument matrix without transformation. I reduce the risk of overfitting by restricting the maximum number of lags of the level variables that are used as instruments for the differenced equation[32] and by collapsing the instrument matrix, which creates an instrument for each variable and lag distance rather than for each variable, time period, and lag distance (Roodman, 2009a, 2009b). To conserve space, I do not report the system GMM results and discuss them only briefly (the full results and a more detailed discussion are available in the working paper version of this article).

The short-term effect of SP aid in both sectors is near zero but is volatile across the different instrument configurations and is estimated imprecisely. As a result, neither the null hypothesis of full fungibility nor the null hypothesis of no fungibility can typically be rejected at conventional significance levels. This volatility and imprecision carry over to the estimate of the long-term effect of education SP aid, $\widehat{\beta}_{SP}^{LR} = \widehat{\beta}_{SP}/(1-\widehat{\alpha})$. This imprecision likely results from the lack of variation in SP aid. The effect of education TC is close to zero, and the null hypothesis of full fungibility is always strongly rejected. No fungibility cannot be rejected, and the point estimate suggests, at most, only minor displacement of public education expenditures by education TC in the short term. Given the persistence in public education expenditures, the estimate of the long-term effect of education TC is more negative (with -0.3 as the lowest estimate), but even in the long term, full fungibility is rejected and no fungibility is

31. Briefly, the GMM estimator differences equation (12) to remove the fixed effect and uses suitably lagged levels of the dependent variable and the right-side variables as instruments for the differenced equation. In addition, the system GMM estimator utilizes the equation in levels, using suitably lagged differences as instruments.

32. I examine a number of different instrument configurations, from the use of a single lag of each variable to instrument the differenced equation to the use of four lags of each variable.

not rejected. In the health sector, full fungibility of TC in the short term is also rejected across the board. In fact, health TC is found to have a positive effect, although the estimate is never significantly different from zero. The average estimated LR effect is approximately 2.6 but has a large standard error. Nonetheless, in all cases except when only a single lag of the variables is used to instrument the differenced equation, full fungibility in the long term can still be rejected. Similar long-term effects are found when a lag of TC aid is added as an explanatory variable in equation (12).

An Alternative Assessment

Finally, it is worthwhile to consider an alternative approach that allows for a broader assessment of the degree of fungibility of education and health aid while allowing for some uncertainty in the measurement of on- and off-budget aid. Beginning with (3), the estimated coefficient of health aid in (2) can be written as follows:[33]

$$\widehat{\beta} = \widehat{\beta}_{ON} w + \widehat{\beta}_{OFF}(1 - w), \tag{13}$$

with weight w

$$w = \frac{1 + \rho\sqrt{\delta}}{1 + \delta + 2\rho\sqrt{\delta}} \tag{14}$$

and with $\rho = \sigma_{ON,OFF}/(\sigma_{ON}\sigma_{OFF})$ as the correlation between on- and off-budget health aid and $\delta = \sigma^2_{OFF}/\sigma^2_{ON}$ as the relative variance of off- versus on-budget health aid. If we impose that on- and off-budget health aid have the same degree of fungibility ($\widehat{\beta}_{OFF} = \widehat{\beta}_{ON} - 1$), then we can rearrange equation (13) to express $\widehat{\beta}_{ON}$ as a function of $\widehat{\beta}$:

$$\widehat{\beta}_{ON} = \widehat{\beta} + 1 - \frac{1 + \rho\sqrt{\delta}}{1 + \delta + 2\rho\sqrt{\delta}}. \tag{15}$$

This equation demonstrates how, for given values of ρ and δ, our naive estimate of β can be used to generate an estimate ($\widehat{\beta}_{ON}$) that can be used to determine the degree of fungibility: a value of $\widehat{\beta}_{ON}$ that is close to 1 indicates that there is little or no fungibility, whereas a value that is closer to 0 suggests a greater degree of fungibility.[34] Table 7 performs this computation for total aid and for each of the 4 aid types in both sectors, beginning with the FE coefficients that are estimated in tables 3 and 4, respectively. For each variable, the

33. As in section I, I focus on the fungibility of health aid for the sake of concreteness.

34. As noted previously, in a model that includes control variables and that is estimated using a FE estimator, ρ refers to the correlation between the partialled-out versions of off- and on-budget aid, and δ refers to the relative variance of the partialled-out versions of off- versus on-budget aid.

TABLE 7. Fungibility of Education and Health Aid

		ρ						
		−1	−3/4	−1/2	0	1/2	3/4	1
(a) Education aid								
δ	1/4	−1.00	−0.25	0.00	0.20	0.29	0.32	0.34
	1/2	−2.41	−0.06	0.19	0.34	0.39	0.41	0.42
	3/4	−6.46	0.23	0.36	0.43	0.46	0.46	0.47
	1	.	0.50	0.50	0.50	0.50	0.50	0.50
	3/2	5.45	0.88	0.70	0.60	0.57	0.56	0.55
	2	3.42	1.07	0.82	0.67	0.62	0.60	0.59
	4	2.00	1.25	1.00	0.80	0.72	0.69	0.67
(b) Health aid								
δ	1/4	−0.74	0.01	0.26	0.46	0.54	0.57	0.59
	1/2	−2.16	0.19	0.44	0.59	0.65	0.66	0.67
	3/4	−6.21	0.48	0.62	0.69	0.71	0.72	0.72
	1	.	0.76	0.76	0.76	0.76	0.76	0.76
	3/2	5.71	1.14	0.96	0.86	0.83	0.82	0.81
	2	3.67	1.33	1.07	0.93	0.87	0.86	0.84
	4	2.26	1.51	1.26	1.06	0.97	0.95	0.93
(c) Education IP								
δ	1/4	−0.88	−0.13	0.12	0.32	0.41	0.44	0.46
	1/2	−2.29	0.05	0.31	0.46	0.51	0.53	0.54
	3/4	−6.34	0.35	0.48	0.55	0.57	0.58	0.59
	1	.	0.62	0.62	0.62	0.62	0.62	0.62
	3/2	5.57	1.00	0.82	0.72	0.69	0.68	0.67
	2	3.54	1.19	0.94	0.79	0.74	0.72	0.71
	4	2.12	1.37	1.12	0.92	0.84	0.81	0.79
(d) Health IP								
δ	1/4	−0.80	−0.05	0.20	0.40	0.49	0.51	0.53
	1/2	−2.21	0.13	0.38	0.53	0.59	0.60	0.61
	3/4	−6.26	0.42	0.56	0.63	0.65	0.66	0.66
	1	.	0.70	0.70	0.70	0.70	0.70	0.70
	3/2	5.65	1.08	0.90	0.80	0.77	0.76	0.75
	2	3.61	1.27	1.02	0.87	0.81	0.80	0.79
	4	2.20	1.45	1.20	1.00	0.91	0.89	0.87
(e) Education SP aid								
δ	1/4	0.21	0.96	1.21	1.41	1.49	1.52	1.54
	1/2	−1.21	1.14	1.39	1.54	1.59	1.61	1.62
	3/4	−5.26	1.43	1.56	1.63	1.66	1.66	1.67
	1	.	1.71	1.71	1.71	1.71	1.71	1.71
(f) Health SP aid								
δ	1/4	−0.16	0.59	0.84	1.04	1.13	1.16	1.18
	1/2	−1.57	0.78	1.03	1.18	1.23	1.25	1.26
	3/4	−5.62	1.07	1.20	1.27	1.30	1.30	1.31
	1	.	1.34	1.34	1.34	1.34	1.34	1.34
(g) Education TC								
δ	1	.	0.49	0.49	0.49	0.49	0.49	0.49
	3/2	5.44	0.87	0.69	0.59	0.56	0.55	0.54
	2	3.41	1.06	0.81	0.66	0.61	0.59	0.58
	4	1.99	1.24	0.99	0.79	0.71	0.68	0.66

(Continued)

TABLE 7. Continued

		ρ						
		−1	−3/4	−1/2	0	1/2	3/4	1
(h) Health TC								
δ	1	.	0.51	0.51	0.51	0.51	0.51	0.51
	3/2	**5.46**	**0.88**	**0.70**	**0.61**	**0.57**	**0.56**	**0.56**
	2	**3.42**	**1.08**	**0.82**	**0.67**	**0.62**	**0.60**	**0.59**
	4	**2.01**	**1.26**	**1.01**	**0.81**	**0.72**	**0.69**	**0.67**
(i) Education ONM aid								
δ	1/4	−0.98	−0.23	0.02	0.22	0.31	**0.33**	**0.35**
	1/2	−2.39	−0.05	0.21	**0.35**	**0.41**	**0.42**	**0.43**
	3/4	−6.44	0.24	**0.38**	**0.45**	**0.47**	**0.48**	**0.48**
	1	.	**0.52**	**0.52**	**0.52**	**0.52**	**0.52**	**0.52**
	3/2	**5.47**	**0.90**	**0.72**	**0.62**	**0.59**	**0.58**	**0.57**
	2	**3.43**	**1.09**	**0.84**	**0.69**	**0.63**	**0.62**	**0.61**
	4	**2.02**	**1.27**	**1.02**	**0.82**	**0.73**	**0.71**	**0.69**
(j) Health ONM aid								
δ	1/4	−0.59	0.16	**0.41**	**0.61**	**0.70**	**0.73**	**0.75**
	1/2	−2.00	0.35	**0.60**	**0.75**	**0.80**	**0.82**	**0.83**
	3/4	−6.05	**0.64**	**0.77**	**0.84**	**0.87**	**0.87**	**0.88**
	1	.	**0.91**	**0.91**	**0.91**	**0.91**	**0.91**	**0.91**
	3/2	**5.86**	**1.29**	**1.11**	**1.01**	**0.98**	**0.97**	**0.96**
	2	**3.83**	**1.48**	**1.23**	**1.08**	**1.03**	**1.01**	**1.00**
	4	**2.41**	**1.66**	**1.41**	**1.21**	**1.13**	**1.10**	**1.08**

Note: The entries in this table are the values of $\widehat{\beta}_{ON}$ computed according to equation (15) for different aid variables, starting from the FE coefficients estimated in tables 3 and 4. ρ is the correlation between the on- and off-budget components of the aid variable, δ the relative variance of the off- versus on-budget component of the aid variable. Bold (underlined) entries indicate that the null of full (no) fungibility is rejected at a 5% significance level.

Source: Author's analysis based on data described in the text.

entries in the table calculate $\widehat{\beta}_{ON}$ for different values of the relative variance of off- versus on-budget aid in the aid type considered (δ, ranging from 1/4 to 4) and the correlation between its on- and off-budget components (ρ, ranging from -1 to 1). A bold (underlined) entry indicates that the null hypothesis of full (no) fungibility can be rejected at a 5% significance level.

After partialling out the fixed effects and the control variables, the correlations between the four different aid types (IP, SP, TC and ONM) are a useful indication of the most plausible values of ρ for total education and health aid. In both sectors, these correlations are close to zero. The most negative correlation is between health SP aid and TC (-0.15), and the most positive correlation is between education TC and ONM (0.14). Hence, ρ is not expected to be far from 0. Meanwhile, it is very likely that most of the variation in total education and health aid is driven by off-budget aid (implying $\delta \geq 1$). Technical assistance, which I have argued is almost entirely off-budget, dominates the variation in health and, especially, education aid (see table 1), while there is

some evidence to suggest that the other non-program components are also not well captured in the budgets of recipient governments (see section III). Hence, the entries in the bottom four rows of tables 7a and 7b are the most plausible. Especially in the health sector, these entries indicate a low degree of fungibility. For health aid, the null hypothesis of no fungibility is never rejected for $\delta \geq 3/2$; even for a δ value as low as 1/2, a fairly low degree of fungibility is found for most values of ρ.

With regard to the aid types, even under the assumption that SP aid is completely on-budget, its estimated FE coefficient in table 4 for both sectors implies low fungibility. Hence, it is not surprising that this conclusion is confirmed in tables 7e and 7f.[35] Tables 7g and 7h relax the assumption that TC is completely off-budget. In almost all cases, the null hypothesis of full fungibility can still be rejected, and most entries suggest limited fungibility. For health TC, the null hypothesis of no fungibility is never rejected. The vast majority of entries in table 7j indicate a low degree of fungibility of health ONM aid, with few rejections of the null hypothesis of no fungibility. The degree of fungibility is higher for education ONM aid and is more difficult to assess. Both null hypotheses are typically rejected; thus, the results suggest partial fungibility, but the exact degree of fungibility depends on the relative variation of off-budget versus on-budget aid, which is difficult to determine. The discussion in section III suggests that aid projects (IP) are frequently not captured in the budgets of recipient governments. Even when the relative variance of off- versus on-budget IP aid is 1 or slightly below 1, the entries in tables 7c and 7d again indicate fairly low degrees of fungibility, especially in the health sector. Hence, unless ρ is very negative, we would only be comfortable concluding that IP is mostly fungible if we believe that the variance in off-budget IP is substantially lower than the variance in on-budget IP.

V. Conclusion

This paper presents new empirical evidence to provide insight into the difficult issue of foreign aid fungibility. I construct data on earmarked education and health aid disbursements that also distinguish between on- and off-budget components of aid. Sector program aid measures on-budget aid, whereas technical cooperation proxies for off-budget aid. I illustrate how a failure to adequately address the presence of off-budget aid may have biased previous estimates of foreign aid fungibility.

Overall, I find little evidence that aid is fully or even largely fungible; rather, most point estimates suggest limited fungibility. In both sectors, technical cooperation leads to, at most, a small displacement of a recipient's own public spending. This effect is estimated relatively tightly, especially in the education sector. Thus, the results suggest a genuine effect rather than merely noise in the

35. For SP aid, I consider only $\delta \leq 1$, and for TC, I consider only $\delta \geq 1$.

data. The effect of technical cooperation is robust across a range of models, whereas the effect of sector program aid is more volatile. In a static panel data model, fixed effects results suggest an approximately one-to-one correlation between sector program aid and public sectoral expenditure, which is robust to a large number of specification changes. However, when system GMM is used to estimate a dynamic model, the effect of sector program aid is imprecisely estimated. Thus, no firm conclusions can be drawn with respect to the fungibility of sector program aid.

Therefore, the result of limited fungibility for education and health aid specifically pertains to technical cooperation. Because technical cooperation is the dominant modality in both sectors, however, it plays a large role in determining the overall degree of fungibility of earmarked education and health aid. The extent to which investment projects and other aid are on- or off-budget is more uncertain, making it more difficult to determine the degree to which these projects are fungible. However, the analysis in section IV suggests that unless we believe that the variance of the on-budget components of investment projects and other aid dominates the variance of their off-budget components, both types of aid are far from fully fungible.

The lack of fungibility may be a consequence of effective donor conditionality. If donors are able to monitor the spending of recipient governments, then they may be able to credibly enforce the condition that aid adds to the resources that are spent in the targeted sector. An additional reason for the low degree of fungibility primarily applies to technical cooperation and is less applicable to other aid types. This explanation is the observation made by Gramlich (1977) that heterogeneity in government expenditures may contribute to reduced fungibility. To the extent that governments in developing countries spend few resources on the type of goods and services that are provided by technical cooperation, it becomes impossible to significantly reduce this class of expenditure because these expenditures rapidly approach the lower bound of zero. If the substitutability between different types of expenditures in a recipient government's utility function is also limited, then low fungibility for technical cooperation may ensue. Finally, a lack of information on the part of a recipient government, which is particularly relevant for off-budget aid, may also reduce the degree of fungibility that is observed in practice.

From the donor perspective, the results in this paper suggest that the costly effort associated with earmarking (e.g., monitoring costs) may not be futile. From the perspective of the population in a recipient country, the limited fungibility of education and health aid can be perceived as a positive result if we believe that better education and health have positive consequences for human welfare. However, this positive interpretation persists only if the aid in these sectors effectively produces valuable outcomes. Moreover, if the low fungibility of off-budget aid arises because a recipient government is not fully aware of this aid, then any positive effects of non-fungible off-budget aid must be balanced against the possible deleterious effects on government capacity and

ownership that are incurred when channeling funds outside of a budget. In general, not a great deal is known about the normative consequences of fungibility (for papers that look at this issue, see McGillivray and Morrissey, 2000; Pettersson, 2007a, 2007b; Wagstaff, 2011), and this constitutes an important area for future study.

References

Adam, Christopher S., Per-Åke Andersson, Arne Bigsten, Paul Collier, and Steve O'Connell. 1994. "Evaluation of Swedish development co-operation with Zambia." Stockholm: Ministry of Foreign Affairs, Secretariat for analysis of Swedish development assistance, Report 6.

Arellano, Manuel. 1987. "Computing robust standard errors for within-groups estimators." *Oxford Bulletin of Economics and Statistics* 49(4): 431–434.

———. 1993. "On the testing of correlated effects with panel data." *Journal of Econometrics* 59(1-2): 87–97.

Arellano, Manuel, and Olympia Bover. 1995. "Another look at the instrumental variable estimation of error-components models." *Journal of Econometrics* 68(1): 29–51.

Baqir, Reza. 2002. "Social sector spending in a panel of countries." Washington, D.C.: International Monetary Fund, Working Paper 02/35.

Baser, Heather, and Peter Morgan. 2001. "The pooling of technical assistance: an overview based on field research experience in six African countries." A study for the Directorate-General for Development Cooperation (DGIS), Ministry of Foreign Affairs of the Netherlands.

Belsley, David A., Edwin Kuh, and Roy E. Welsch. 1980. *Regression diagnostics: identifying influential data and sources of collinearity*. New York: John Wiley & Sons.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How much should we trust differences-in-differences estimates." *Quarterly Journal of Economics* 119(1): 249–275.

Billor, Nedret, Ali S. Hadi, and Paul F. Velleman. 2000. "BACON: blocked adaptive computationally efficient outlier nominators." *Computational Statistics & Data Analysis* 34: 279–298.

Blundell, Richard, and Stephen Bond. 1998. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics* 87(1): 115–143.

Bollen, Kenneth A., and Robert W. Jackman. 1990. "Regression diagnostics: an expository treatment of outliers and influential cases." In John Fox, and J. Scott Long, eds., *Modern methods of data analysis*, London: Sage.

Chatterjee, Santanu, Paola Giuliano, and Ilker Kaya. 2007. "Where has all the money gone? Foreign aid and the quest for growth." Bonn: Institute for the Study of Labor (IZA), Discussion Paper No. 2858.

Depetris Chauvin, Nicolas, and Aart Kraay. 2005. "What has 100 billion dollars worth of debt relief done for low-income countries." Washington, D.C.: World Bank, mimeo.

Devarajan, Shantayanan, Andrew Sunil Rajkumar, and Vinaya Swaroop. 2007. "What does aid to Africa finance." In Sajal Lahiri, ed., *Theory and practice of foreign aid*, Amsterdam: Elsevier.

Dikhanov, Yuri. 2004. "Historical PV of debt in developing countries, 1980-2002." Washington, D.C.: World Bank, mimeo.

Dreher, Axel. 2006. "The influence of globalization on taxes and social policy: an empirical analysis for OECD countries." *European Journal of Political Economy* 22(1): 179–201.

Drukker, David M. 2003. "Testing for serial correlation in linear panel-data models." *Stata Journal* 3(2): 168–177.

Fagernäs, Sonja, and John Roberts. 2004a. "The fiscal effects of aid in Uganda." London: Overseas Development Institute (ODI), Economics and Statistics Analysis Unit, Working Paper 9.

———. 2004b. "The fiscal effects of aid in Zambia." London: Overseas Development Institute (ODI), Economics and Statistics Analysis Unit, Working Paper 10.

Fagernäs, Sonja, and Cedrik Schurich. 2004. "The fiscal effects of aid in Malawi." London: Overseas Development Institute (ODI), Economics and Statistics Analysis Unit, Working Paper 7.

Feeny, Simon. 2007. "Foreign aid and fiscal governance in Melanesia." *World Development* 35(3): 439–453.

Feyzioglu, Tarhan, Vinaya Swaroop, and Min Zhu. 1998. "A panel data analysis of the fungibility of foreign aid." *World Bank Economic Review* 12(1): 29–58.

Franco-Rodriguez, Susana, Oliver Morrissey, and Mark McGillivray. 1998. "Aid and the public sector in Pakistan: evidence with endogenous aid." *World Development* 26(7): 1241–1250.

Gerdtham, Ulf-G., and Bengt Jönsson. 2000. "International comparisons of health expenditure: theory, data, and econometric analysis." In Anthony J. Culyer, and Joseph P. Newhouse, eds., *Handbook of health economics*, Amsterdam: Elsevier Science.

Gramlich, Edward M. 1977. "Intergovernmental grants: a review of the empirical literature." In Wallace E. Oates, ed., *The political economy of fiscal federalism*, Lexington, Massachusetts: Lexington Books.

Gupta, Sanjeev, Benedict Clements, and Erwin Tiongson. 1998. "Public spending on human development." *Finance and Development* 35(3): 10–13.

Hadi, Ali S. 1992. "Identifying multiple outliers in multivariate data." *Journal of the Royal Statistical Society: Series B* 54(3): 761–771.

———. 1994. "A modification of a method for the detection of outliers in multivariate samples." *Journal of the Royal Statistical Society: Series B* 56(2): 393–396.

Heller, Peter S. 1975. "A model of public fiscal behavior in developing countries: aid, investment and taxation." *American Economic Review* 65(3): 429–445.

Hepp, Ralf. 2005. "Health expenditures under the HIPC initiative." Davis, California: University of California, Davis, Department of Economics, mimeo.

IDD and Associates. 2006. "Evaluation of general budget support: synthesis report." Birmingham: International Development Department, School of Public Policy, University of Birmingham.

International Monetary Fund. 2006. "Government Finance Statistics." June 2006 CD-ROM.

Johnson, Alison, and Matthew Martin. 2005. "Empowering developing countries to lead the aid partnership: a background paper for the UNDP human development report." New York: UNDP, Human Development Report Office, Occassional Paper 2005/16 (February 2005).

Khilji, Nasir M., and Ernest M. Zampelli. 1991. "The fungibility of US assistance to developing countries and the impact on recipient expenditures: a case study of Pakistan." *World Development* 19(8): 1095–1105.

———. 1994. "The fungibility of US military and non-military assistance and the impacts on expenditures of major aid recipients." *Journal of Development Economics* 43(2): 345–362.

Lichtenberg, Frank R. 1990. "Aggregation of variables in least-squares regression." *American Statistician* 44(2): 169–171.

Lu, Chunling, Matthew T. Schneider, Paul Gubbins, Katherine Leach-Kemon, Dean Jamison, and Christopher J.L. Murray. 2010. "Public financing of health in developing countries: a cross-national systematic analysis." *Lancet* 375(9723): 1375–1387.

Marshall, Monty G., and Keith Jaggers. 2007. "Polity IV project: dataset users' manual." Vienna: Center for Systemic Peace, mimeo.

Mavrotas, George. 2002. "Foreign aid and fiscal response: does aid disaggregation matter." *Weltwirtschaftliches Archiv* 138(3): 534–559.

McGillivray, Mark, and Oliver Morrissey. 2000. "Aid fungibility in Assessing Aid: red herring or true concern." *Journal of International Development* 12(3): 413–428.

———. 2001. "Aid illusion and public sector fiscal behaviour." *Journal of Development Studies* 37(6): 118–136.

———. 2004. "Fiscal effects of aid." In Tony Addison, and Alan Roe, eds., *Fiscal policy for development: poverty, reconstruction and growth*, Basingstoke: Palgrave Macmillan.

McGuire, Martin. 1982. "US assistance, Israeli allocation, and the arms race in the Middle East: an analysis of three interdependent resource allocation processes." *Journal of Conflict Resolution* 26(2): 199–235.

———. 1987. "Foreign assistance, investment, and defense: a methodological study with an application to Israel, 1960-1979." *Economic Development and Cultural Change* 35(4): 847–873.

Mokoro. 2008. "Putting aid on budget: synthesis report." A study for the Collaborative Africa Budget Reform Initiative (CABRI) and the Strategic Partnership with Africa (SPA) (April 2008), http://www.mokoro.co.uk/files/13/publication/cabri_synthesis_eng_proof_6.pdf.

OECD. 2000a. "DAC Statistical Reporting Directives."

———. 2000b. "Handbook for reporting debt reorganisation on the DAC questionnaire."

———. 2002. "Reporting Directives for the Creditor Reporting System."

———. 2007a. "DAC Statistical Reporting Directives."

———. 2007b. "Reporting Directives for the Creditor Reporting System."

———. 2008. "2008 survey on monitoring the Paris declaration: making aid more effective by 2010."

Ouattara, Bazoumana. 2006. "Aid, debt and fiscal policies in Senegal." *Journal of International Development* 18(8): 1105–1122.

Pack, Howard, and Janet Rothenberg Pack. 1990. "Is foreign aid fungible? The case of Indonesia." *Economic Journal* 100(399): 188–194.

———. 1993. "Foreign aid and the question of fungibility." *Review of Economics and Statistics* 75(2): 258–265.

———. 1999. "Foreign aid and fiscal stress." In Gary R. Saxonhouse, and T.N. Srinivasan, eds., *Development, duality, and the international economic regime: essays in honor of Gustav Ranis*, Ann Arbor, Michigan: University of Michigan Press.

Pettersson, Jan. 2007a. "Child mortality: is aid fungibility in pro-poor expenditure sectors decisive." *Review of World Economics* 143(4): 673–693.

———. 2007b. "Foreign sectoral aid fungibility, growth and poverty reduction." *Journal of International Development* 19(8): 1074–1098.

The Political Risk Services Group. 2008. "International Country Risk Guide methodology." http://www.prsgroup.com/PDFS/icrgmethodology.pdf (consulted 7 July 2008).

Ravishankar, Nirmala, Paul Gubbins, Rebecca J. Cooley, Katherine Leach-Kemon, Catherine M. Michaud, Dean T. Jamison, and Christopher J.L. Murray. 2009. "Financing of global health: tracking development assistance for health from 1990 to 2007." *Lancet* 373(9681): 2113–2124.

Republic of Liberia Ministry of Finance. 2009. "Draft national budget for the fiscal year July 1, 2009 – June 30, 2010."

Rodrik, Dani. 1998. "Why do more open economies have bigger governments." *Journal of Political Economy* 106(5): 997–1032.

Roodman, David. 2007. "The anarchy of numbers: aid, development, and cross-country empirics." *World Bank Economic Review* 21(2): 255–277.

———. 2009a. "How to do xtabond2: an introduction to difference and system GMM in Stata." *Stata Journal* 9(1): 86–136.

———. 2009b. "A note on the theme of too many instruments." *Oxford Bulletin of Economics and Statistics* 71(1): 135–158.

Schaffer, M.E., and S. Stillman. 2006. "Stata module to calculate tests of overidentifying restrictions after xtreg, xtivreg, xtivreg2 and xthtaylor." http://ideas.repec.org/c/boc/bocode/s456779.html.

Sundberg, Mark, and Alan Gelb. 2006. "Making aid work." *Finance and Development* 43(4): 14–17.

Swaroop, Vinaya, Shikha Jha, and Andrew Sunil Rajkumar. 2000. "Fiscal effects of foreign aid in a federal system of governance: the case of India." *Journal of Public Economics* 77(3): 307–330.

Wagstaff, Adam. 2011. "Fungibility and the impact of development assistance: evidence from Vietnam's health sector." *Journal of Development Economics* 94(1): 62–73.

Weber, Sylvain. 2010. "bacon: an effective way to detect outliers in multivariate data using Stata (and Mata)." *Stata Journal* 10(3): 331–338.

Windmeijer, Frank. 2005. "A finite sample correction for the variance of linear efficient two-step GMM estimators." *Journal of Econometrics* 126(1): 25–51.

Wooldridge, Jeffrey M. 2002. *Econometric analysis of cross section and panel data*. Cambridge, Massachusetts: MIT Press.

World Bank. 2006a. "Edstats."

———. 2006b. "Global Development Finance."

———. 2006c. "World Development Indicators."

Industry Switching in Developing Countries

Carol Newman, John Rand, and Finn Tarp *

Firm turnover (i.e., firm entry and exit) is a well-recognized source of sector-level productivity growth. In contrast, the role and importance of firms that switch activities from one sector to another is not well understood. Firm switchers are likely to be unique, differing from both newly established entrants and exiting firms that are closing down operations. In this study, we develop an empirical model that examines switching behavior using data from Vietnamese manufacturing firms during the 2001–2008 period. The diagnostic shows that switching firms exhibit different characteristics and behavior than do entry and exit firms. Switchers tend to be labor intensive and to seek competitive opportunities in labor-intensive sectors in response to changes in market environments. Moreover, resource reallocation resulting from switching forms an important component of productivity growth. The topic of switching merits attention in the future design of firm surveys across developing countries and in associated analytical studies. JEL codes: D21, L6, O14 firm dynamics, sector switching, productivity, Vietnam

Disentangling and defining the contribution of firm turnover to sector-level productivity growth is an important challenge in development economics research[1] that is also relevant to policy making. This primary motivation for this study was reinforced when we discovered that some contributions to the literature (e.g., Aw et al. 2001) categorize all firms leaving a particular sector as exit firms even if these firms do not actually close but instead switch to production

* Carol Newman is a professor at Trinity College Dublin, email address: cnewman@tcd.ie; John Rand is a professor at University of Copenhagen, email address: jra@foi.ku.dk; and Finn Tarp (corresponding author) is the director of UNU-WIDER and a professor at the University of Copenhagen, email address: Finn.Tarp@econ.ku.dk.

The authors are grateful for their stimulating collaboration with the staff at the Central Institute of Economic Management and the Institute of Labor Science and Social Affairs in Hanoi, Vietnam. Thanks are also due to the participants at the 2007 and 2011 Nordic Conferences in Development Economics. Three anonymous referees provided valuable critiques, and expert guidance by the WBER editors is highly appreciated. Financial support from Danida is acknowledged, and the usual caveats apply.

1. See Caves (1998), Bartelsman and Doms (2000), Tybout (2000), and Syverson (2011) for background.

THE WORLD BANK ECONOMIC REVIEW, VOL. 27, NO. 2, pp. 357–388 doi:10.1093/wber/lhs030
Advance Access Publication November 29, 2012

in a different sector. Similarly, new entrants in a given sector regularly include both genuinely new firms and firms that previously operated in other sectors. Some studies of firm dynamics (e.g., Bernard et al. 2006) have followed establishments over time, treating sector switchers as incumbents. The classification of firms that change their main sector of production potentially affects the understanding of the reallocation processes that enhance productivity.

This thesis is substantiated by the literature on firm capabilities, including Sutton (2005) and Bernard et al. (2010), which suggests that the average productivity of switchers is likely to differ from that of real entrants/exits if the switching firms have underlying capabilities (e.g., know-how and working practices held within the firm setup) that affect the firm's level of productivity. Switchers have already incurred sunk costs and gained knowledge related to their capabilities when they initially established production. For a given postentry/switching productivity draw (in the spirit of learning models on the evolution of industry), this situation suggests that the average productivity level of switchers exceeds that of new entrants. If switching firms exhibit specific characteristics and are motivated differently than real entrants/exits, then these dynamics must be understood when assessing changes in economic policy or the external environment to determine their effect on productivity.[2]

Accordingly, rather than examining the drivers of productivity growth within firms, this study describes the dispersion of productivity across manufacturing firms in Vietnam (using census data for the period 2001–2008), with a special focus on disentangling the contribution of firm turnover to sector-level productivity growth. An industry switcher is defined as a firm whose main year t product and main year $t - 1$ product belong to different four-digit industries. Although this definition may classify marginal changes in the product mix as industry switching, the data show that most industry switchers are one-product firms. Consequently, this characterization of industry switching corresponds to extreme changes in production.

This study aims to analyze the processes by which firms decide to enter, exit, or switch sectors, to make predictions about the expected relative productivity differences between firms, and to establish the contribution to productivity growth of firm turnover and sector switching (as distinguished from the "real" entry and exit of firms). Given that one of our core objectives is to diagnose the extent to which differences exist between entry, exit, and switching firms, we consider the association between firm turnover and the observed characteristics of firms and sectors as well as the association with exogenous shocks, such as regulatory changes and trade liberalization.[3] Finally, factors correlated with

2. Recent literature has focused on changes in the product mix by surviving firms as the main channel of productivity growth; see Bernard et al. (2009, 2010), Goldberg et al. (2010), and Eckel and Neary (2006).

3. On the impact of trade liberalization on productivity growth, see Melitz (2003) and Pavcnik (2002), and see Eslava et al. (2004), Olley and Pakes (1996), and Stiroh and Strahan (2003) on how deregulation induces resource allocations within an industry.

TABLE 1. Conceptual Framework and Definition of Switchers

	Time ($t-1$)	Time (t)		Time ($t+1$)
1	Firm not operating/not established	Manufacture of furniture (ISIC 3610)	(entry)	Manufacture of furniture (ISIC 3610)
2	Manufacture of furniture (ISIC 3610)	Manufacture of furniture (ISIC 3610)	(exit)	Closed
3	Sawmilling and planning of wood (ISIC 2010)	Manufacture of furniture (ISIC 3610)	(switch in)	Manufacture of furniture (ISIC 3610)
4	Manufacture of furniture (ISIC 3610)	Manufacture of furniture (ISIC 3610)	(switch out)	Sawmilling and planning of wood (ISIC 2010)
5	Manufacture of furniture (ISIC 3610)	Manufacture of furniture (ISIC 3610)	(incumbent)	Manufacture of furniture (ISIC 3610)

Source: Authors' illustration.

switching behavior are explored empirically and compared to factors related to entry and exit in Vietnam.

Vietnam is a populous Southeast Asian economy that has experienced rapid economic growth since 1986, when the comprehensive economic reform process known as Doi Moi was initiated. During the past decade, Vietnam has been one of the fastest-growing economies in the world, with GDP growing at an annual rate of 7.3 percent from 2000 to 2009 (World Bank 2011). The Doi Moi process has included wide-ranging reforms of enterprise, commercial, and investment laws, especially since 2000. These reforms have been coupled with extensive trade liberalization, accession to the WTO in 2007, and significant inflows of foreign direct investment. These features offer further motivation for understanding industry evolution in a dynamic transition economy.

Section I below defines sector switching and specifies three testable hypotheses. Section II presents the empirical approach, and section III describes the data. Empirical results follow in section IV, and section V concludes that correctly assessing the respective contributions of entry, exit, and switching firms to productivity growth has important analytical and policy implications.

I. DEFINING INDUSTRY SWITCHING

Industry switching can be defined at various levels of sector aggregation. Switching is defined in this study at the four-digit International Standard Industrial Classification (ISIC) level, and the conceptual differences between switchers, entrants, exits, and incumbents are illustrated in table 1.[4] Differences in the classification of firms are considered from the perspective of a firm that manufactures furniture (ISIC 3610) and switches in or out of sawmilling and planing of wood (ISIC 2010). In previous studies, switchers-in

4. In this study, the four-digit industry definition is our focus. For analysis at the two-digit level, see Newman et al. (2011).

(captured in row 3) were classified as either new entrants (row 1), as in Aw et al. (2001), or as incumbents (row 5), as in Eslava et al. (2004). Similarly, switchers-out (captured in row 4) have been labeled exits (row 2) or incumbents (row 5).[5] Next, we focus on the differences in the way the literature has treated sector switchers.[6]

The above definition of industry switching has implications for multiproduct firms. The data contain information on whether firms produce more than one product (six-digit) but do not specify what products these firms produce. Thus, a multiproduct firm producing two products will be regarded as a sector switcher when its percentage rates of production marginally change from 49 percent of product 1 and 51 percent of product 2 to 51 percent of product 1 and 49 percent of product 2. Therefore, producers of more than one product are controlled for in the empirical analysis, and multiproduct firms are excluded as a robustness check. The data clearly show that switching in Vietnam is not limited to changes in the product mix. In fact, the majority of switching, as defined in this study, is associated with single-product firms.

Several empirical papers in the literature (see Bartelsman and Doms (2000) and Syverson (2011) for reviews) have shown that enterprise turnover contributes significantly to sector-level productivity growth, and various models seek to explain why productivity should enhance reallocation. Vintage models of industry dynamics suggest that, on average, new entrants are more productive because their productivity is enhanced by the latest technology. It is assumed that firm productivity remains constant over time unless it is affected by a random shock. Therefore, it is expected that the productivity of entrants will dominate that of incumbents and switchers, and firm exit occurs when productivity relative to entrants drops below a certain threshold. Eslava et al. (2004) and Foster et al. (2008) provide recent empirical evidence supporting these vintage model predictions.

Learning models on the evolution of industry yield similar insights regarding the mean productivity differences between entrants and exits, but they differ from vintage models with respect to the expected relative productivity between incumbents and entrants. For a given distribution function summarizing the heterogeneity in productivity among firms that are strictly decreasing in current productivity, Hopenhayn (1992) shows that the productivity distribution of incumbents stochastically dominates the productivity distribution of entrants.

5. Other examples of switching firms observed in the data include switchers from the manufacture of three-wheeled, no-motor vehicles (ISIC 3591) to the manufacture of motor vehicles (ISIC 3410), an example of a switch to a more technology-intensive sector. Switchers are also observed from the building of ships (ISIC 3511) to the repairing of ships and boats (ISIC 3513); these are closely related sectors for which the latter is more service driven.

6. It is possible that entering firms may immediately switch out/exit or that switching-in firms may immediately exit or switch out. In the analysis, these firms are treated as both switchers and entrants/exits.

Learning models have received empirical support in several papers, including Bartelsman and Doms (2000).

The seminal literature on the evolution of industry and industry dynamics does not consider the possibility that firms may choose to reallocate resources into new sectors to maximize profits or to avoid exiting the industry. At first glance, the importance of this possibility may not be clear. Changes to the main area of production may require a firm to undergo a learning period to gain knowledge of new production processes, and this change may be so fundamental that it may be similar to closing the business and opening a new firm in another sector. However, recent literature on the importance of firm/owner entrepreneurial capabilities (see Sutton [2005] for initial thoughts along these lines) suggests that productivity differences between switchers and entrants/exits should be expected. Given that switching firms have already incurred sunk costs when they initially established production (and have acquired, in the process, knowledge of different business procedures, access to public utilities and general market conditions), the sunk costs of switching sectors are arguably lower than the sunk costs facing new firms entering a sector.[7] Moreover, the variance of the postswitch productivity draw is expected to be smaller than that of entrants into the same sector due to the business experience gained from previous production (i.e., knowledge of the firm's underlying capabilities).

The above considerations lead to the following three testable hypotheses regarding productivity differences between entrants, exits, and incumbents, on the one hand, and switchers, on the other:

(i) Switchers-in versus Entrants

The aggregate sunk costs for switching firms will be below those of entrants. Switchers-in will have greater knowledge of both their productivity potential (entrepreneurial capabilities) and general market conditions and will therefore be likely to have a higher level of productivity than entrants.

(ii) Switchers-out versus Exits

Switchers-out will be more productive than exits. Switchers-out would also have exited if their expectations about future profitability (which depends on productivity) were equal to or less than that of exit firms. Switching depends on the observed characteristics of one sector relative to those of other sectors because expectations about future market conditions play an important role in the decision to switch and/or exit.

(iii) Switchers versus Incumbents

Switchers will be less productive than incumbents. They face a sunk cost of switching, and their information regarding their productivity

7. Switching costs may be interpreted, for example, as a depletion of capital stock as a result of certain machinery or equipment becoming obsolete. Investment in new machinery or equipment would be considered an investment that adds to the firm's capital stock.

and existing market conditions is related to their former sector rather than the new sector.

II. Empirical Approach

This section first outlines the method used to measure productivity and then describes the empirical model detailing the firm-, sector-, and industry-specific factors included in the analysis.

Productivity Measurement

Following Aw et al. (2001), we used an index number approach to estimate total factor productivity for firms in each manufacturing subsector.[8] Productivity is measured relative to the mean level of productivity in a given sector and year. To analyze productivity over time, this productivity differential is linked to changes in the reference levels of productivity from year to year.

The sector-specific total factor productivity (TFP) index is given in equation (1). The productivity of a firm is compared, in any given time period, relative to the average productivity of the sector.[9]

$$\begin{aligned}\omega_{ijt} = &\left(\ln Y_{ijt} - \overline{\ln Y_{jt}}\right) + \sum_{\tau=2}^{t}\left(\overline{\ln Y_{jt}} - \overline{\ln Y_{jt-1}}\right)\\ &- \sum_{m=1}^{k}\frac{1}{2}\left(s_{mijt} + \overline{s_{mjt}}\right)\left(\ln X_{mijt} - \overline{\ln X_{mjt}}\right)\\ &+ \sum_{\tau=2}^{t}\sum_{m=1}^{k}\frac{1}{2}\left(\overline{s_{mjt}} + \overline{s_{mjt-1}}\right)\left(\overline{\ln X_{mjt}} - \overline{\ln X_{mjt-1}}\right),\end{aligned} \tag{1}$$

where Y_{ijt} measures the output of firm i in sector j in year t, X_{mjit} is the amount of input m used by the firm, and s_{mjit} is the expenditure of the firm on input m as a share of the total expenditure.

8. Data are only available for the value of inputs and output, so it is not possible to estimate physical productivity measures as suggested by Foster et al. (2008).

9. The construction of the index is complicated by the fact that new four-digit sectors emerged over the course of the sample period. This situation prevents us from linking to the reference productivity level for each year. Where this occurs, the reference productivity level for the two-digit sector as a whole is used. The reduced sample that does not use this correction yields very similar results (available on request).

Firm-specific productivity scores are used to compute a measure of productivity for each subsector in each year in equation (2).

$$wpr_{jt} = \sum_{i=1}^{n} w_{ijt}\omega_{ijt}, \tag{2}$$

where $w_{ijt} = \sum_{i=1}^{n}\left(y_{ijt} \Big/ \sum_{i=1}^{n} y_{ijt}\right)$.

Equation (3) shows how this weighted productivity measure can be decomposed into the average unweighted productivity level of each subsector $\bar{\omega}_{jt} = \frac{1}{n}\sum_{i=1}^{n}\omega_{ijt}$ and a term that captures how the allocation of resources contributes to productivity (Olley and Pakes 1996).

$$wpr_{jt} = \omega_{jt} + \sum_{i=1}^{n}\left(w_{ijt} - \bar{w}_{jt}\right)\left(\omega_{ijt} - \bar{\omega}_{jt}\right). \tag{3}$$

This measure is used to consider the proportional contribution of switchers to aggregate productivity growth and to compare productivity between incumbents, entrants, exits, and switchers.

Empirical Model

Productivity is critical to firm decision making, but many other firm characteristics and sector-specific characteristics can be expected to play a role in switching behavior. Equations (4a) and (4b) show the empirical models of the decisions to switch into a sector and out of a sector, respectively.

$$\Pr\left(switch\_IN_{ijt}\right) = f\begin{pmatrix} \omega_{it}, k_{it}/l_{it}, \log(l_{it}), SOE_{it}, FOE_{it}, multiprod_{it}, \\ \omega_{jt}, k_{jt}/l_{jt}, \log(l_{jt}), CR_{jt}, FR_{jt}, SR_{jt}, TR_{jt}, EX_{jt} \\ \gamma_i, \theta_j, \tau_t, \vartheta_p \end{pmatrix}, \tag{4a}$$

$$\Pr(switch\_OUT_{imt}) = f\begin{pmatrix} \omega_{it}, k_{it}/l_{it}, \log(l_{it}), SOE_{it}, FOE_{it}, multiprod_{it}, \\ \omega_{mt}, k_{mt}/l_{mt}, \log(l_{mt}), CR_{mt}, FR_{mt}, SR_{mt}, TR_{mt}, EX_{mt} \\ \gamma_i, \theta_m, \tau_t, \vartheta_p \end{pmatrix}, \tag{4b}$$

where ω_{it} is the productivity of firm i in period t, k_{it}/l_{it} is the capital-labor ratio, $\log(l_{it})$ is the log number of employees, SOE_{it} is a dummy indicator for whether the firm is state owned, FOE_{it} is a dummy indicator for whether the firm is foreign owned, and $multiprod_{it}$, is a dummy indicator for whether the firm

produces more than one product. The subscript j in equation (4a) refers to the new sector entered, and the subscript m in equation (4b) refers to the old sector exited. For example, ω_{jt} is the average productivity level in sector j that the firm switches into in period t, and ω_{mt} is the average productivity level in sector m that the firm switches out of in period t. Other sector-specific variables include the following: k_{jt}/l_{jt} refers to the average capital-labor ratio; $\log(l_{jt})$ is the log average size of firms; CR_{jt} is the concentration ratio; FR_{jt} is the concentration of foreign-owned firms; SR_{jt} is the concentration of state-owned firms; TR_{jt} is the tariff rate; and EX_{jt} is the level of exports. Additionally, γ_i represents firm-specific fixed effects, θ_j and θ_m are sector-specific fixed effects, τ_t represents time dummies, and ϑ_p represents province fixed effects.

The decisions to enter, exit, or switch sectors are modeled in a similar fashion. For entry, as in equation (4a), the characteristics of the new sector are included, and for exit, as in equation (4b), the characteristics of the former sector are used. The inclusion of firm-specific fixed effects eliminates any time invariant unobserved heterogeneity influencing firms' decisions. As such, the identification of the effect of the firm-specific effects included in the model comes from the within-firm variation in firm characteristics. It is possible that some time-varying factors remain unobserved, so the (i, t) indexed variables are endogenous to firm behavior and therefore to the entry, exit, and switching outcomes. Consequently, no claim is made that results are causal.[10] Before presenting the data in section III, the motivation for the choice of firm-, sector-, and industry-specific variables included in the model is explained along with expectations regarding their influence.

Firm-Specific Factors

First, a firm's productivity level is a critical factor in decisions regarding whether to stay in production, switch sectors, or exit production altogether.[11] Firms evaluate their expected current and future profits on the basis of their own observed productivity level.

Second, capital accumulation is a key mechanism for increasing profitability. Thus, capital-intensive firms should increase profitability over time and should be less likely to exit and more likely to enter as an industry evolves. Bernard et al. (2006) have found that a firm's capital-labor ratio is an important determinant of decisions to switch sectors in the United States. In Vietnam, where labor-intensive firms arguably have a comparative advantage, labor-intensive firms are unlikely to have a higher probability of exit.

10. The authors are grateful to Alain de Janvry and Elisabeth Sadoulet for emphasizing this point.

11. Because there are difficulties in comparing productivity levels across sectors, the firm's rank in the productivity distribution is considered an alternative measure.

Third, the link between firm size and probability of survival has long been considered important.[12] This link may affect a firm's decision to switch sectors because larger firms may find it more difficult to retrain employees. To capture the effect of this factor, firm size, measured by number of employees, is included.

Fourth, arguably, ownership structure influences firm decision making in Vietnam, even after firm productivity, capital accumulation, and size are controlled for. The political hierarchy in the management structures of Vietnamese state-owned enterprises (SOEs) is likely to limit intersector dynamics; consequently, SOEs are unlikely to switch sectors. In contrast, the ongoing reform/privatization process has led to the dismantling of many SOEs, so one might expect a positive association between state ownership and firm closure. Foreign-owned enterprises (FOEs), or enterprises with some foreign participation, are also expected to be more "locked into" specific sectors because of legal constraints.[13]

Fifth, Bernard et al. (2010) and Goldberg et al. (2010) explain reallocations that enhance productivity through changes in product mix. Thus, multiproduct firms will be in a better position to churn products (i.e., dropping inefficient products to produce more efficient products) in response to changes in the economic environment and thus may be less likely to switch sectors.

Sector-Specific Factors

We will now consider a range of sector-specific measures for establishing the competitiveness of a sector relative to other sectors in the industry.

First, higher levels of average productivity in a sector make it more difficult for firms to compete. Because of difficulties in comparing productivity across sectors, we also consider the dispersion in productivity distribution a unit-neutral measure of productivity. It is easier to survive in sectors with a larger dispersion in productivity because low productivity levels are more likely to be tolerated. Accordingly, it is expected that firms are less likely to leave, and more likely to enter, sectors with a wider dispersion of productivity distribution.

Second, the average capital intensity of the sector, as measured by the capital-labor ratio, may be influential in determining a sector's competitiveness. Audretsch (1991) found that firm survival is much less likely when there is a high capital-labor ratio, but Bernard et al. (2006) found that firms in

12. Firm age is an important predictor of firm survival, but the data do not identify this characteristic. The inclusion of firm-specific fixed effects controls for any initial differences in the survival or switching probability of firms that are attributable to differences in age at the start of the sample period.

13. Until recently, foreign and domestic investors were governed by two separate laws. A new investment law came into effect in July 2006 (CIEM 2006). It aims to equalize opportunities for domestic and foreign investors.

labor-intensive sectors switched into more capital-intensive sectors when exposed to competitive pressures from imports.

Third, the average size of firms within a sector may affect switching decisions. It is more difficult for entering and switching firms to compete in a sector where the size of the average firm is large, given the economies of scale already enjoyed by incumbents. Consequently, it is expected that firms are more likely to enter or switch into sectors where firms are smaller, on average.

Fourth, the sector concentration ratio (CR), defined as the ratio of the accumulated revenue of a sector's four largest firms to total revenue in the sector, is likely to be influential because it is a proxy for sector competition. Siegfried and Evans (1994) document that a high CR may strengthen collusion efforts among incumbent firms and increase the likelihood that firms will attempt to prevent entry and maintain higher expected profits. In contrast, Audretsch (1991) shows that a high CR helps the survival rates of new entrants in the short run. On balance, it is likely that a high CR reduces incentives to move out of a given sector and is indicative of barriers to entry.

Industry-Specific Factors

Changes in exogenous conditions may "shock" enterprises, leading to different productivity outcomes. Trade liberalization, in the form of a tariff reduction, may lead to low-productivity firms exiting sectors or switching to sectors that remain protected. Because trade liberalization may lead to opportunities in new export markets, however, it is hypothesized that more productive firms will switch to sectors where these opportunities emerge.

In the manufacturing industry in Vietnam, it is critical to explore the impact of the privatization of SOEs and the entry of FOEs into the market. Where an entire industry is undergoing deregulation, a significant amount of productivity-enhancing reallocation is likely to take place. The dominance of state enterprises (SR), as measured by the share of total sector output controlled by SOEs, is likely to play a role in exit and switching decisions. Preferential treatment of SOEs makes it difficult for non-SOEs to compete and may force efficient non-SOE firms to exit (or to decide not to enter) industries with high concentrations of SOEs.

In contrast, during the transition from a planned to a market economy, the SOE share of material inputs bought at market conditions may increase the attractiveness of industries with high concentrations of SOEs for smaller, private enterprises acting as producers of intermediates for SOEs, as suggested by Jefferson and Rawski (1994) in their study of China. An added dimension is that deregulating SOEs may increase competition as a result of the decline in the level of protection and barriers to entry, thereby inducing firms to switch.

Similar arguments apply when considering the dominance of foreign enterprises (FR) in a sector, or the FOE share of total sector output. Aitkin and Harrison (1999) emphasize that preferential treatment of foreign-owned firms may distort competition, forcing out equally efficient domestically owned

counterparts. However, governments grant special treatment to FOEs to promote technology transfer, so FR may create a basis for domestically owned firms to produce intermediate inputs, as is the case with SOEs. Whether FR is positively or negatively related to sector switching and firm exit depends on which of these contrasting effects dominate.

III. Data

The data originate from the 2001–2008 Enterprise Surveys collected annually by the General Statistics Office of Vietnam. These surveys include all enterprises with 30 employees or more registered in Vietnam, a representative sample of smaller firms,[14] and all firms whose main activity is in the manufacturing sector.[15] Therefore, the sample includes approximately 50,000 firms.

The extent of diversification within the manufacturing sector is substantial, as illustrated by the number of four-digit subsectors within each two-digit sector (table 2). There are no data available to examine product diversification at the six-digit level, as in Bernard et al. (2010) and Goldberg et al. (2010), but we know that between 5 and 19 percent of manufacturing firms produce more than one product. For these firms, diversification across a product mix is a potential source of productivity growth that cannot be captured in this analysis. However, changes in the product mix that make one activity relatively more important than another when defined at aggregated levels are captured by the definition of switching. The factors that determine the decision of multiproduct firms to change their product mix are likely to differ from those that determine the decision of single-product firms to completely switch sectors, making this an important control variable in the analysis.

The manufacturing sector in Vietnam is characterized by significant enterprise dynamics (table 2). This characterization is consistent with much of the existing literature on firm dynamics that finds a positive correlation between exit and entry rates at the sector level (Dunne et al. 1988, 1989; Disney et al. 2003; Roberts and Tybout 1996). Following Bernard et al. (2006), our focus here is

14. Trade and tariff data are from the World Integrated Trade Solutions database. Trade data at the four-digit level for Vietnam with the rest of the world are taken from the UN COMTRADE database. Tariff data refer to the four-digit weighted average Most Favored Nation tariff applied to imports collected from the UNCTAD-TRAINS database for all imports into Vietnam. For further details on the data and descriptive statistics, see Newman et al. (2011).

15. One caveat of the data is that they do not indicate whether an enterprise has more than one establishment, so it cannot be determined whether multiestablishment enterprises differ from single-establishment enterprises in terms of their switching behavior. Moreover, given that only a representative sample of small firms is included, it is possible that the entry, exit, and switching behavior of small firms is not fully observed. As a robustness check, firms with 30 employees or more are also analyzed in isolation.

TABLE 2. Sector Diversification and Dynamics

	Number of four-digit subsectors	Multiproduct (%)	Exits (%)	Entrants (%)	Switch (four-digit) (%)	Switch excl. multiprod. (four-digit) (%)
15 Food products and bev.	21	10.83	49.63	68.35	13.42	12.48
16 Tobacco products	3	19.05	42.86	33.33	68.57	67.86
17 Textiles	8	8.26	46.06	78.01	27.83	26.13
18 Wearing apparel	3	5.09	53.35	84.36	6.44	5.90
19 Tanning/ dressing leather	4	7.98	48.02	74.98	13.07	12.00
20 Wood and wood products	8	12.79	53.05	81.13	19.51	16.57
21 Paper and paper products	3	9.10	48.15	77.67	24.64	22.79
22 Publishing, printing, etc.	7	5.06	49.70	88.60	15.17	14.33
23 Coke, refined petroleum	4	7.14	55.71	82.86	14.71	14.29
24 Chemicals and chem. prod.	9	11.47	47.92	78.59	15.28	13.60
25 Rubber and plastics	4	8.47	42.22	80.92	15.14	14.08
26 Other nonmetallic mineral	9	16.77	49.64	69.16	15.43	13.96
27 Basic metals	4	16.58	44.19	81.80	26.27	24.47
28 Fabricated metals	8	10.11	48.89	86.88	26.72	24.03
29 Machinery and equipment	15	9.40	48.52	81.31	32.26	29.57
30 Office equipment	4	6.09	61.74	93.91	23.01	22.64
31 Electrical machinery	7	12.12	51.01	79.60	28.12	25.82
32 Radio, television, etc.	3	7.23	51.98	82.10	19.61	17.15
33 Medical, precision and opt.	5	7.48	47.64	80.31	19.37	18.80
34 Motor vehicles, transport	4	13.58	57.21	72.96	28.46	25.88
35 Other transport equip.	8	17.79	46.54	73.91	35.61	33.30
36 Furniture	18	9.86	52.77	82.98	17.13	16.16
37 Recycling	2	10.96	42.47	90.41	20.83	19.53

Source: Authors' calculations using Vietnam Enterprise Surveys 2001–2008.

Note: Statistics are based on all enterprises operating in 2008 whose main output was in the manufacturing sector at some stage between 2001 and 2008.

exclusively on four-digit switching, which ranges from 6 to 35 percent.[16] Moreover, the majority of this switching is attributable to single-product firms.

In comparison, Bernard et al. (2006) find that approximately 8 percent of U.S. manufacturing firms switched activities during five-year periods between 1977 and 1997. The fact that switching is common in Vietnam is indicative of an evolving industrial sector where new opportunities are emerging as a result of deregulation, trade liberalization, and on-going structural transformation. The extent of observed switching behavior adds further weight to the argument that the effects of switching on productivity should be separated from the effects of standard exit/entry. A better understanding of the forces responsible for switching may be helpful in designing effective economic policy in developing countries.

IV. Empirical Results

Our empirical results are grouped under the following four headings: productivity growth, switching, firm entry and exit, and robustness checks.

Productivity Growth

Productivity is first estimated for each subsector of the manufacturing industry. Output is the total revenue of the firm deflated by the two-digit sector-level GDP deflator. Inputs are composed of the following elements: (i) labor, measured as the total number of persons employed at the end of the year; (ii) capital, measured as the total assets of the firm at the end of the year deflated by a capital price series; and (iii) other costs of production deflated by the two-digit sector-level producer price index. The cost of labor is the firm's reported wage bill deflated by the producer price index, and the cost of capital is charged at the average annual commercial bank lending rate in each year plus an estimated depreciation rate of 2 percent per annum.[17]

Each sector's trend in productivity from 2001 to 2008 is computed using the index number approach outlined in section II.[18] In all sectors in all years, the covariance between output and productivity is positive, indicating that more productive firms account for a larger share of output (table 3). The size of the covariance term does not change significantly over time, suggesting that the main source of productivity growth can be traced to changes in the productivity level of firms rather than to increases in the concentration of output in

16. As many as 68 percent of firms switched activity across four-digit sectors in the tobacco products sector (table 2), but this sector consists of very few firms and is not included in the main analysis.

17. Summary statistics for each of the variables are provided in table A1.

18. The figures are computed based on a two-digit level of aggregation for presentation purposes. The results are very similar when productivity is measured at the four-digit level and using the Olley and Pakes (1996) approach for each two-digit subsector.

TABLE 3. Weighted Productivity Estimates and Decomposition (Manufacturing Sample)

Activity 15	*n*	Weighted	Covariance	Activity 17	*n*	Weighted	Covariance
2001	2,715	.174	.115	2001	385	.299	.093
2002	2,962	.198	.124	2002	471	.301	.118
2003	3,046	.215	.140	2003	551	.643	.120
2004	3,280	.494	.149	2004	635	.495	.127
2005	3,665	.530	.163	2005	784	.528	.117
2006	4,383	.587	.162	2006	957	.601	.179
2007	4,506	.385	.153	2007	1,058	.374	.124
2008	5,306	.721	.142	2008	1,293	.288	.086
Activity 18	*n*	Weighted	Covariance	Activity 19	*n*	Weighted	Covariance
2001	522	.413	.146	2001	236	.203	.098
2002	766	.487	.172	2002	290	.258	.165
2003	934	.363	.156	2003	316	.281	.095
2004	1,206	.466	.166	2004	395	.360	.130
2005	1,320	.537	.149	2005	453	.389	.072
2006	1,548	.525	.150	2006	441	.479	.109
2007	1,783	.338	.096	2007	524	.340	.106
2008	2,580	.568	.071	2008	659	.387	.039
Activity 20	*n*	Weighted	Covariance	Activity 21	*n*	Weighted	Covariance
2001	724	.116	.118	2001	429	.107	.075
2002	818	.173	.111	2002	494	.187	.072
2003	914	.486	.104	2003	596	.241	.074
2004	1,096	.354	.111	2004	689	.210	.087
2005	1,261	.440	.113	2005	871	.261	.089
2006	1,440	.387	.087	2006	960	.185	.077
2007	1,755	.411	.100	2007	1,033	.257	.085
2008	2,366	.554	.091	2008	1,301	.598	.062
Activity 22	*n*	Weighted	Covariance	Activity 24	*n*	Weighted	Covariance
2001	333	.250	.136	2001	428	.280	.101
2002	454	.271	.182	2002	509	.323	.142
2003	580	.571	.190	2003	598	.538	.149
2004	820	.496	.202	2004	688	.579	.189

2005	1,023	.536	.182	2005	840	.673	.200
2006	1,564	.466	.170	2006	1,026	.563	.200
2007	1,457	.310	.182	2007	1,120	.442	.167
2008	1,890	.754	.139	2008	1,376	.674	.155
Activity 25	*n*	Weighted	Covariance	Activity 26	*n*	Weighted	Covariance
2001	523	.154	.077	2001	1,044	.680	.672
2002	683	.182	.104	2002	1,069	.336	.257
2003	769	.420	.096	2003	1,122	.476	.232
2004	971	.371	.105	2004	1,282	.547	.206
2005	1,222	.499	.105	2005	1,425	.587	.206
2006	1,427	.404	.098	2006	1,479	.558	.182
2007	1,664	.185	.082	2007	1,641	.462	.187
2008	2,000	.556	.051	2008	2,004	.667	.178
Activity 27	*n*	Weighted	Covariance	Activity 28	*n*	Weighted	Covariance
2001	142	.180	.139	2001	722	.157	.134
2002	185	.299	.130	2002	941	.182	.133
2003	223	.329	.115	2003	1,239	.395	.122
2004	267	.363	.096	2004	1,611	.405	.111
2005	351	.408	.082	2005	2,057	.473	.119
2006	388	.331	.093	2006	2,551	.367	.114
2007	510	.363	.111	2007	3,009	.294	.101
2008	634	.090	.079	2008	3,984	.495	.091
Activity 29	*n*	Weighted	Covariance	Activity 31	*n*	Weighted	Covariance
2001	264	.296	.088	2001	167	.179	.087
2002	342	.306	.115	2002	195	.110	.118
2003	404	.285	.125	2003	238	.530	.123
2004	466	.604	.130	2004	316	.673	.207
2005	559	.451	.125	2005	372	.335	.130
2006	628	.439	.136	2006	398	.204	.132
2007	758	.337	.125	2007	385	.251	.110
2008	887	.249	.091	2008	498	.764	.076

(Continued)

TABLE 3. Continued

Activity 15	*n*	Weighted	Covariance	Activity 17	*n*	Weighted	Covariance
Activity 32	*n*	Weighted	Covariance	Activity 33	*n*	Weighted	Covariance
2001	76	.280	.109	2001	37	.408	.089
2002	101	.299	.213	2002	49	.428	.066
2003	125	.471	.206	2003	53	.213	.048
2004	151	.378	.246	2004	60	.595	.119
2005	169	.664	.201	2005	83	.249	.163
2006	187	.458	.150	2006	95	.245	.091
2007	229	.249	.133	2007	101	.551	.096
2008	299	.395	.227	2008	125	.450	.111
Activity 34	*n*	Weighted	Covariance	Activity 35	*n*	Weighted	Covariance
2001	182	.214	.293	2001	247	.362	.182
2002	222	.249	.267	2002	298	.308	.175
2003	222	.568	.269	2003	344	.521	.200
2004	247	.426	.219	2004	363	.529	.207
2005	301	.625	.193	2005	452	.587	.195
2006	239	.514	.146	2006	452	.484	.179
2007	275	.331	.164	2007	558	.483	.195
2008	348	.446	.178	2008	618	.914	.141
Activity 36	*n*	Weighted	Covariance				
2001	601	.175	.132				
2002	679	.307	.208				
2003	901	.618	.130				
2004	1,118	.512	.141				
2005	1,373	.515	.123				
2006	1,552	.475	.097				
2007	1,793	.405	.110				
2008	2,510	.542	.090				

Source: Authors' calculations using Vietnam Enterprise Surveys 2001–2008.

Note: Sectors 16, 23, 30, and 37 are excluded because of the small number of firms that operate in these sectors.

TABLE 4. Contribution of Switchers to Productivity before and after Switching

	Contribution of Switchers to TFP			Contribution of Switchers to WTFP		
	After switch	Before switch	*p* value *t*-test of difference	After switch	Before switch	*p* value *t*-test of difference
2002	14.41	14.40	.9954	21.41	22.76	.4436
2003	15.71	12.67	.0000	12.32	11.09	.0412
2004	11.35	8.97	.0000	9.87	7.23	.0000
2005	12.57	11.07	.0000	9.12	7.89	.0000
2006	24.64	21.53	.0000	20.77	19.63	.0002
2007	17.63	15.26	.0000	16.02	12.58	.0000
2008	5.53	4.51	.0000	3.90	3.13	.0000

Source: Authors' calculations using Vietnam Enterprise Surveys 2001–2008.

TABLE 5. Testing the Differences in Productivity Rankings

	Number of firms	Average rank in productivity distribution	*t*-statistic (difference = 0)	*p* value
Incumbents	84,942	.546		
Entry	21,438	.400	67.215	.000
Incumbents	76,967	.548		
Switchers-in	9,010	.512	11.219	.000
Incumbents	84,942	.546		
Exit firms	9,230	.436	34.740	.000
Incumbents	64,723	.546		
Switchers-out	6,849	.520	7.029	.000
Entry	21,438	.400		
Exit	9,230	.436	−10.929	.000
Entry	27,600	.393		
Switchers-in	10,813	.499	−34.559	.000
Exit	9,230	.436		
Switchers-out	9,532	.478	−10.324	.000

Source: Authors' calculations using Vietnam Enterprise Surveys 2001–2008.

Note: Switchers-in and entry firms are compared to firms in the receiving sector, whereas switchers-out and exit firms are compared to firms in the expelling sector.

more productive firms.[19] This does not mean, however, that reallocations of output within and across sectors do not contribute to productivity growth.

To determine whether switching is economically important, the proportional contribution of switchers to aggregate productivity in the sector that they switch out of is compared with their proportional contribution to the sector that they switch into (table 4). Moreover, productivity differences between incumbents, entrants, exits, and switchers are described in a manner that

19. This is similar to the findings of Aw et al. (2001) for Taiwanese manufacturing.

follows Aw et al. (2001) (table 5). This description facilitates the testing of the hypotheses outlined in section I.

Switching firms generally contribute significantly more to productivity in the sector that they switch into than in the sector that they left (table 4). This pattern holds for both weighted and unweighted TFP, with the only exception occurring between 2001 and 2002. Although the magnitude of the gains is not substantial, these gains are achieved in the year immediately following the switch, when firms have undoubtedly incurred some costs in switching activities.[20]

A firm fixed-effects regression for switching firms was conducted for the year of the switch, where firm TFP is regressed on dummy indicators controlling for time, industry, province, and sector-level productivity. These results, which are not reported here, confirm the findings presented in table 4 and suggest that there is a break in a positive direction after the switch.

The comparison of productivity performance between switching and other firms is not straightforward. Relying on productivity levels is not appropriate because, in the construction of the productivity index, firms are positioned relative to the average within the sector, making comparisons across sectors impossible to interpret. The same situation arises when we attempt to interpret growth rates in productivity levels for switching firms.

To overcome this obstacle, the productivity performance of each firm is ranked within each sector in each year. Simple *t*-tests are then performed of the differences between incumbent, switching, entry, and exit firms in terms of their productivity ranking. Incumbent firms rank higher in the productivity distribution than entry and exit firms on both measures (table 5). Similarly, exit firms rank higher than entry firms. These results are not surprising because entry firms may incur high sunk costs that negatively affect their productivity performance in the year of entry.[21]

The results also show that switchers-in rank higher in terms of productivity than entry firms in the sector into which they enter, and switchers-out rank higher than exit firms in the sector that they switch from. This result suggests that switching firms are indeed a separate and important source of productivity growth, which is consistent with the predictions in section I above and with the emerging literature on firm capabilities, which we have referenced. Finally, incumbents have a higher productivity ranking than switchers in the

20. When switching is defined at the two-digit level, the results are mixed; in some years, firms contribute more to the sector that they switch into than the sector that they switch out of, but in other years, the opposite is true. These mixed results are not surprising given that switching between two-digit sectors requires a significant change in production activities, and it may take time for firms to adjust. Overall, this finding suggests that there are more gains to aggregate productivity from firms that switch between more closely related sectors than at the two-digit level. The results are not presented here but are available on request.

21. When the performance of entry firms is compared to that of exit firms over their lifecycles, entry firms outperform exit firms. This finding is consistent with the extensive literature on firm turnover that emphasizes resource reallocations involving the exit and entry of firms into sectors as an important source of productivity growth.

sectors that they switch into and out of, which is also consistent with our expectations.[22]

Given the difficulties in comparing productivity levels across sectors and, consequently, of comparing productivity in switching firms before and after they switch sectors, it is not possible to conclude definitively that switching firms become more productive as a result of moving from one sector to another. Two observations may be made, however. First, switching firms contribute more to overall productivity in the sector that they switch into than in the sector that they switch from, suggesting that switching leads to a reallocation of resources that enhances productivity (table 4). Second, switching firms have higher productivity levels and rank higher in the productivity distribution than entry and exit firms in the same sector (table 5).

Switching

The first set of switching results relates to the decisions of firms to switch into and out of particular sectors (table 6). The former identifies sector-specific pull factors because sector characteristics in the year after the switch are included. The latter identifies sector-specific push factors because the sector characteristics in the year before the switch are included. In other words, although we are not estimating a dynamic model, the key focus is to identify the push and pull factors through the timing of their inclusion in the model. Switching and sector-specific variables are defined at the four-digit industry classification level, and sector-level variables are computed separately for each firm i by excluding information on firm i in the computation of the sector-level aggregates. These definitions and computations ensure that the individual characteristics of the switchers do not drive the sector-specific effects.

For firm-specific factors, switching firms rank higher in the productivity distribution of the sectors that they switch into and have higher productivity levels than other firms (i.e., after the switch), and they rank below incumbents and other firms in the sectors that they switch out of (i.e., before the switch). There is some evidence that switching firms are not the worst performing firms in the sectors that they switch out of (column 6), which is consistent with the hypothesis that switching firms are more productive than exit and entry firms. Switching firms have a relatively higher level of productivity compared to firms in the sector that they switch into (column 2) than firms in the sector that they switched out of (column 6), providing further evidence that firms that switch sectors manage to improve their relative performance. Thus, expectations seem to hold; switching sectors to exploit profitable opportunities is a viable alternative to exiting production altogether.

Switching firms tend to be larger than other firms in the sectors that they switch into and smaller than the firms in the sectors that they switch out of,

22. Comparing productivity levels across groups of firms leads to the same conclusions. Moreover, the results are similar when switching is defined at the two-digit level. Results are available on request.

which is consistent with Dunne et al. (1988). Given that much of the literature suggests a positive association between size and survival probability, this finding indicates that firms switch to sectors in which they are larger than other firms when compared to the firm sizes in the sectors that they left. These firms will therefore survive longer. There is also weak evidence that switchers tend to have a lower capital-labor ratio than nonswitching firms. Given the fact that switching firms are generally larger, labor-intensive firms are more likely to be switchers than are capital-intensive firms. Moreover, there is evidence that multiproduct firms registered in Vietnam are less likely to switch, in accordance with the findings of Bernard et al. (2010) and Goldberg et al. (2010), who show that multiproduct firms can change their product mix in response to changing market conditions without switching sectors.

For sector-specific factors, our results show that although firms registered in Vietnam switch between low productivity sectors, the magnitude of the coefficient is lower for the sectors that they switch into than for the sectors that they switch out of, suggesting that firms switch into lower productivity sectors (table 6). Switching firms tend to be more productive than the firms that are already operating in the sectors that they switch into, demonstrating that switching firms seek productivity enhancing opportunities when deciding to switch sectors. The average size of the sector, as measured by the number of employees, and the average capital-labor ratio play important roles. Firms switch out of sectors that have high capital-labor ratios where they are smaller in size (push factors) and into sectors with low average capital-labor ratios (pull factors) where they are relatively larger. Because the analysis is set in the context of a developing country with low labor costs, it is not surprising that firms switch into sectors where they can potentially exploit low labor cost advantages. Moreover, because our earlier results indicated that switchers are more likely to be labor-intensive firms, it is not surprising for these firms to switch into labor-intensive sectors.

For market structure variables, switching is less likely to occur into sectors with a large proportion of FOEs (i.e., a high FR). This result is consistent with previous literature showing that firms avoid sectors dominated by FOEs (Aitkin and Harrison 1999; Tybout 2000).[23] Firms are also more likely to switch into sectors with lower CRs. It might be expected that a high CR would reduce firms' incentives to move out of a sector, but firms are more likely to switch

23. When switching is defined at the two-digit level, firms switch out of sectors with low levels of foreign ownership and into sectors with high levels of foreign ownership, which is different than the results presented here under a more disaggregated definition of industry switching. This is indicative of a general move by firms toward more competitive sectors at an aggregate level, presenting the possibility that the presence of learning spillover effects and technology diffusion will make a sector attractive to domestic enterprises. The results at the four-digit level suggest that switching firms choose not to compete directly with foreign-owned firms that produce similar products but that within two-digit sectors, firms seek opportunities to benefit from vertical technology spillovers from foreign-owned firms. For further details, see Newman et al. (2011).

TABLE 6. Determinants of Switching

	Switch in				*Switch out*			
Firm-specific	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Productivity	.018***	.039***		.039***	.005	.021***		.021
	(.005)	(.005)		(.015)	(.005)	(.005)		(.013)
Rank productivity			.030***				−.014*	
			(.007)				(.008)	
lnK/L	.000	.001	−.001	.001	−.002	−.001	−.003*	−.001
	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)
Size	.010***	.010***	.006**	.010***	−.006**	−.005*	−.006**	−.005**
	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)
State-owned	.014	.018*	.018*	.018	.004	.009	.008	.009
	(.011)	(.011)	(.011)	(.012)	(.012)	(.011)	(.011)	(.011)
Foreign-owned	.075*	.052	.050	.052	.002	−.027	−.027	−.027
	(.042)	(.040)	(.040)	(.035)	(.058)	(.055)	(.055)	(.040)
Multiproduct	.010*	.011*	.010*	.011**	−.010*	−.017***	−.016***	−.017***
	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)
Sector-specific								
Productivity[a]		−.020***		−.020***		−.007***		−.007*
		(.004)		(.007)		(.002)		(.004)
IQR productivity[b]			−.011***				−.002***	
			(.003)				(.001)	
lnK/L[a]		−.037***	−.032**	−.037		.056***	.052***	.056*
		(.013)	(.013)	(.037)		(.013)	(.013)	(.031)
Size[a]		−.010	−.003	−.010		−.042***	−.044***	−.042**
		(.007)	(.007)	(.018)		(.007)	(.007)	(.017)
CR[a]		−.114***	−.121***	−.114		.047	.037	.047
		(.030)	(.030)	(.079)		(.032)	(.032)	(.082)
FR[a]		−.061**	−.069***	−.061		−.022	−.024	−.022

(Continued)

TABLE 6. Continued

		Switch in				*Switch out*		
		(.025)	(.025)	(.070)		(.025)	(.025)	(.072)
SR[a]		.012	.011	.012		−.014	−.013	−.014
		(.020)	(.020)	(.048)		(.021)	(.021)	(.056)
Tariff level[c]		−.250***	−.237***	−.250***		.186***	.193***	.186
		(.025)	(.025)	(.087)		(.045)	(.044)	(.144)
Export intensity[c]		.022***	.021***	.022***		.004	.004	.004
		(.003)	(.003)	(.008)		(.003)	(.003)	(.008)
Observations	136,235	126,683	126,683	126,683	115,189	107,323	107,323	107,323
Firms	49,409	46,790	46,790	46,790	39,820	37,968	37,968	37,968
Clustering	Firm	Firm	Firm	Industry	Firm	Firm	Firm	Industry

Source: Authors' calculations using Vietnam Enterprise Surveys 2001–2008.

Notes: All models include firm, sector (four-digit), time, and provincial fixed effects. Robust standard errors clustered at the firm/industry level reported in parenthesis. *, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.

[a] Average for each four-digit sector computed separately for each firm *i* excluding information on firm *i*.

[b] Interquartile range for the four-digit sector.

[c] Average for each four-digit sector.

TABLE 7. Determinants of Entry and Exit

	Entry				*Exit*			
Firm-specific	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Productivity	−.063***	−.092***		−.092***	−.003	−.004		−.004
	(.004)	(.005)		(.010)	(.003)	(.004)		(.006)
Rank productivity			−.242***				.004	
			(.008)				(.006)	
lnK/L	.015***	.013***	.014***	.013***	.000	.000	.001	.000
	(.002)	(.002)	(.002)	(.002)	(.001)	(.002)	(.002)	(.001)
Size	−.046***	−.047***	−.027***	−.047***	−.025***	−.026***	−.026***	−.026***
	(.003)	(.003)	(.003)	(.003)	(.002)	(.002)	(.002)	(.002)
State-owned	−.073***	−.071***	−.067***	−.071***	.048***	.047***	.047***	.047***
	(.010)	(.010)	(.010)	(.011)	(.008)	(.008)	(.008)	(.007)
Foreign-owned	−.168***	−.154***	−.147***	−.154***	.035	.036	.036	.036
	(.043)	(.044)	(.040)	(.038)	(.031)	(.033)	(.033)	(.026)
Multiproduct	.009**	.009*	.015***	.009**	.006	.005	.005	.005
	(.005)	(.005)	(.005)	(.004)	(.004)	(.004)	(.004)	(.004)
Sector-specific								
Productivity[a]		.015***		.015***		.003		.003
		(.002)		(.003)		(.002)		(.002)
IQR productivity[b]			.000				.002***	
			(.001)				(.001)	
lnK/L[a]		.003	.014	.003		.006	.006	.006
		(.010)	(.009)	(.022)		(.008)	(.008)	(.013)
Size[a]		.080***	.044***	.080***		−.006	−.006	−.006
		(.006)	(.006)	(.025)		(.005)	(.005)	(.012)
CR[a]		−.055**	−.035**	−.055		.008	.009	.008
		(.024)	(.023)	(.050)		(.020)	(.020)	(.037)
FR[a]		−.111***	−.086***	−.111**		.056***	.056***	.056*
		(.020)	(.020)	(.056)		(.016)	(.016)	(.032)
SR[a]		.000	−.007	.000		.025*	.025*	.025
		(.017)	(.017)	(.049)		(.014)	(.014)	(.027)
Tariff level[c]		−.034	−.078***	−.034		.011	.009	.011
		(.026)	(.026)	(.078)		(.014)	(.014)	(.048)
Export intensity[c]		−.011***	−.011***	−.011*		.003*	.003*	.003

(Continued)

TABLE 7. Continued

	Entry				Exit			
		(.003)	(.003)	(.006)		(.002)	(.002)	(.004)
Observations	146,058	135,976	135,976	135,976	146,058	135,976	135,976	135,976
Firms	50,807	48,179	48,179	48,179	50,807	48,179	48,179	48,179
Clustering	Firm	Firm	Firm	Industry	Firm	Firm	Firm	Industry

Source: Authors' calculations using Vietnam Enterprise Surveys 2001–2008.

Notes: All models include firm, sector (four-digit), time, and provincial fixed effects. Robust standard errors clustered at the firm/industry level reported in parenthesis. *, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.

[a] Average for each four-digit sector computed separately for each firm *i* excluding information on firm *i*.

[b] Interquartile range for the four-digit sector.

[c] Average for each four-digit sector.

TABLE 8. Determinants of Switching – Robustness Checks

	Switch in						*Switch out*					
	Large firms	Private firms	Excluding multiprod.	Excluding re-entrants	Relative to optimum	Olley & Pakes	Large firms	Private firms	Excluding multiprod	Excluding re-entrants	Relative to optimum	Olley & Pakes
Firm-specific	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Productivity	.020***	.055***	.036***	.037***			.014**	.031***	.017***	.018***		
	(.006)	(.006)	(.005)	(.005)			(.006)	(.006)	(.005)	(.005)		
Rank productivity					.033***	.010*					−.010	−.008
					(.008)	(.006)					(.008)	(.007)
lnK/L	.003	.001	.000	.001	−.002	−.001	.002	.000	−.001	.000	−.003	−.002
	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)	(.003)	(.002)	(.002)	(.002)	(.002)	(.003)
Size	.006*	.013***	.011***	.011***	.006**	.011***	−.006*	−.003	−.003	−.005*	−.006**	−.009**
	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.004)
State-owned	.015		.020*	.020*	.018*	.019	.004		.021*	.004	.008	−.007
	(.011)		(.012)	(.011)	(.011)	(.013)	(.011)		(.012)	(.011)	(.011)	(.014)
Foreign-owned	.054		.056	.055	.049	.009	−.027		−.027	−.029	−.027	−.006
	(.041)		(.043)	(.041)	(.040)	(.046)	(.054)		(.058)	(.057)	(.055)	(.039)
Multiproduct	.015**	.010*		.010*	.010*	.018**	−.010*	−.021***		−.016***	−.016***	−.015**
	(.006)	(.006)		(.006)	(.005)	(.007)	(.006)	(.006)		(.006)	(.005)	(.007)
Sector-specific												
Productivity[a]	−.006*	−.013***	−.013***	−.020***	.004**	.044	−.005***	−.002	−.006**	−.006***	−.006***	.029
	(.003)	(.003)	(.004)	(.004)	(.002)	(.058)	(.002)	(.002)	(.002)	(.002)	(.002)	(.067)
lnK/L[a]	−.044***	−.048***	−.031**	−.038***	−.034***	−.020	.052***	.070***	.045***	.053***	.051***	.080***
	(.015)	(.015)	(.013)	(.013)	(.013)	(.016)	(.015)	(.016)	(.013)	(.013)	(.013)	(.016)
Size[a]	−.005	−.008	−.012*	−.007	−.007	−.005	−.040***	−.050***	−.044***	−.043***	−.043***	−.044***
	(.008)	(.008)	(.007)	(.007)	(.007)	(.008)	(.008)	(.009)	(.007)	(.007)	(.007)	(.009)
CR[a]	−.113***	−.069*	−.142***	−.120***	−.128***	−.115***	.081**	.054	.057*	.056*	.026	.090**
	(.037)	(.037)	(.032)	(.031)	(.030)	(.037)	(.039)	(.039)	(.034)	(.033)	(.032)	(.040)
FR[a]	−.004	−.075**	−.064**	−.056**	−.059**	−.066**	.013	−.023	−.013	−.017	−.030	−.002
	(.028)	(.030)	(.026)	(.025)	(.025)	(.031)	(.028)	(.030)	(.025)	(.025)	(.025)	(.031)

(Continued)

TABLE 8. Continued

	Switch in						*Switch out*					
	Large firms	Private firms	Excluding multiprod.	Excluding re-entrants	Relative to optimum	Olley & Pakes	Large firms	Private firms	Excluding multiprod	Excluding re-entrants	Relative to optimum	Olley & Pakes
SR[a]	.014	.023	−.007	.013	.012	−.014	−.005	−.010	−.003	−.012	−.015	−.020
	(.024)	(.024)	(.022)	(.021)	(.020)	(.025)	(.025)	(.026)	(.022)	(.022)	(.021)	(.026)
Tariff level[b]	−.166***	−.314***	−.244***	−.260***	−.233***	−.239***	.145**	.123**	.199***	.204***	.187***	.141**
	(.029)	(.030)	(.027)	(.026)	(.025)	(.032)	(.056)	(.053)	(.047)	(.045)	(.045)	(.056)
Export intensity[b]	.013***	.025***	.023***	.023***	.022***	.020***	−.001	.006	.006*	.003	.004	.006
	(.003)	(.004)	(.003)	(.003)	(.003)	(.004)	(.003)	(.004)	(.003)	(.003)	(.003)	(.004)
Observations	69,717	103,918	117,262	123,504	126,683	93,230	62,692	86,588	98,767	104,337	107,323	77,907
Firms	19,356	41,125	45,360	45,665	46,790	41,664	17,845	32,725	36,630	36,874	37,968	33,758

Source: Authors' calculations using Vietnam Enterprise Surveys 2001–2008.

Notes: All models include firm, sector (four-digit), time, and provincial fixed effects. Robust standard errors clustered at the firm/industry level reported in parenthesis. *, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.

[a] Average for each four-digit sector computed separately for each firm *i* excluding information on firm *i*.

[b] Interquartile range for the four-digit sector.

into sectors that are more competitive than the sectors that they leave. Therefore, switching firms may seek less regulated sectors, suggesting that more competitive sectors with lower levels of concentration attract firms. It should be noted, however, that the statistical significance of the results for foreign-ownership and sector-level concentration is not robust to clustering the standard errors at the industry level (column 4).

We have identified strong evidence that firms switch into sectors with low tariff levels and high levels of trade exposure (pull factors) and out of sectors with high tariff levels (a push factor). The results show that firms are willing to switch to different four-digit sectors to exploit potential opportunities from trade reform.

Firm Entry and Exit

There is a significant amount of heterogeneity in the characteristics of entry firms (table 7). The results show that entry firms have lower productivity levels than incumbents and firms that switch into a sector. Entry firms are more capital intensive and usually have fewer employees than other firms. They are more likely to be private domestic firms as opposed to state- or foreign-owned firms. These patterns reflect the deregulation in the manufacturing sector over the analysis period, which provided many new opportunities for private domestic firms. There is also some evidence that entry firms are more likely to be multiproduct firms, implying that a new type of flexible enterprise has emerged.

Consistent with the literature on firm survival, smaller firms are more likely to exit production (table 7). Exit is also associated in a statistically significant way with state ownership and is rooted in the ongoing reform process discussed by the Central Institute of Economic Management (2003). It is clear from this analysis that the firm-specific characteristics of entry and exit firms are very different from those of switching firms.

In summary, our entry and exit results are consistent with the existing literature on industry evolution, but the expanded analysis in this study, which has identified switching as separate from exit and entry, suggests an unexplored dimension in the context of developing countries.

Further differences are embedded in sector-specific characteristics. The results show that entry firms are more likely to enter sectors with high productivity levels and with larger average firm sizes than sector switchers. New firms are also less likely to enter into trade-intensive sectors than switchers, who are more likely to seek trade opportunities. Similar to switching firms, entry firms are deterred by high CRs and are less likely to enter sectors with a high concentration of foreign-owned firms. The latter finding is consistent with the idea that preferential treatment of foreign-owned firms might distort competition by deterring domestic firms from entering sectors with a high foreign ownership presence (Aitkin and Harrison 1999; Tybout 2000). Finally, we found evidence to indicate that both entry and switching firms are less likely to enter sectors

with high tariff levels, which may mean that barriers to entry exist in these sectors.

Robustness Checks

To check robustness, we first consider a subset of data that includes large manufacturing firms (defined as having more than 30 employees) for which a full population of firms is available. Second, a subsample of private firms is studied. Third, all multiproduct firms are excluded because productivity measurements are complicated for firms that produce more than one product. Fourth, firms that exit and re-enter the sample are omitted. Fifth, we consider two alternative measures of productivity. The first is an index number approach based on a relative measure of productivity that compares firms with the best-producing firm in a sector. The second measure was constructed using the Olley and Pakes (1996) semiparametric approach to productivity measurement. To enhance comparability in the latter two robustness checks, a firm's rank in the productivity distribution is used rather than the firm's productivity level. The findings from these robustness checks are broadly consistent across all specifications (table 8).

The most notable results are that switching firms (i) are more productive than other firms in their new sectors, (ii) are larger in terms of the number of employees, (iii) are unlikely to switch out of a sector if they are multiproduct firms, (iv) are likely to switch into sectors with lower average levels of productivity, (v) are more likely to switch into sectors with low capital-labor ratios, and (vi) are more likely to switch into sectors with low tariff levels and high levels of export intensity. Overall, our key findings above are confirmed.

V. Conclusion

This study began by observing that sector switchers are likely to have different characteristics and behavior from "real" entry and exit firms. This subject has never been empirically studied or established in the context of a developing country. This is arguably an important omission in the existing literature on firm dynamics and in the understanding of the impact of firm turnover on both productivity and resource reallocation at the firm and sector level.

Using a unique panel data set from Vietnam that covers a large number of manufacturing firms for the 2001–2008 period, we found solid evidence that the reallocation of resources within and across sectors accounts for a significant proportion of total productivity growth in the Vietnamese manufacturing sector. Firm switchers make an important contribution to these reallocations. They contribute more to productivity in the sector that they switch into than the sector that they switch out of, and they are more productive than firms entering or exiting the sector.

The analysis also revealed that switching firms have characteristics that differ from both entry and exit firms, and switching firms appear to be

motivated by different sector-specific factors. Switching firms tend to be larger than firms in the sector that they switch into, while entry firms tend to be smaller than other firms in the sector. Our analysis also shows that multiproduct firms are less likely to switch sectors.

Firms seek competitive opportunities when deciding which sectors to switch into, and they appear to avoid sectors with large concentrations of foreign firms. Moreover, trade liberalization is positively associated with switching behavior; firms are more likely to switch into sectors with better opportunities to trade (i.e., sectors with lower tariffs and a greater proportion of exported output).

Our analytical approach and the empirical findings of this study may serve as a starting point for similar analyses in other developing countries and may inspire a redesign of enterprise surveys across the developing world. The correct assessment of the respective contributions of entry, exit, and switching firms to productivity growth has important analytical and policy implications. At this stage, however, empirical evidence is almost nonexistent.

We can draw several conclusions regarding Vietnam from this work. First, there are much more complex firm dynamics underlying the robust economic progress achieved in this dynamic East Asian economy than those found in standard firm entry and exit explanations. Second, in practice, firms have adjusted to changing circumstances in sometimes innovative and not always predictable ways. Finally, government policy should pay careful attention to both the potential for productivity enhancement and the reduction of reallocations through sector switching. This policy should also help discourage firms from switching into sectors that are not associated with comparative advantage.

APPENDIX

TABLE A1. Descriptive Statistics for Productivity Analysis

	Revenue (million VND)	Labor (number employees)	Capital (million VND)	Other costs (million VND)
15 Food products and beverages	26,999	101	22,052	33,691
	(144,003)	(300)	(114,730)	(160,912)
17 Textiles	24,379	203	42,700	27,633
	(172,394)	(543)	(235,551)	(223,110)
18 Wearing apparel	13,322	342	15,006	15,279
	(47,825)	(708)	(43,741)	(66,379)
19 Tanning/dressing leather	46,321	1,096	56,457	49,528

(Continued)

TABLE A1. Continued

	Revenue (million VND)	Labor (number employees)	Capital (million VND)	Other costs (million VND)
	(202,041)	(3,369)	(235,526)	(227,280)
20 Wood and wood products	4,895	70	5,821	4,799
	(15,252)	(168)	(20,526)	(15,576)
21 Paper and paper products	11,917	72	17,144	10,958
	(41,428)	(165)	(78,593)	(35,798)
22 Publishing, printing, etc.	5,465	37	7,315	5,893
	(24,318)	(82)	(36,396)	(27,069)
24 Chemicals and chem. products	35,228	99	36,902	38,788
	(156,910)	(256)	(143,717)	(171,274)
25 Rubber and plastics	17,197	87	20,591	14,276
	(54,274)	(200)	(60,577)	(45,220)
26 Other nonmetallic mineral	18,933	137	40,637	17,796
	(99,188)	(268)	(249,302)	(123,819)
27 Basic metals	60,926	119	60,665	69,740
	(239,265)	(522)	(237,293)	(269,482)
28 Fabricated metals	10,328	54	12,625	14,003
	(42,842)	(143)	(46,959)	(59,604)
29 Machinery and equipment	13,514	90	20,692	15,788
	(53,221)	(210)	(75,188)	(63,813)
31 Electrical machinery	59,244	234	61,172	66,396
	(189,640)	(794)	(160,859)	(215,311)
32 Radio, television, etc.	64,853	199	70,867	91,420
	(189,703)	(493)	(173,636)	(280,609)
33 Medical, precision, and optical	26,029	155	32,231	35,781
	(186,179)	(388)	(87,192)	(289,953)
34 Motor vehicles, transport	59,747	127	64,975	91,141
	(283,758)	(241)	(260,270)	(456,288)
35 Other transport equipment	62,714	190	80,341	68,996
	(459,070)	(468)	(356,834)	(477,028)
36 Furniture	12,789	164	16,506	12,989
	(41,393)	(425)	(54,033)	(42,162)
Manufacturing average	22,375	146	25,852	25,960
	(141,009)	(641)	(140,823)	(169,534)

Source: Authors' calculations using Vietnam Enterprise Surveys 2001–2008.

Note: Means with standard deviations presented in parentheses. Descriptive statistics are presented in levels but included in logs in the productivity analysis.

References

Aitkin, Brian J., and Ann E. Harrison. 1999. "Do Domestic Firms Benefit from Direct Foreign Investment? Evidence from Venezuela." *American Economic Review* 89(3): 605–618.

Audretsch, David B. 1991. "New-firm Survival and the Technological Regime." *Review of Economics and Statistics* 73(3): 441–50.

Aw, Bee Yan, Xiaomin Chen, and Mark J. Roberts. 2001. "Firm-level Evidence on Productivity Differentials and Turnover in Taiwanese Manufacturing." *Journal of Development Economics* 66: 51–86.

Bartelsman, Eric J., and Mark Doms. 2000. "Understanding Productivity: Lessons from Longitudinal Micro Data." *Journal of Economic Literature* 38: 569–594.

Bernard, Andrew B., J. Bradford Jenson, and Peter K. Schott. 2006. "Survival of the Best Fit: Exposure to Low-Wage Countries and the (Uneven) Growth of U.S. Manufacturing Plants." *Journal of International Economics* 68: 219–237.

Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott. 2009. "Products and Productivity." *Scandinavian Journal of Economics* 111(4): 681–709.

———. 2010. "Multi-product Firms and Product Switching." *American Economic Review* 100: 70–97.

Caves, Richard E. 1998. "Industrial Organization and New Findings on the Turnover and Mobility of Firms." *Journal of Economic Literature* 36: 1947–1982.

CIEM (Central Institute of Economic Management). 2003. *Developing the Socialist-Oriented Market Economy in Vietnam*, edited by Dinh Van An. Ha Noi: Statistical Publisher.

———. 2006. *Vietnam's Economy in 2005 (A Reference Book)*. Ha Noi: Publishing House of Political Theory.

Disney, Richard, Jonathan Haskel, and Ylva Heden. 2003. "Restructuring and Productivity Growth in UK Manufacturing." *Economic Journal* 113: 666–694.

Dunne, Timothy, Mark J. Roberts, and Larry Samuelson. 1988. "Patterns of Firm Entry and Exit in the U.S. Manufacturing Industries." *RAND Journal of Economics* 19(4): 495–515.

———. 1989. "The Growth and Failure of U.S. Manufacturing Plants." *Quarterly Journal of Economics* 104(4): 671–698.

Eckel, Carsten, and J. Peter Neary. 2006. "Multi-product Firms and Flexible Manufacturing in the Global Economy." Discussion Paper 5941. London: CEPR.

Eslava, Marcela, John Haltiwanger, Adriana Kugler, and Maurice Kugler. 2004. "The Effects of Structural Reform on Productivity and Profitability Enhancing Reallocation: Evidence from Columbia." *Journal of Development Economics* 75: 333–371.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review* 98(1): 394–425.

Goldberg, Pinelopi K., Amit K. Khandelwal, Nina Pavcnik, and Petia Topalova. 2010. "Multi-product Firms and Product Turnover in the Developing World: Evidence from India." *Review of Economics and Statistics* 92(4): 1042–1049.

Hopenhayn, Hugo, A. 1992. "Entry, Exit and Firm Dynamics in Long Run Equilibrium." *Econometrica* 60(5): 1127–1150.

Jefferson, Gary H., and Thomas G. Rawski. 1994. "Enterprise Reform in Chinese Industry." *Journal of Economic Perspectives* 8(2): 47–70.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71(6): 1695–1725.

Newman, Carol, John Rand, and Finn Tarp. 2011. "Industry Switching in Developing Countries." Working Paper 2011/49. Helsinki: UNU-WIDER.

Olley, G., and A. Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64(6): 1263–1297.

Pavcnik, Nina. 2002. "Trade Liberalization, Exit and Productivity Improvements: Evidence from Chilean Plants." *Review of Economic Studies* 69: 245–276.

Roberts, Mark J., and James R. Tybout. 1996. *Industrial Evolution in Developing Countries: Micro Patterns of Turnover, Productivity and Market Structure*. Oxford: Oxford University Press.

Siegfried, John J., and Laurie Beth Evans. 1994. "Empirical Studies of Entry and Exit: A Survey of the Evidence." *Review of Industrial Organization* 9(2): 121–155.

Stiroh, K., and P. Strahan. 2003. "Competitive Dynamics of Deregulation: Evidence from US Banking." *Journal of Money, Credit and Banking* 35(5): 801–828.

Sutton, John. 2005. *Competing in Capabilities*, Clarendon Lectures. Oxford: Oxford University Press.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature* 49(2): 326–365.

Tybout, James R. 2000. "Manufacturing Firms in Developing Countries: How Well Do They Do, and Why?" *Journal of Economic Literature* 38: 11–44.

World Bank. 2011. *World Development Indicators*. Washington DC: World Bank.

THE WORLD BANK
1818 H Street, NW
Washington, DC 20433, USA
World Wide Web: http://www.worldbank.org/
E-mail: wber@worldbank.org

OXFORD OPEN

ISBN 978-0-19-967961-4
9 780199 679614