INSIDE

Knight on Inequality in China
Devarajan, Khemani, and Walton on Can Civil Society Overcome Government Failure in Africa?
McKenzie and Woodruff on Business and Entrepreneurship Evaluations
Das Gupta on Population, Poverty, and Climate Change
Canuto, Pinto, and Prasad on Orderly Sovereign Debt Restructuring
The World Bank Research Observer is intended for anyone who has a professional interest in development. Observer articles are written to be accessible to nonspecialist readers; contributors examine key issues in development economics, survey the literature and the latest World Bank research, and debate issues of development policy. Articles are reviewed by an editorial board drawn from across the Bank and the international community of economists. Inconsistency with Bank policy is not grounds for rejection.

The journal welcomes editorial comments and responses, which will be considered for publication to the extent that space permits. On occasion the Observer considers unsolicited contributions. Any reader interested in preparing such an article is invited to submit a proposal of not more than two pages to the Editor. Please direct all editorial correspondence to the Editor, The World Bank Research Observer, 1818 H Street, NW, Washington, DC 20433, USA.

The views and interpretations expressed in this journal are those of the authors and do not necessarily represent the views and policies of the World Bank or of its Executive Directors or the countries they represent. The World Bank does not guarantee the accuracy of data included in this publication and accepts no responsibility whatsoever for any consequences of their use. When maps are used, the boundaries, denominations, and other information do not imply on the part of the World Bank Group any judgment on the legal status of any territory or the endorsement or acceptance of such boundaries.

Subscriptions
A subscription to *The World Bank Research Observer* (ISSN 0257-3032) comprises 2 issues. Prices include postage; for subscribers outside the Americas, issues are sent air freight.

Annual Subscription Rate (Volume 29, 2 issues, 2014)

**Academic libraries**
Print edition and site-wide online access: US$220/£147/€220
Print edition only: US$202/£134/€202
Site-wide online access only: US$177/£117/€177

**Corporate**
Print edition and site-wide online access: US$334/£221/€334
Print edition only: US$306/£203/€306
Site-wide online access only: US$268/£177/€268

Please note: US$ rate applies to US & Canada, Euros applies to Europe, UK£ applies to UK and Rest of World.

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.

There may be other subscription rates available; for a complete listing, please visit www.wbro.oxfordjournals.org/subscriptions.

Full pre-payment in the correct currency is required for all orders. Payment should be in US dollars for orders being delivered to the USA or Canada: Euros for orders being delivered within Europe (excluding the UK); GBP sterling for orders being delivered elsewhere (i.e., not being delivered to USA, Canada, or Europe). All orders should be accompanied by full payment and sent to your nearest Oxford Journals office. Subscriptions are accepted for complete volumes only. Orders are regarded as firm, and payments are not refundable. Our prices include Standard Air as postage outside of the UK. Claims must be notified within four months of despatch/order date (whichever is later). Subscriptions in the EEC 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. Subscribers in Canada, please add GST to the prices quoted. Personal rate subscriptions are only available if payment is made by personal cheque or credit card, delivery is to a private address, and is for personal use only.

**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.

Contact information:
Journals Customer Service Department, Oxford University Press, Great Clarendon Street, Oxford OX2 6DP, UK. E-mail: jnlscust.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, 2011 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, 4-5-10-8F Shibata, Minato-ku, Tokyo, 108-8356, Japan. E-mail: custservjp@oup.com. Tel: +81 3 5444 5858; Fax: +81 3 3454 2929.


**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/jnl/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: ijsadvvertising@oup.com.

**Disclaimer:** Statements of fact and opinion in the articles in *The World Bank Research Observer* are those of the respective authors and contributors and not of the International Bank for Reconstruction and Development/The World Bank or Oxford University Press. Neither Oxford University Press nor the International Bank for Reconstruction and Development/The World Bank make any representation, express or implied, in respect of the accuracy of the material in this journal and cannot accept any legal responsibility or liability for any errors or omissions that may be made. The reader should make her or his own evaluation as to the appropriateness or otherwise of any experimental technique described.

**Paper used:** *The World Bank Research Observer* 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 Research Observer* is indexed and/or abstracted by ABI/INFORM, CAR Abstracts, Current Contents/Social and Behavioral Sciences, Journal of Economic Literature/EconLit, PALS International, RePEc (Research in Economic Papers), Social Services Citation Index, and Wilson Business Abstracts.

**Copyright:** © 2014 The International Bank for Reconstruction and Development/The World Bank. All rights reserved; no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise without prior written permission of the publisher or a license permitting restricted copying issued in the UK by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1P 9HE, or in the USA by the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. Typeset by Techset Composition Limited, Chennai, India. Printed by The Sheridan Press, USA.
Inequality in China: An Overview
   John Knight

Can Civil Society Overcome Government Failure in Africa?
   Shantayanan Devarajan, Stuti Khemani, and Michael Walton

What Are We Learning from Business Training and Entrepreneurship Evaluations around the Developing World?
   David McKenzie and Christopher Woodruff

Population, Poverty, and Climate Change
   Monica Das Gupta

Orderly Sovereign Debt Restructuring: Missing in Action! (And Likely To Remain So)
   Otaviano Canuto, Brian Pinto, and Mona Prasad
Inequality in China: An Overview

John Knight

This paper provides an overview of research on income inequality in China over the period of economic reform. It presents the results of two main sources of evidence on income inequality and, assisted by various decompositions, explains the reasons income inequality has increased rapidly and the Gini coefficient is now almost 0.5. This paper evaluates the degree of income inequality from the perspectives of people’s subjective well-being and government concerns. It poses the following question: has income inequality peaked? It also discusses the policy implications of the analysis. The concluding comments of this paper propose a research agenda and suggest possible lessons from China’s experience that may be useful for other developing countries. JEL codes: D31, D63, O15.

When China embarked on economic reform, it had too much income equality. The egalitarian arrangements in the communes and factories stifled incentives and produced enormous inefficiency. The new Chinese leadership recognized that greater income inequality was necessary to provide the incentives essential to an economy that was in the process of making the transition from a central planned economy to a market-driven, private-sector-based economy. Inequality increased rapidly over the reform period. The Gini coefficient of household income per capita was 0.49 in 2007 (Li et al. 2013), and China was found to have the joint highest inequality in Asia (Asian Development Bank 2007: figure 1). Income inequality had become a matter of concern to the Chinese leadership.

It is notoriously difficult to make reliable intercountry comparisons of income inequality or its change. Nevertheless, table 1 reports the Gini coefficient in the 15 largest developing countries (for which data are available) in the late 1980s and the late 2000s. The table suggests that China is outstripped in its recent inequality only by Brazil and South Africa and in the rise of its inequality only by the Russian Federation.
This paper is a timely and reflective overview of recent economic changes in inequality. It is not designed to provide a comprehensive and thorough empirical survey; rather, it concentrates on the aspects of China’s inequality and its rise that are likely to be of most interest beyond China. This paper focuses on the period of economic reform beginning in 1978 and is largely concerned with inequality of income and factors that generate this inequality.

We address a series of important and serious questions. How well can China’s income inequality be measured? Can discrepancies in the evidence from alternative sources be explained? How and why has wealth inequality increased? What are the dimensions and components of increasing income inequality? How do the different components help to explain the remarkable rise in income inequality? What is the relationship of inequality to poverty? Why and how is income inequality of concern to people and to the government? Can China’s past and likely future income inequality be interpreted in terms of the inverted-U of the Kuznets curve? What are the implications for Chinese policy? Are there lessons for research and for other developing countries?

### Table 1. The Gini Coefficient of Income Inequality for Selected Large Developing Countries, Circa 1988 and Circa 2007

<table>
<thead>
<tr>
<th>Country</th>
<th>Circa 1988</th>
<th>Circa 2007</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>China (World Bank)</td>
<td>0.30</td>
<td>0.43</td>
<td>0.13</td>
</tr>
<tr>
<td>(CHIP surveys)</td>
<td>0.38</td>
<td>0.49</td>
<td>0.11</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>0.29</td>
<td>0.32</td>
<td>0.03</td>
</tr>
<tr>
<td>Brazil</td>
<td>0.61</td>
<td>0.56</td>
<td>-0.05</td>
</tr>
<tr>
<td>Egypt</td>
<td>0.32</td>
<td>0.31</td>
<td>-0.01</td>
</tr>
<tr>
<td>India</td>
<td>0.32</td>
<td>0.33</td>
<td>0.01</td>
</tr>
<tr>
<td>Indonesia</td>
<td>0.29</td>
<td>0.34</td>
<td>0.05</td>
</tr>
<tr>
<td>Iran</td>
<td>0.44</td>
<td>0.38</td>
<td>-0.06</td>
</tr>
<tr>
<td>Nigeria</td>
<td>0.39</td>
<td>0.49</td>
<td>0.10</td>
</tr>
<tr>
<td>Pakistan</td>
<td>0.33</td>
<td>0.30</td>
<td>-0.03</td>
</tr>
<tr>
<td>Philippines</td>
<td>0.41</td>
<td>0.43</td>
<td>0.02</td>
</tr>
<tr>
<td>Russian Federation</td>
<td>0.24</td>
<td>0.43</td>
<td>0.19</td>
</tr>
<tr>
<td>South Africa</td>
<td>0.59</td>
<td>0.63</td>
<td>0.04</td>
</tr>
<tr>
<td>Thailand</td>
<td>0.44</td>
<td>0.41</td>
<td>-0.03</td>
</tr>
<tr>
<td>Turkey</td>
<td>0.44</td>
<td>0.39</td>
<td>-0.05</td>
</tr>
<tr>
<td>Vietnam</td>
<td>0.36</td>
<td>0.36</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Notes: All earlier figures fall within the 1986–1990 period except Vietnam (1993) and South Africa (1993), and all later figures fall within the 2005–2010 period. The (alternative) CHIP estimates for China will be explained below.

Source: worldbank.org/indicator/SI.POV.GINI; Griffin and Zhao (1993), Li et al. (2013).

The World Bank Research Observer, vol. 29, no. 1 (February 2014)
Measuring Inequality

Some knowledge and understanding of the data sources and their limitations is necessary. There are two main sources of information on income inequality over time: the annual national household income and expenditure surveys of the National Bureau of Statistics (NBS) and the periodic national household surveys of the China Household Income Project (CHIP). The NBS surveys contain many observations but a limited number of questions. They cannot be used as a panel (namely, longitudinal study over two or more points in time), and they are generally not available to researchers at disaggregated household and individual levels. Therefore, measures of income inequality such as the Gini coefficient, when derived from official statistics, must be based on province-level or percentile data. The CHIP surveys relate to the years 1988, 1995, 2002, and 2007, they use a subsample of the NBS surveys, and they ask many more questions. There is an edited volume on each of the CHIP surveys (Griffin and Zhao 1993; Riskin et al. 2001; Gustafsson et al. 2007a; Li et al. 2013). The two sources use different definitions of income; the CHIP definition is more comprehensive.

Because of the sharp administrative and economic divide between urban and rural China and the need for different survey questionnaires, measures of inequality are generally reported for urban and rural areas separately as well as with a weighted national measure. The NBS surveys are based on urban or rural residence registration (hukou), so they exclude most rural-urban migrants (who normally retain rural hukou) from the urban sample. By 2002, the number of rural-urban migrants exceeded 100 million. The 2002 and 2007 CHIP surveys added a separate sample of rural hukou households in urban China.

China’s poverty and inequality decreased dramatically in 1978–1985 during the years of rural reform, when farming was decollectivized, household production was restored, and farm incomes responded. It is possible to obtain a fairly consistent set of Gini coefficients from the 1988–2007 CHIP surveys. In 1988, the urban Gini (0.24) was very low by international standards, the rural Gini (0.33) reflected regional income disparities, and the national Gini (0.38) was higher than both the urban and the rural Gini coefficients because of the high ratio of urban to rural household income per capita. There appeared to be a lull in this increase because the national Gini was 0.45 in both 1995 and 2002. However, in 2007, the urban Gini was 0.34, the rural Gini was 0.36, and the national Gini was no less than 0.50. Adjusted for regional price differences, the Gini was 0.43 in 2007, having risen from 0.40 in 2002 (Li et al. 2013).

Ravallion and Chen (2007), who had partial access to the NBS microdata, found growing income inequality: all three Gini coefficients increased by 3 percentage points over the six years between 1995 and 2001. These authors’ estimate of the national Gini in 2001 was 0.45. NBS statistics for urban China show that income
inequality continued to rise after 2001. The share of the lowest three quintiles fell monotonically over the period of 2000–2008, whereas that of the highest quintile increased sharply. Moreover, the national Gini coefficient based on grouped NBS data was estimated to rise by 5 percentage points between 2000 and 2008 (Lin et al. 2010).

Although there is no indication that income inequality in urban areas stopped increasing in the NBS data, the CHIP data suggest that it was no higher in 2007 than it had been in 1995. The explanation for this discrepancy is likely to be found in the definition of income. In contrast to the NBS’s definition, CHIP’s definition includes various regressive subsidies received by urban hukou residents, particularly housing subsidies. The phasing out of subsidies over this period may have reduced urban income inequality.

The various estimates of income discussed so far are for disposable income, which includes various private and public transfers as well as factor incomes (derived from productive activities). In fact, taxes and subsidies have done nothing to remedy factor income inequality, although the degree of fiscal regressivity (in which the effect on income is disproportionally greater on poorer than richer people) has fallen as reforms have progressed (Khan and Riskin 2007). In 2007, the urban Gini for income after the deduction of direct taxes was only 1 percentage point lower than its pretax counterpart (Xu and Yue 2013).

A rich entrepreneurial class emerged remarkably rapidly in China. There were above-normal profits to be earned, and the combination of a semimarketized economy, weak legal system, and ill-defined or insecure property rights provided room for corruption, cronyism, and rent seeking. Because not all income derived from such activities is detectable in the NBS or CHIP surveys, incomes at the top of the income distribution are likely to be understated. An ingenious attempt to measure this effect claimed to find much “grey income” in the highest income group (Wang and Woo 2010).²

During the period of central planning, there was almost no personal wealth in China. Economic reform brought not only rapid accumulation but also considerable inequality of wealth. China thus provides an excellent case study of the various processes that generate wealth inequality. The Gini coefficient of wealth in 2002 was 0.55 (rural 0.40, urban 0.48), which was considerably higher than the coefficient of income per capita (Zhao and Ding 2007). The main cause of the higher Gini coefficient in both rural and urban areas was differences in the quality and value of housing, which, in the latter case, represented two-thirds of the inequality of net wealth. Urban dwellers who acquired ownership of the houses that they had occupied (while paying nominal rents) made huge capital gains; the housing subsidy was merely capitalized. Rationed access to cheap loans from state-owned banks provided opportunities for capital accumulation. More generally, the acquisition or appropriation of state assets at below-market prices was a powerful disequalizing force.
The divergence of wealth was assisted by the fact that the household saving rate increases sharply with income.

Decomposing Inequality

To better understand income inequality in China, we need to decompose the complexity of the data. Thus, we consider the various dimensions and components of income inequality, starting with the urban sector. Under central planning, the work unit (danwei) served as a mini welfare state, providing lifetime employment, housing, pensions, and medical care to its members. Workers were allocated bureaucratically; wages were determined administratively and were highly egalitarian. As an urban labor market gradually emerged, the wage structure widened and wage inequality increased. For instance, the Gini coefficient was 0.21 in 1988, 0.33 was in 1995, but was still 0.33 in 2007 (Deng and Gustafsson 2013).

The increase was partly due to increasing rewards for productive characteristics and incentives for efficiency. For instance, the wage premium of a college degree over primary schooling was 9 percent, 39 percent, and 88 percent in 1988, 1995, and 2002, respectively. However, the increase was partly due to new or growing forms of discrimination and segmentation (Knight and Song 2007). For instance, these authors found that wages were increasingly sensitive to enterprise profitability due to informal profit sharing. Knight and Yueh (2008) identified an important and continuing role for social connections despite the growing strength of market forces. Appleton et al. (2004) showed that urban workers who had been retrenched as a result of the reform, privatization, and closure of state-owned enterprises had to enter a difficult new labor market and, if reemployed, were at a considerable wage disadvantage in comparison with nonretrenched urban workers. The same was true of rural-urban migrants, who generally retained their rural hukou (residence registration) and were treated as second-class citizens in the cities. Such distinctions were grounds for perceptions of distributive injustice.

The presence of rural-urban migrants complicates the accuracy of estimates of urban inequality. Many of these people are temporary migrants who retain close links with their rural households and expect to return to them. Measures of rural household income include migrant remittances. Only the recent CHIP surveys include samples of urban resident households with rural hukou. The 2002 CHIP survey permitted an estimate of the urban Gini coefficient both with and without migrants. The inclusion of migrants raised the Gini by 2 percentage points (Khan and Riskin 2007), but this result may be an understatement if migrants living in households have higher incomes than independent sojourners.

Nonfarm employment is important for rural household income and its distribution. Both average and, especially, marginal income are higher in local nonfarm
and migrant activities than in farming. The share of wages in rural income rose sharply as rural industry burgeoned and migration accelerated. Processes of cumulative causation were initially at work as some villages industrialized and some became migration villages. Wage income contributed 21 percent of rural income inequality in 1988, 40 percent in 1995, and 41 percent in 2007. The slowdown was due to the reduction in rural spatial income inequality as wage employment opportunities spread more widely across provinces and counties. In principle, migration can either increase or decrease inequality depending on whether poor households, which have the greatest incentive to send migrants, have the ability to do so. An analysis of the effect of migrant members on the income of rural households using the 2007 CHIP survey showed that it reduced rural poverty and, by implication, inequality (Luo and Yue 2010). Benjamin et al. (2005) analyzed rural income inequality using a Ministry of Agriculture annual survey of 100 villages. Between 1987 and 1999, the (spatially price-deflated) Gini coefficient rose from 0.29 to 0.35. Most of this increase was at the local level. Whereas household access to local nonfarm employment increased inequality in this period, temporary migrant employment decreased it.

Under central planning, China was characterized by a severe rural-urban divide. This divide was not reduced by the reform and marketization of the economy; the ratio of urban to rural household income per capita was greater than ever in 2007, at 4.10 according to the CHIP survey. However, it decreased to 2.91 after adjustment for spatial differences in prices. The corresponding CHIP ratios in 2002 were 3.35 and 2.28 (Li et al. 2013: table 2.8). Including various disguised subsidies (for health care, education, and pension contributions), the 2002 ratios were 4.35 and 3.10, respectively (Li and Luo 2010: 119). Rural-urban differences in the cost of living were offset by subsidies to urban people. The explanation for the high ratio is the underlying political economy that favors urban dwellers and the control of migration (Knight and Song 1999). The contribution to overall inequality made by the mean difference in rural and urban incomes rose from 37 percent in 1988 to 54 percent in 2007. Even adjusting for spatial price differences, which reduces the 2007 figure to 41 percent (Li et al. 2013), this is far higher than that in most other developing countries. Much of China’s income inequality would vanish if mean income per capita in rural and urban China were equal.

Regional Inequality

It is inevitable that a country as large as China will have large spatial or geographical differences in income levels. The more interesting question is whether there is regional divergence or convergence over time—that is, whether the processes of cumulative causation that produce “polarization” are outweighed by “spread effects.” The former are likely to be significant in the initial stages of economic development
but eventually give way to the latter as competitive advantages are eroded by rising costs. There is a good deal of research on this question in relation to China. The evidence tends to favor absolute divergence, but in line with economic theory, conditional convergence exists. This pattern was found by Lau (2010) in an examination of the GDP per capita among provinces over the 1978–2005 period.

Unfortunately, the use of province-level GDP per capita as the dependent variable is liable to produce biased results (Tsui 2007; Li and Gibson 2012). Whereas GDP data relate to production in the province, population data generally refer to the population registered in the province and exclude rural-urban migrants from other provinces who retain their rural hukou registration. This approach overstates the GDP per capita in the richer provinces that attract migrants. Because migration has grown rapidly, the GDP per capita growth rates of these provinces are exaggerated. Thus, evidence of absolute divergence might be an artifact.

Income or consumption based on representative household surveys is therefore probably a more reliable guide to changes in regional income inequality. Kanbur et al. have studied regional income inequality in China using consumption per capita figures derived from the NBS annual household surveys (for instance, Kanbur and Zhang 2005; Fan et al. 2008, 2011). These authors generate a series for China’s income inequality aggregated to the province level (and thus exclude income inequality among households within a province). Over the period from 1980 to 2007, the Gini coefficient rose from 0.27 to 0.34 (Fan et al. 2011). Econometric analysis shows that fiscal decentralization and trade liberalization contributed to the rise in inequality (Kanbur and Zhang 2005). Fiscal decentralization enabled the richer coastal provinces to increase their revenues and thus to promote economic development. Trade liberalization enabled the coastal provinces to grow more rapidly through both their geographic advantage and preferential treatment from the central government (with respect, for instance, to infrastructure and foreign direct investment). Fan et al. (2011:50) found that inequality attributable to income differences between the coastal and inland regions increased from 3 percent to 10 percent of the total province-level inequality between 1980 and 2007.

A further reason for the rise in income inequality among provinces in recent years involves the fiscal relationship between central and provincial governments. After the fiscal recentralization of 1994, the central government had greater power to redistribute revenue to the poorer provinces. Rule-based transfers tend to be equalizing, but the two-thirds of transfers that are specific and subject to negotiation are found to be disequalizing because they require matching funds or produce rent seeking (Huang and Kang 2012). In recent years, fiscal transfers from the center to the provinces have done nothing to correct the income divergence among provinces.

Inequality among provinces makes a larger contribution to inequality among households in rural China than in urban China. Using the CHIP surveys, Gustafsson et al. (2007b) found that the proportion of household inequality in urban China due
to between-province inequality fell from 29 percent in 1988 to 19 percent in 2002. The main gain came from within eastern China, where this more developed economy was becoming more spatially integrated. The contribution of between-province inequality to rural income inequality rose from 22 percent in 1988 to 39 percent in 1995 and remained at 39 percent in 2002 (Gustafsson et al. 2007b). It appears that the initial polarization effects were offset by the spread effects that were created by the growing scarcity of local resources.

Evaluating Inequality

It is well known that absolute income poverty in China has fallen dramatically. For instance, Ravallion and Chen (2007) report that the proportion of households who were under the official absolute poverty line decreased from 53 percent in 1980 to 18 percent in 1988 and to 8 percent in 2001. In each year, the great majority of the poor were rural. However, if poverty is defined in terms of relative income, it did not decrease. For instance, the proportion of households with no more than half of the national median household income per capita edged up from 13.2 percent to 13.3 percent in the five years from 2002 to 2007 (Li et al. 2013). Ultimately, the choice of how to measure poverty requires a value judgment. Concern for income inequality implies the introduction, wholly or partly, of a relative concept. Sen (1983 and elsewhere) has argued that concern for absolute poverty in terms of people’s “capabilities” (to be and to do things of intrinsic worth) can imply concern for more income equality—a reduction in relative poverty. The dramatic fall in absolute poverty in China over the reform period, reflecting the overall rise in personal income, strengthens the case for switching to the use of a poverty line that is expressed in relative terms.

The evaluation of income inequality requires a normative judgment. The economic literature on inequality commonly proceeds from the judgment that income inequality at the national level is the appropriate criterion and that the degree of inequality measured in this way is too high and should be reduced. This section delves more deeply into the basis for such a judgment. How much concern is there about income inequality in China? We focus first on the people and then on the government.

Research on subjective well-being in China shows why people are concerned about income inequality: “relative deprivation” is a common phenomenon (for instance, Knight and Gunatilaka 2011, using the 2002 CHIP survey). Regression analyses of happiness (or life satisfaction, or subjective well-being—we use the terms interchangeably) in China produce well-fitting equations with understandable and significant coefficients. Two consistent findings are the importance of relative income and the importance of the chosen reference group.
In rural China, where the happiness score ranges from 0 to 4 with a mean of 2.7, those who report being “much below” the mean income of the village (five answers are possible) have a happiness score that is 1.06 below those who report being “much above” the mean (Knight and Gunatilaka 2010). Respondents were asked with whom they compared themselves; in rural China, two-thirds of respondents claimed to compare themselves with others living within the village. This finding indicates that the reference groups are narrow. In rural China, the “relevant others” are households in the same village, and in urban China, the “relevant others” are households in the same city. In the cities, where a doubling of income raises the happiness score by 0.10 units, the happiness of respondents whose households fall into the lowest quarter of city income per capita is 0.81 points below those in the highest quarter (Knight and Gunatilaka 2010). Thus, income inequality matters, but it is inequality at the local and not the national or regional level that matters to people.

The finding that subjective well-being in China is sensitive to relative income is in line with the evidence for many countries (surveyed by Clark et al. 2008, and by Graham and Felton 2006). The coefficient on group income is generally negative, but there are cases in which it is positive. For instance, Senik (2004) found a positive coefficient for Russia (on income in the region), and Kingdon and Knight (2007) found a positive coefficient for South Africa (on income in the close neighborhood). The (usual) negative coefficient is normally interpreted to indicate feelings of relative deprivation, and the (unusual) positive coefficient is normally interpreted to indicate local opportunities for improvement (Russia) or fellow feeling (South Africa). The reference group may be crucial. In rural China, the negative effect of being low in the village’s income distribution coexists with a positive effect of income inequality in the county in which it is located, as measured by the Gini coefficient (Knight et al. 2009). The former may represent relative deprivation, and the latter may represent perceived room for self-advancement.

Over the period from 1990 to 2010, China’s income grew substantially, yet life satisfaction was no higher in 2010 than it had been in 1990 (Easterlin et al. 2012). This finding is based on an examination of the six available time series of life satisfaction in China over that period (including, for instance, the World Values Survey and the Gallup World Poll). The explanation for the stagnation in life satisfaction can be found in the increase in the reference group’s income, which offsets the effect of the increase in one’s own income, as well as the powerful socioeconomic changes that accompanied China’s rapid growth (Knight and Gunatilaka 2011). These changes included higher unemployment and redundancy, greater economic uncertainty and insecurity, and changing reference groups.

Easterlin et al. (2012) found that inequality of income and inequality of the life satisfaction score rose over those two decades. This finding is consistent with the positive association between income and happiness found in the cross-section.
However, it is plausible that subjective well-being becomes more sensitive to income inequality and that average subjective well-being decreases as income inequality at the national level rises and social cohesion is weakened. These hypotheses have yet to be tested for China, but if they were correct, they would strengthen the case for the central government to use policies to address and reduce income inequality.

The failure of subjective well-being to rise across two decades should be a matter of concern for the Chinese government. China does not rank high in various recent international rankings reported in the World Happiness Report, being in the 27th percentile for quality of life, the 28th percentile for life satisfaction, and the 30th percentile for happiness (Helliwell, Layard, and Sachs 2012: figures 2.3, 2.5, 2.9). Inequality is very likely to explain, in part, this relatively lower subjective well-being.

The Chinese government’s overriding objective of rapid economic growth has two implications for its policy on income inequality. On the one hand, there is some evidence that spatial income inequality has been bad for economic growth. For instance, Ravallion (1998) showed for rural China that asset inequality in the locality retarded the growth of individual household consumption, and Ravallion and Chen (2007) found that provinces with higher income inequality experienced slower economic growth. On the other hand, economic reforms, marketization, and institutional arrangements that promoted economic growth contributed to increases in inequality in various dimensions.

The government’s unwillingness or inability to prevent income inequality from increasing signals a clear risk of rising social discontent. In common with leaders in many other countries, China’s leadership is concerned with its own political survival. Specifically, in recent years, China’s leaders have expressed their concern about the possibility of “social instability.” Social instability, in turn, can impede China’s continued rapid growth (Knight and Ding 2012: 295–306). One potential source of social instability is income inequality.

We have seen that the increase in China’s income inequality takes three main forms: among households, across regions, and between rural and urban areas. However, comparisons in these dimensions of inequality at the national level may not be important. Because of the narrowness of people’s reference groups, it may be more important for a government concerned about social instability to remedy the causes of income inequality at the local level. However, extra-local orbits of comparison are widening owing to the increasing use of the Internet and the growth of “the greatest migration in human history.” The analysis of rural-urban migrants living in households showed that this group had the lowest mean happiness—owing to the transfer of their reference group to the city, with its higher incomes, and the unequal treatment of rural hukou households in matters of employment, residence, education, and other services (Knight and Gunatilaka 2011). A qualification is in order: unhappiness does not necessarily translate into social discontent.
This relationship might depend on the extent to which people perceive their unhappiness to be manmade and capable of being remedied by the government.

Government strategies for much of the reform period can be summed up in the words of a high official, Du Runsheng (1989: 192): “Prosperity to few, then to many, then to all.” Wherever there has been a tradeoff, efficiency considerations have taken precedence over equity considerations. In creating a “developmental state,” the government overwhelming prioritized economic growth. However, it appears that the leadership became more sensitive to rising income inequality in the mid-2000s, when policies to promote a more “harmonious society” were introduced. We examine the policies that have been introduced and those that seem promising for the future.

Policy Implications

Policy can be addressed at two levels: the redistribution of primary income through income transfers and alterations in the primary income generation processes. We focus on both of these in turn. “Harmonious society” policies have concentrated on the former and on people at the bottom of the national income distribution. In 2007, no less than 97 percent of poor households (defined as those with real income per capita of less than 1.25 dollars a day [purchasing power parity]) were rural dwellers (Li et al. 2013). There was a series of pro-rural policies. One of these policies concerned agricultural taxes and fees, which had been oppressive and regressive, averaging 5.3 percent of rural household income overall and 13.9 percent for the lowest income decile in the 1995 CHIP survey. Agricultural taxes and fees were abolished in 2006, so the average tax rate on rural household incomes was only 0.3 percent in the 2007 CHIP survey (Li 2012). Several other policies were introduced to benefit farmers during the first decade of the new century. These included compensation policies to return farmland to forest, a farm support program involving agricultural subsidies, and rural infrastructure development. From 2004 to 2011, the growth rate of central government funds to support agriculture grew by nearly 30 percent per annum (Li 2012).

In 2002, the poorest quintile of rural households spent a quarter of their income on education (Knight et al. 2009:317). An important redistributive policy with short- and long-term consequences was the abolition of all school fees in compulsory (nine-year) education. This policy was introduced in poor rural areas in 2005 and was extended to all rural areas in 2007.

The minimum income guarantee (dibao) system became important in rural China only after 2005, reaching 52 million people in 2010. The dibao system had been introduced earlier in the cities; by 2010, it covered 23 million urban people (Li 2012). Dibao helped the unemployed, those in ill health, and the elderly. However,
because of poor coverage and low benefit levels, it had a limited effect on urban poverty (Ravallion 2012) and even less of an effect on income inequality. Another form of intervention in urban China that was intended to be redistributive was the introduction and extension of minimum wages in many cities. Real minimum wages have risen rapidly in recent years, reflecting the central government’s directions and incentives.

Direct taxation is low in China. Because it is based on the individual and not the household and is open to evasion by high-income groups, it has little effect on urban income inequality. In 2008, personal income tax represented less than 0.01 percent of the household income of those in the lowest income decile, 0.12 percent in the sixth decile, and 2.1 percent in the highest decile (Li 2012). Personal income tax accounted for less than 7 percent of government revenue in 2010; indirect taxation was much more important. There is room to make direct taxation a more important source of revenue and to make it more progressive.

There are institutional reasons why China’s social security provision remains highly segmented. Under central planning, the social security system was largely confined to urban residents, who enjoyed an “iron rice bowl” provided by employing enterprises. With enterprise reform, which began in earnest in the late 1990s, the enterprise provision of social security disintegrated. Unemployment insurance, health care insurance, and pension schemes were belatedly and incompletely taken over by broader groupings that were normally based on locality or an ownership sector. Urban informal sector workers and rural-urban migrants were poorly covered. Social security provision in rural China remains limited in both coverage and quality, although rural health care insurance expanded rapidly in just a few years to achieve a participation rate of 95 percent in 2010 (Li 2012). Although this type of inequality is not reflected in the measure of income inequality, its inclusion could be expected to exacerbate rather than diminish the extent of inequality in economic welfare. Movement toward a comprehensive system of social security provision within a common and progressive framework would reduce inequalities in Chinese society.

Income inequality among provinces can be addressed by increasing the importance of (the equalizing) rule-based general revenue transfers from the central government and by reducing the importance of (the disequalizing) specific transfers. However, there should be specific transfers solely or preferentially to the poorer provinces for development-promoting expenditures, such as infrastructure investment, education, and health care. The stimulus package introduced in response to the world financial crisis of 2007–2008 marks some movement in that direction (Fan et al. 2011).

We now focus on the policies that may be needed to equalize the distribution of factor income. The institutional arrangements that divide China’s society into urban (households with urban residence registration, hukou), rural-urban migrants
(most of whom retain rural hukou), and rural (with rural hukou) create unequal access to various income-earning opportunities, including jobs and human capital acquisition. Some of this income inequality could be reduced by permitting rural-urban migrants the freedom to settle and to compete on equal terms with urban residents in the labor market. Some of the inequality, however, is deep rooted and long lasting and requires attention to education policies. There is a great disparity in the quantity and quality of education that urban and rural children receive (for instance, Knight and Song 1995: ch.4). Moreover, there is inequality in access to education within rural China based mainly on the income and educational attainment of households and on the locality (Knight et al. 2009).

Given the importance of education for income generation, this difference in access to education can give rise to a poverty trap (Knight et al. 2010). The transmission of education from one generation of a household to another is a powerful phenomenon in reform-era China and has become stronger in recent years (Knight et al. 2013). This phenomenon tends to carry forward educational inequality, and thus income inequality, from one generation to the next. Although the abolition in 2007 of school fees for compulsory schooling in rural China helped to equalize educational opportunities, policy measures are needed to address unequal access to high-quality education at all levels, including high school and higher education.

Little policy attention has been paid to inequality at the top of the income distribution. China’s system of governance is open to rent seeking and corruption and to profit opportunities for those with power or influence. According to the World Bank’s Worldwide Governance Indicators, China was ranked 148th on “control of corruption” and 220th on “voice and accountability” out of 235 countries in 2009 (World Bank 2011). Policies to reduce the inequalities that arise in these ways would require reforms in China’s governance, such as the creation of a powerful anticorruption agency, the strengthening of the rule of law, greater press or media freedom, and arrangements that accord more “voice” to citizens.

Primary income distribution also depends on the relative demand for and supply of the factors of production, including labor and human capital. Although China’s rapid economic growth and marketization have contributed to the rise in income inequality, they may eventually generate equalizing market forces. Between 1995 and 2007, the labor force, affected by the one-child policies that were introduced at the beginning of economic reform, rose by only 1.3 percent per annum, and urban employment rose from 28 percent to 37 percent of the labor force. Much of this increase was due to increased rural-urban migration: the employment of rural migrants in urban areas rose from 30 million to 132 million over that period (Knight et al. 2011). When unskilled labor eventually becomes scarce, unskilled wages can be predicted to rise relative to other incomes. The priority accorded to
promoting rapid economic growth may thus eventually prove to be the policy responsible for the most dramatic equalization of income.

China’s proportion of labor with higher education has been small by international standards, and access to higher education has been rationed. This scarcity raised the premium on higher education as market forces began to operate in the labor market. However, in the late 1990s, a dramatic change took place in China’s higher education policy. In 1998, higher education enrollment was 3.4 million; in 2008, it was 20.1 million, nearly six times its level a decade earlier. Short-term labor market consequences take the form of a rise in unemployment among graduates and the gradual acceptance of jobs previously entered by nongraduates or of “graduate” jobs at lower pay. The long-term graduate wage premium is also affected. The demand for university graduates is likely to grow rapidly as China responds to the rising price of unskilled labor and with industrial upgrading to technologically advanced processes and products. However, a policy of rapidly expanding the supply of graduates in relation to their demand will likely narrow the wage structure.

Whither Inequality?

Inequality decreased during the brief period of dramatic rural reform but rose rapidly as urban reform progressed. The initial changes in national inequality were related much more closely to economic reforms than to the level of income, but the rise was consistent with the upward-sloping part of the hypothesized Kuznets curve relating inequality to income level (Kuznets 1955). Can China be predicted to follow the downward-sloping part of the hypothesized Kuznets curve as well? That is, will inequality decrease as income increases in the future?

The answer to this question will depend on the balance of countervailing forces. On the one hand, various processes that have increased China’s income inequality during the reform period will continue to operate. On the other hand, there are three main equalizing forces that may weaken or entirely offset these processes. It is predictable that the labor market will tighten as China enters the second stage of the Lewis model and the fruits of economic development are extended. It appears very likely from projections of the labor force and of urban employment that this transition will occur in the 2011–2020 decade (Knight et al. 2011). The growing scarcity of labor and other resources can be predicted to transfer production from the coastal provinces to the poorer interior provinces. These processes may have already begun; the NBS household surveys show that since 2009, rural household income per capita has grown faster than its urban counterpart and that overall provincial income inequality appears to have leveled off and even slightly declined since 2005 (Fan et al. 2011). There are signs that Chinese society is becoming more sophisticated and better informed and that people’s aspirations are rising. In this
situation, the government may selectively introduce stronger policies to diminish various dimensions of inequality as a protection against social disorder or instability.

Concluding Comments

It was inevitable that income inequality would increase significantly as China moved from a centrally planned economy, in which egalitarianism was a cornerstone, to a market-based economy. Material incentives were needed to induce greater effort, saving, investment in physical and human capital, and entrepreneurship. Similarly, economic efficiency was likely to be enhanced by disequalizing processes of cumulative causation. Nevertheless, some of the increase in income inequality was difficult to justify in terms of either efficiency or equity. Much of this unjustified inequality stemmed from the institutional framework within which China’s semimarketized economy operated.

The research agenda on income inequality in China (and other developing countries) could productively move in the following direction. A distinction can be made between inequality that is based on rewards for productive characteristics and inequality that is based on market discrimination or segmentation and unequal access to income opportunities. In 2004, Whyte (2010) conducted a sociological survey of Chinese attitudes toward income inequality and concluded that Chinese people were not averse to the degree of inequality that they observed, particularly if it was based on merit, effort, or risk taking. Indeed, income inequality appeared to offer people incentives or other opportunities for improving their economic positions. This interpretation corresponds to the first stage of the “tunnel effect” (see below) hypothesized by Hirschman and Rothschild (1973).

By contrast, inequality based on unfairness or inequity in access to opportunities was generally disliked. Whyte (2010) found that farmers, despite being the poorest group, were the least discontented. Actual income is not necessarily a good guide to perceived distributional injustice because people’s information sets and aspirations not only matter but also vary. It is an important question whether China will enter the second stage of Hirschman’s tunnel effect—that is, whether or when a critical mass of people will begin to see inequality not as a sign of available opportunities but as a sign of unequal opportunities and distributional injustice.

For the first quarter century of economic reform, China’s leaders gave overriding priority to the achievement of rapid economic growth, even at the cost of rising income inequality. When there was a policy tradeoff between equity and efficiency, the efficiency objectives normally won out. One of the few exceptions—to be explained by the government’s concern for maintaining social stability—is the retention of leasehold arrangements in farming and the continued refusal to permit land ownership in rural China. The system of fiscal decentralization and the *nomenklatura*
system of state appointments created incentives at all levels to promote economic growth. Thus, China became a “developmental state” (Knight and Ding 2012). Only within the last decade have efforts to promote a “harmonious society” brought issues of income inequality—other than landlessness—to the policy agenda. We have argued that the new policies to redress inequality can be taken further by strengthening transfers of income and by equalizing opportunities for income generation.

The powers given to officials in pursuit of economic growth and their lack of accountability generated rent seeking, corruption, and procedural injustice, all of which contributed to the growth of localized and national income inequality. Can other developing countries follow China’s example to create a developmental state that drives rapid economic growth and yet avoids the rise in income inequality that this process has produced in China? The most important lesson that China’s experience offers other countries lies in the answer to this question.

Notes

John Knight is an Emeritus Professor in the Department of Economics, Manor Road Building, University of Oxford, OX1 3UQ, and China Institute of Income Distribution, School of Economics and Business Administration, Beijing Normal University; email: john.knight@economics.ox.ac.uk. I am grateful to Ravi Kanbur, Li Shi, Martin Ravallion, Terry Sicular, and the editor and three referees for their helpful comments.

1. Migrants are excluded for comparison with earlier years. Including migrants, the urban and national Gini were 0.33 and 0.49 in 2007 (Li et al. 2013).

2. The authors’ methodology was criticized in Luo et al. (2012), but their general conclusion was not disputed.

3. These figures are derived from the CHIP volumes for the 1995, 2002, and 2007 surveys. There are discrepancies among the sources, but it is clear that the percentage rose strongly and then remained fairly constant.

4. Five categories of happiness are converted into a cardinal score, with “very happy” having a value of four and “not at all happy” having a value of zero.

5. It is possible to distinguish between absolute and relative income because of the wide range of mean household income per capita among cities and among villages.

6. Figures 2.3 and 2.5 are derived from the Gallup World Poll, and figure 2.9 is derived from the World Values Survey.

References


Can Civil Society Overcome Government Failure in Africa?

Shantayanan Devarajan, Stuti Khemani, and Michael Walton

Government failures are widespread in Africa. Symptoms include absentee teachers, leakage of public funds, monopolized trucking, and employment-restricting regulations. Can civil society do anything about these failures? Would external donor support to civil society help? We argue that the challenge for civil society is to improve government functioning by strengthening political incentives—the underlying cause of government failure—rather than bypassing or supplanting the state. This paper reviews the available evidence on civil society interventions from this perspective. Although the current increase in political competition and extensive citizen engagement in Africa seems to create the potential for civil society influence, we find that there are large knowledge gaps regarding what works, where, and how. Some rigorous evaluations find significant impacts of civil society involvement on development outcomes, but these studies typically pay insufficient attention to the mechanisms. For example, are impacts due to overcoming government failure or to changing private household behavior, leaving the wasteful allocation of public resources untouched? We conclude that donor support to civil society should take an approach of learning by doing through ongoing experimentation backed by rigorous, data-based evaluations of the mechanisms of impact.

JEL codes: H41, O19, P26

Consider the following facts:

- Teachers in public primary schools in Uganda are absent 20 percent of the time; when present, they are in class teaching 18 percent of the time (Chaudhury et al. 2006; World Bank 2007).
- Only 1 percent of nonsalary expenditures allocated to primary health clinics in Chad actually reach the clinics (Gauthier and Wane 2009).
- There are huge leakages from fertilizer subsidies in Tanzania, with elected officials receiving 60 percent of the vouchers (Pan and Christiansen 2011).
• Efforts to hire workers below the minimum wage in South Africa—a country whose unemployment rate is 25 percent—are met with widespread, sometimes violent, protest.
• Because of regulations that prohibit entry, monopoly rents to trucking companies cause Africa to have the highest transport prices in the world (Teranavithorn and Raballand 2009).

In addition to being emblematic of the problems facing Africans, especially poor Africans, these facts can be explained by failures of “accountability” at various points in the chain of relationships for the implementation of public sector policies. This chain goes from the preferences and needs of citizens, through the political process, and then passes to the bureaucracies and front-line actors charged with delivering services. It also passes to the regulators, judges, and others with responsibility for monitoring and enforcing public behavior. Teacher absenteeism and the leakage of public funds are examples of politicians or central-level bureaucrats’ inability or unwillingness to hold lower-level bureaucrats and service providers accountable. The capture of fertilizer subsidies or wage premiums by politically powerful groups are instances of how special interests are able to shape the behavior of state actors, whereas the broader citizenry, who are unorganized, remain unable to make politicians accountable to their needs. We call these accountability failures “government failures.” They are analogous to market failures in that public officials act in their own interest, leading to an equilibrium that is socially suboptimal.

The development financing community of bilateral and multilateral donors, including the World Bank, has recently begun to explore how these problems of state accountability might be overcome by civil society organizations. Does a substantially strengthened role for civil society have a sound conceptual and empirical basis, or is it just another fad? Even if there is a sound, substantiated case for some forms of civil society action, can these external actors play a useful role in supporting it, or will they only add a new set of distortions?

This paper attempts to answer these questions by reviewing the available literature and development experience through the lens of accountability relationships. We emphasize the fundamental importance of the political economy in explaining government failures in African societies. We then assess the extent to which civil society interventions have improved outcomes by reducing government failures rather than bypassing or supplanting the state.

The analytics and evidence support a common conclusion: there is a strong prima facie case for a strengthened role of civil society in both democratic and semiauthoritarian regimes. However, strategies and interventions in this regard need to focus on mechanisms for reducing government failures rather than increasing the burden on citizens to help themselves in ways that leave state failures largely intact.
More specifically, in attempting to improve government performance and accountability, civil society interventions face the challenge of pervasive rent seeking and “clientelism” in Africa, where politics revolves around providing narrowly targeted benefits (often to particular ethnic, religious, or regional groups) in exchange for political support at the expense of broad public goods. In this setting, unqualified faith in civil society as a force for good is likely to be misplaced. Historically created institutions of inequality or ethnic identity can often inhibit collective action in the broader public interest, promoting narrow sectarian interests and nourishing clientelistic political competition. Furthermore, public-interest action by civil society is heavily constrained by existing systems and institutions of the state.

Nevertheless, civil society action can achieve incremental, and possibly transformational, success in addressing accountability failures. We focus here on the potential for incremental change, building on existing developments in Africa such as greater political contestation and citizen organization.

What does this potential for civil society action mean for external donors? It implies that support for civil society should be both organic and experimental: organic in the sense that interventions contribute to change in existing political and societal structures as opposed to seeking to bring “best practice” ideas from outside; and experimental in the sense that there should be structured monitoring, information generation, and evaluation in the process, with the techniques depending on the nature of the intervention.

In the next section, we examine government failure in the form of breakdowns in accountability relationships and explore how civil society action may or may not help. The following section summarizes our review of the existing evidence on the impact of civil society-related initiatives to improve accountability. A final section discusses the implications for external donors.

Accountability Relationships and Civil Society Action in an African Context

This section develops the theoretical framework underpinning the arguments regarding the role of civil society in Africa.

The Accountability Framework

We use the accountability framework of the 2004 WDR, *Making Services Work for Poor People*, presented in figure 1, to examine government failure as breaks in the “long route of accountability.” The argument is as follows. In a private market
transaction where market institutions are reasonably competitive and free from information problems—such as buying a sandwich from a vendor—there is direct accountability of the sandwich provider to the client or consumer. The client pays the vendor directly, she can observe whether she receives the sandwich, and if the market is reasonably competitive, the client can go elsewhere if she does not like the sandwich—and the vendor knows this. This direct market relationship is the “short route of accountability” in figure 1, exercised through client power.

When the state is involved (for example, in response to market failures or redistributive goals), the relationship between the client and the service provider is mediated by the institutions that shape the incentives and behavior of state actors. In the case of publicly provided education, for instance, public school teachers are managed by and answerable to state bureaucrats, termed the “compact” in this framework, and are only indirectly answerable to citizens. Problems of teacher absenteeism are consequences of weak compacts in which teachers are not held accountable by state bureaucrats for showing up and teaching effectively. Why are teachers not held accountable by bureaucrats? The reason, we argue, has to do with politics: teachers’ jobs are often used as political patronage. Teachers help politicians get elected, in return for which they obtain jobs from which they can be absent. In other words, weak political incentives in the long route of accountability lead state actors to select compacts that deliver protected teaching jobs in the public sector rather than the broad public service of quality education.

Our main argument, therefore, is that government failure is fundamentally shaped by the first link in the long route, the relationship between citizens and...
politicians, which, in turn, determines the behavior of other state actors. A majority of citizens may prefer teachers to show up and teach effectively. However, imperfections in the political process may result in the election of politicians who provide protected jobs to teachers as opposed to those with an interest in reforming teacher incentives. In the case of education, in addition to jobs as patronage, teachers may be politically powerful. For example, the South African Democratic Teachers Union is both part of the governing coalition and one of the most powerful trade unions in the country (New York Times 2009). In the legislature of the Indian state of Uttar Pradesh in the early 2000s, teachers constituted 20 percent of the assembly, and former teachers constituted another 20 percent (Kingdon and Muzammil 2001).

Substantial public spending on the salaries of absent teachers is difficult to explain by policymakers’ lack of knowledge of service delivery conditions or lack of access to technology and mechanisms to reduce absenteeism. In this setting, civil society action, such as citizen-based school management committees, may fail to reduce teacher absenteeism if higher-tier bureaucrats do not have political incentives to make teachers or school administrators responsive to local citizens. Success is unlikely if incentives are geared toward protecting jobs in the public sector rather than improving the quality of education. Local school committees may also be captured by local political elites who prefer to safeguard the power of patronage in public sector jobs.

Political economy problems exist everywhere, but recent literature on the persistence of underdevelopment in regions such as Africa argues that political failures in these contexts are severe, institutionalized, and self-perpetuating. One argument is that the unequal distribution of endowments and power leads to state institutions that encourage elites to organize to extract rents from state resources (North et al. 2009; Acemoglu et al. 2001, 2005; Engerman and Sokoloff 1997; Rajan 2009; Acemoglu and Robinson 2012). Political elites share rents with economic elites but fail to deliver general public goods, such as credible commitment to all investors, the rule of law, market institutions that support entry and competition, and social and other services for the general population.

Elite privileges can be sustained through state repression, especially under authoritarian regimes but also through “clientelist” strategies, in politically competitive or democratic regimes. Clientelism involves the provision of narrowly targeted benefits to particular voters in return for their political support; it allows political elites to get away with high rents and low provision of broad public services (Keefer and Khemani 2005; Kitschelt and Wilkinson 2007; Vicente and Wantchekon 2009). This combination of rent sharing among elite groups and clientelism is pervasive in Africa (van de Walle 2001; Azam 2001). Elite rent sharing and clientelism are often linked to identity-based politics on the basis of ethnicity, religion, or region (or all three overlapping, as they do in Nigeria and Sudan, for example). Ethnic fragmentation and polarization have been blamed
for enabling politicians to win and remain in public office despite the underprovision of broad public goods (Easterly and Levine 1997; Alesina, Baqir and Easterly 1999; Montalvo and Reynal-Querol 2010). However, ethnicity is not necessary for clientelism. In Tanzania, for example, identity-based links are much less salient than in many other countries, in part because of a concerted nation-building strategy, but clientelism remains rampant (Kelsall 2002). Such identity-based links are not peculiar to Africa. India is a consolidated democracy in which identity of caste and religion is highly politically salient (Chandra 2004). Widespread poverty can allow clientelist strategies, such as vote buying, to be successful and can hinder collective citizen organization to demand broad public goods (Stokes 2007).

Politicized bureaucracies can perpetuate themselves. Even if a reform-minded politician wants to deliver public goods, he may be unable to do so given the pervasive weaknesses in the state’s bureaucratic infrastructure. A clientelist political strategy may be forced upon him for political survival. A highly committed senior education officer can do little if teachers are hired for reasons of political patronage and not to teach. A mining official determined to meet citizens’ complaints of flouting environmental standards will not be tolerated if politicians who are party to rent sharing address the mining companies.

Elite rent extraction and clientelist political strategies tend to sustain and reproduce inequalities. State benefits are targeted to those groups with greater capacity to organize (e.g., special interest groups, unions) or those with ethnic, geographic, or other identity-based ties to elites in power. Local elites are selected and supported by clientelistic networks, inhibiting the development of other forms of local organization by citizens, or by programmatic political parties, who may demand broader public goods from the state.

Can civil society action break this vicious cycle and improve the accountability of the state to its broader citizenry? How can civil society organizations mobilize demand for better quality public policies and reduce the power of narrowly targeted patronage in winning political support?

**Emerging Potential for Civil Society Action in Africa**

Civil society action in Africa is emerging at a time of unprecedented political contestation and citizen engagement. There is evidence of extensive political participation and associational activity in both democratic and (semi)authoritarian polities. Data from the Polity IV effort, which uses the expert opinions of political scientists to categorize the extent of autocracy and democracy of a regime, illustrate the scale of changes. There is a continuum of regimes, from completely autocratic to fully consolidated democracies, with a large range of intermediate types (called “anocracies”) that have elements of both autocratic and democratic processes. Their synthetic classification uses a numerical system, in which −10 to −7 is classified...
Figure 2 illustrates the difference between 1985 and 2009. In the mid-1980s, 31 out of a total of 40 countries in the database were fully autocratic, with only two full-fledged democracies (Gambia and Mauritius) and six intermediate regimes, most of which had a heavy predominance of autocracy. By late 2009, the Polity IV classification listed 12 fully democratic regimes (Benin, Botswana, Comoros, Ghana, Kenya, Lesotho, Mali, Mauritius, Senegal, Sierra Leone, South Africa, and Zambia) and only three fully autocratic regimes (Eritrea, Somalia, and Swaziland). The majority are classified as intermediate (anocratic) regimes, most with significant democratic elements (including, for example, Burkina Faso, Côte d’Ivoire, the Democratic Republic of Congo, Ethiopia, Mozambique, Nigeria, Rwanda, Uganda, Tanzania, and Zimbabwe).

Data from the Afrobarometer surveys show large-scale citizen participation and mobilization across countries. Table 1 provides a few indicators of the scale of citizen participation across countries. Interestingly, reported participation is not strikingly different between countries classified as “full democracies” and those

---

**Figure 2.** Democratic and Autocratic Regimes in Sub-Saharan Africa, 1985 and 2009

Source: Polity IV database.
Averagely self-reported voter turnout in the 2008 round was more than 70 percent in “full democracies” and 60–70 percent for the intermediate cases. Among the full democracies, the highest-ranked countries (South Africa and Botswana) reported active group membership of only 16 and 17 percent compared to 43 and 45 percent for Kenya and Liberia, both of which are lower on the “democracy” scale. Among intermediate regimes, group membership and attendance in meetings is particularly high for Tanzania.7

Even among the countries with the lowest Polity IV scores, an average of 41 percent state that they have often joined others to raise an issue, and 45 percent state that they are willing to protest or demonstrate as part of citizen action. These findings indicate both the potential and pitfalls of civil society action. Even in less

### Table 1. Civil Society Activity across Countries

<table>
<thead>
<tr>
<th>Country</th>
<th>Voted in last elections</th>
<th>Active member of a group</th>
<th>Often attended group meetings</th>
<th>Often joined others to raise an issue</th>
<th>Willing to demonstrate or protest</th>
<th>Punishment likely if people complain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polity IV Score between 10 and 6 (“full democracies”)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Africa</td>
<td>65</td>
<td>16</td>
<td>29</td>
<td>20</td>
<td>45</td>
<td>46</td>
</tr>
<tr>
<td>Botswana</td>
<td>59</td>
<td>17</td>
<td>55</td>
<td>39</td>
<td>52</td>
<td>19</td>
</tr>
<tr>
<td>Lesotho</td>
<td>61</td>
<td>28</td>
<td>67</td>
<td>50</td>
<td>52</td>
<td>74</td>
</tr>
<tr>
<td>Ghana</td>
<td>81</td>
<td>35</td>
<td>40</td>
<td>38</td>
<td>30</td>
<td>32</td>
</tr>
<tr>
<td>Senegal</td>
<td>79</td>
<td>28</td>
<td>60</td>
<td>55</td>
<td>62</td>
<td>60</td>
</tr>
<tr>
<td>Mali</td>
<td>77</td>
<td>29</td>
<td>57</td>
<td>55</td>
<td>59</td>
<td>46</td>
</tr>
<tr>
<td>Kenya</td>
<td>79</td>
<td>43</td>
<td>44</td>
<td>35</td>
<td>40</td>
<td>37</td>
</tr>
<tr>
<td>Benin</td>
<td>91</td>
<td>23</td>
<td>53</td>
<td>50</td>
<td>65</td>
<td>25</td>
</tr>
<tr>
<td>Zambia</td>
<td>59</td>
<td>24</td>
<td>45</td>
<td>33</td>
<td>45</td>
<td>47</td>
</tr>
<tr>
<td>Liberia</td>
<td>78</td>
<td>45</td>
<td>50</td>
<td>42</td>
<td>24</td>
<td>40</td>
</tr>
<tr>
<td>Malawi</td>
<td>75</td>
<td>22</td>
<td>60</td>
<td>54</td>
<td>50</td>
<td>27</td>
</tr>
<tr>
<td><strong>Avg.</strong></td>
<td><strong>73</strong></td>
<td><strong>28</strong></td>
<td><strong>51</strong></td>
<td><strong>43</strong></td>
<td><strong>48</strong></td>
<td><strong>41</strong></td>
</tr>
<tr>
<td>Polity IV Score between 6 and 2 (“higher scoring intermediate democracies”)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mozambique</td>
<td>60</td>
<td>18</td>
<td>49</td>
<td>45</td>
<td>57</td>
<td>47</td>
</tr>
<tr>
<td>Nigeria</td>
<td>60</td>
<td>27</td>
<td>30</td>
<td>26</td>
<td>43</td>
<td>57</td>
</tr>
<tr>
<td><strong>Avg.</strong></td>
<td><strong>60</strong></td>
<td><strong>23</strong></td>
<td><strong>40</strong></td>
<td><strong>36</strong></td>
<td><strong>50</strong></td>
<td><strong>52</strong></td>
</tr>
<tr>
<td>Polity IV Score between 2 and –2 (“lower scoring intermediate democracies”)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>61</td>
<td>16</td>
<td>51</td>
<td>38</td>
<td>42</td>
<td>44</td>
</tr>
<tr>
<td>Madagascar</td>
<td>70</td>
<td>4</td>
<td>75</td>
<td>29</td>
<td>20</td>
<td>61</td>
</tr>
<tr>
<td>Burkina</td>
<td>72</td>
<td>18</td>
<td>59</td>
<td>52</td>
<td>63</td>
<td>53</td>
</tr>
<tr>
<td>Uganda</td>
<td>70</td>
<td>31</td>
<td>44</td>
<td>33</td>
<td>35</td>
<td>41</td>
</tr>
<tr>
<td>Tanzania</td>
<td>82</td>
<td>35</td>
<td>64</td>
<td>51</td>
<td>65</td>
<td>57</td>
</tr>
<tr>
<td><strong>Avg.</strong></td>
<td><strong>71</strong></td>
<td><strong>21</strong></td>
<td><strong>59</strong></td>
<td><strong>41</strong></td>
<td><strong>45</strong></td>
<td><strong>51</strong></td>
</tr>
</tbody>
</table>

*Source: Afrobarometer Round 4, 2008.*
democratic societies, the relatively high degree of citizen participation could be a basis for greater participation and, possibly, political contestation. At the same time, those who participate may face obstacles, such as fear of retribution. Furthermore, citizens may participate to serve ethnic interests, for example, leading to even more perverse politics.

Further analysis of the Afrobarometer data suggests that both hypotheses are possible. Local-level organization of citizens into groups can include the poor. For six countries—Benin, Botswana, Ghana, Senegal, South Africa, and Zimbabwe—an indicator of poverty in the Afrobarometer (when a respondent says the household has gone without adequate food many times in the recent past) is significantly associated with greater attendance at group meetings (after controlling for respondent age, gender, education, media access, household size, and neighborhood availability of infrastructure).

At the same time, citizens report fear of punishment or reprisals should they complain about the poor quality of government services or the misuse of government funds. The results from seven countries—Benin, Lesotho, Mozambique, South Africa, Tanzania, Uganda, and Zimbabwe—indicate that poor respondents are more likely to report these fears.

Furthermore, in five countries—Benin, Ghana, Senegal, Zambia, and Zimbabwe—respondents who report frequent attendance in group meetings are also more likely to have ethnic grievances. This finding suggests that when civil society groups pursue sectional interests, they are potentially a source of distortions. In the public-choice literature, the power of organized “special interests” has been blamed for distortions to growth-promoting or efficient or equitable fiscal policies, even in advanced industrial democracies such as the United States (Tullock 1959; Becker 1985). An extreme and tragic example is the Hutu mobilization (and the use of radio within this mobilization) in the Rwandan genocide (Yanagizawa-Drott 2012).

Yet, detailed research on the political behavior of citizens in Africa suggests that there is significant potential for change. Citizens do not respond only to clientelistic appeals and ethnic identity; government performance in managing the economy and delivering development matters to them as well (see, for example, Bratton et al. 2011; Young 2009). There are examples of civil society organizations transcending the “special-interest” type of distorting action and trying to mobilize citizens to demand broader public goods. Effective mobilization may bring sanctions upon non-performing public agents and may reward those with good performance. Civil society action may lead to changes through the electoral process (in which nonperforming incumbents are voted out of power) or by activating other institutions for sanctions within the state, such as internal bureaucratic structures, or formal independent institutions within the state, such as the judiciary and auditing departments.

The critical point is that to successfully reduce government failures, these actions need to address the underlying political incentives. For example, mobilizing citizens...
around the quality of education in a community where politics revolves around patronage and where the mobilization effort has not taken that into account is likely to fail. Providing citizens with information about teacher absenteeism in this context may be superfluous; citizens are already aware of it, but they are unable or unwilling to do anything because they know teachers’ jobs are politically protected. Opportunities for successful civil society action in the public interest may emerge in a community where local institutions have facilitated local collective action or in a sector where broad citizen demand for improved public services has emerged. In contrast, opportunities may arise at a higher level, around national or provincial state entities, because of broader-based political competition, for example. Civil society action at that level may facilitate institutional changes that alter practices or discipline errant local providers.

In sum, the changing political context in Africa suggests that there is significant scope and potential for civil society action. Is there evidence of effectiveness? In the next section, we consider evidence of the success of civil society interventions in mobilizing citizens to demand better performance from the state. We continue the education example, among others, to assess whether available impact evaluations that have improved education outcomes have done so by reducing the political power of teachers. Did previously errant public school teachers become more responsive to citizen demands? Or did outcomes improve because parents contributed more of their own resources to education, including greater financial support to teachers (a form of higher local taxation), leaving patronage teachers unscathed?

Evidence of the Mechanisms of the Impact of Civil Society Interventions

We focus on interventions that have been rigorously assessed, organizing them on the basis of which “arm” of the accountability triangle (fig. 1) they principally address. The most common interventions assessed in the policy literature involved attempts to strengthen the client power relationship, typically via some combination of providing information to communities on the performance of local service providers and direct support for local collective action. Broader transparency and information interventions, where civil society organizations track and monitor government programs, budgets, and policy performance, are attempts to influence the compact relationship. Finally, a few interventions have directly focused on the political relationship by providing information on politicians. Regardless of which arm the interventions directly targeted, we consider whether the pattern of evidence suggests that any impact on development outcomes was achieved by overcoming government failure rather than by private efforts bypassing the state.
Client Power Interventions

One source of inspiration for the recent trend toward civil society engagement is the experience of a civil society organization based in Bangalore, India, the Public Affairs Centre, which pioneered the use of “citizen report cards” as a tool for client power. In the anecdotal evidence provided by Paul (2002), this initiative aggregated citizens’ perceptions of the quality of urban services in the city of Bangalore and then publicized these perceptions through local media, serving as a massive, collective complaint. The quality of services improved over time. The intervention either “shamed” the agencies into improving services or sent a strong signal to local politicians that citizens care about service delivery, leading politicians, in turn, to pressure the providers to improve performance.

Two studies of client-power interventions in health and education in Africa, both from Uganda, stand out in the literature for their identification of significant positive impacts on outcomes within public delivery systems (Björkman and Svensson 2009; Barr et al. 2012). However, as we argue below, neither is able to clarify whether this impact was achieved because a government failure was overcome or because communities were mobilized to make additional private contributions to public services—a form of local taxation—without addressing government failure. A third study from a different context, rural India, which is able to say something on mechanisms, finds that very similarly structured client-power interventions (as in Uganda) had no impact on the public system (that is, on government failure) but improved outcomes through greater private efforts of communities that bypassed the public schools (Banerjee et al. 2010c). This suggests that much more evidence is needed on whether or what types of citizen mobilization interventions might work in Africa to address government failures.

Björkman and Svensson (2009) undertake a randomized control trial of a citizen report card intervention in the health sector in Uganda. In this study, local civil society organizations were trained to compile data on citizens’ perceptions of the quality of services at local health clinics and to use these as the basis of discussions between selected community members and the health providers. On average, the intervention communities experienced significant improvements in health services and in health outcomes (as measured by weight for age and under-five mortality). However, there is insufficient evidence on the mechanisms by which this impact was achieved. Did the interventions generate such striking impacts on health outcomes because they increased private demand for health services such as immunization? Or did they overcome government failures by providing incentives to health providers to improve the quality of service delivery?

One piece of evidence is consistent with the latter interpretation—the interventions may have made the directly elected Health Users Management Committees more responsive to citizens for the quality of services. In intervention
areas, on average, the Health Users Management Committees experienced significant turnover in elected positions after the citizen campaign. This finding reinforces the importance of addressing local political incentives if outcomes are to be improved.

Follow-up work in this area in Uganda lends support to some of the specific hypotheses in section 2, which suggest that particular institutions in Africa, such as ethnic identity and historical inequality, can thwart collective action to overcome government failures. Björkman and Svensson (2009) find that the impact within intervention communities was particularly sensitive to the degree of ethnic heterogeneity and wealth inequality. There were significant impacts only in those communities that were relatively homogeneous along both ethnolinguistic and economic dimensions. How civil society interventions might overcome these divisions remains an area for future research and policy learning.

The second strand of evidence from Uganda comes from education. Previous research has revealed that a first-order accountability failure in education is teacher absenteeism. In surveys across several developing countries in 2002, Uganda had the highest rate of teacher absenteeism, at 27 percent (Chaudhury et al. 2006). Subsequent work supports the view that this was not due to poor working conditions (Chaudhury et al. 2006; Habyarimana 2004). Government teacher salaries in Uganda are significantly higher than the market wages of individuals with similar qualifications, and variation in the infrastructure conditions of schools and communities is not correlated with variations in teacher absence. More detailed research from other regions, notably South Asia and Latin America, suggests that poor public school teacher performance is sustained because of the political power of teachers as an organized interest group and because of the value of extending patronage through teaching jobs (Grindle 2004; Pritchett and Murgai 2006; Béteille 2009). In this context, can civil society interventions mobilize communities to demand better performance from teachers?

Barr et al. (2012) provide evidence that a community monitoring intervention, albeit implemented by government agencies rather than civil society, reduced teacher absenteeism and improved learning outcomes in Uganda. These authors attempt to address the question of mechanisms of impact through direct evidence that the successful intervention resolved collective action problems. The results of behavioral experiments—which involve playing games with participants in a laboratory setting—suggest that parents in intervention villages were more willing to make voluntary contributions to public goods. This is an intriguing and valuable result, but it does not, on its own, answer the question of whether the improvements occurred because the government failure was addressed. Did outcomes improve because parents contributed more of their own resources to education, including greater financial support to teachers, or because
previously errant public school teachers became more responsive to citizens’
demands?

Furthermore, the result hinges on the behavior of those teachers who contin-
ued their tenure at the sampled schools in the two years between baseline and
follow-up and is particularly sensitive to the length of tenure. Frequent teacher
turnover is a significant issue in Uganda—36 percent of teachers who were on the
schools’ books at baseline were no longer there at follow-up. If politically connect-
ed teachers are able to move to a different school (and continue to get away with
poor performance), then the intervention may have had no real effect on the gov-
ernment failure. Turnover rates were no different in treatment and control
schools.

Evidence on whether the mechanism of impact worked through changes in the
public sector, or private action, is available from another context—rural India—and
has implications for the design and evaluation of future work in Africa. A similar
school-based citizen report card exercise as in Uganda was organized in rural India,
in a province known for local patronage politics (Uttar Pradesh), and had no impact
on either teacher effort or learning outcomes in public schools (Banerjee et al.
2010c). The intervention also had no impact on local school committees, the
Village Education Committees, which had nominally been created for greater local
agency, monitoring, and participation but were found to be defunct at baseline. In
contrast to this complete lack of impact on the public system, the initiative success-
fully mobilized private action by local youth volunteers to hold remedial classes
outside school. Children who attended the volunteer-led classes made dramatic im-
provements in learning. In short, some people were able to take private action to
improve education, but they were not able to hold the state accountable for better
quality public education.

A review of the larger literature on the performance of citizen monitoring and
participation through local committees shows more systematically that success
depends upon the nature of social organization within communities (particularly
the degree of inequality) and, importantly, on whether higher-tier state agencies
provide the power and resources to local committees to be effective (see Mansuri
and Rao 2013). The outstanding question is whether higher-tier agencies have
sufficient political incentives to devolve power and to structure local institutions
appropriately. Moreover, such committees are often not autonomous civil society
organizations outside state structures because they are designed to be part of state
institutions, with representation of politicians and bureaucrats.12 A recent study
from Latin America argues specifically that state representatives can undermine
parental participation in community-managed schools (Altschuler 2013).13 In
sum, more research and evidence is needed on whether civil society can “activate”
existing state institutions to perform better.
Client Power and the Compact as Alternatives or Complements

Some studies have sought to directly compare the effectiveness of civil society initiatives that work through client power with those that work through strengthened sanctions in the compact between bureaucrats and service providers. Is the exertion of client power through civil society engagement more effective than technocratic changes to the compact between governments and their local agents? For Indonesia, Olken (2005) compares the impact of information provision to local citizens about corruption in local roads projects against the impact of the provision of information on a technocratic government-led initiative—credible audits of local projects by higher tier authorities. The state-led audits are substantially more effective at reducing corruption. The unaddressed question is whether civil society interventions might complement government incentives to undertake credible audits to overcome corruption.

By contrast, a study with a similar design in Madagascar that compares interventions designed to encourage district education bureaucrats to improve their monitoring of schools (versus directly engaging school communities to do so) shows that local-level monitoring is more effective (Nguyen and Lassibille 2008). A key difference of the audit intervention in Indonesia, however, is that the Madagascar intervention did not require district officials to improve their monitoring. Again, the question remains whether civil society action can more efficiently achieve outcomes by changing bureaucratic behavior or by encouraging community participation.

A new study from Kenya provides an important insight on this issue (Bold et al. 2012). It evaluates a policy of using contract teachers—lesser-paid teachers without the security of tenure compared to regular civil service teachers. In previous work with NGO-managed schools in Kenya (Duflo et al. 2012), the use of contract teachers is reported to be highly cost effective because contract teachers are paid lower salaries than regular civil service teachers, yet they have lower absence rates and are associated with higher test scores. Bold et al. (2012) provide a follow-up study in Kenya in which the intervention involved a government-supported contract teacher pilot program administered under two alternative management regimes, one run by an NGO and one by the government. They find significant positive effects of contract teachers on student learning in schools administered by the NGO, but none in those directly managed by the government. They provide descriptive evidence that the lack of impact in the government-managed intervention was associated with a fierce reaction of the teachers’ union to the use of contract teachers, which led to both salary delays and eventual agreement by the Ministry of Education to make the contract teachers permanent civil servants at the end of the two-year period.

This evidence is consistent with our arguments in section 2 that political incentives (teachers’ unions are politically powerful in Kenya) can thwart efforts (such as
hiring contract teachers) to overcome government failures, even if these efforts were shown in other settings to yield superior results in terms of cost effectiveness and student learning compared with civil servant teachers. Once again, the unanswered question relates to the role of civil society not as direct providers or managers of services (such as the NGO in this Kenya case) but as agents that can help change government incentives.

**Compact Interventions: Country- or Sector-wide Information and Accountability Initiatives**

Many civil society initiatives seek to bring about greater transparency in government budgets and programs. The assumption is that when civil society brings more information to light, governments will be pressured to address their failures. Citizen action and donor pressure on governments have contributed to the adoption of legislation for citizens’ “Right to Information” and various initiatives to facilitate citizens’ ability to monitor public budgets and the allocation of public resources, such as the Extractive Industries Transparency Initiative and the Open Budget Index.

Much of the evidence on how these initiatives work and their impact is qualitative and focuses on intermediate outcomes, such as whether citizens became more informed and engaged in budget processes (McGee and Gaventa 2010, and Gaventa and Barrett 2010, provide reviews). Some quantitative studies reviewed by Carlitz (2010) provide mostly cross-country correlations of indicators of greater budget transparency with indicators of governance, corruption, socioeconomic and human development, and even political participation. However, correlations say nothing about causation; it is not possible to draw conclusions about whether specific budget transparency initiatives lead to better outcomes or whether other underlying changes in government accountability to citizens drive the outcomes.

A small number of studies provide more rigorous results, highlighting the importance of understanding mechanisms of impact. Keefer and Khemani (2011a, b) examine the role of community radio, a medium that is particularly suited to conveying information to poor citizens in Africa. They find that in Benin, greater access to community radio is not associated with the ability of citizens to extract greater benefits from their government. Villages with greater access to information through community radio did not receive more or better inputs for their public schools, nor did they receive more bed nets to protect their populations from malaria. Instead, households were persuaded by the public interest programming on the radio to increase their private investments in the education of their children and the health of their family. Although this is a useful role for mass media to play in development, it is not evidence that this mechanism will address government failures. The finding in Benin contrasts with evidence from more mature democracies (the United States and India) of greater government responsiveness to more informed
citizens (Stromberg 2004; Besley and Burgess 2002). It may be that the information provided by community radio in Benin was not politically salient or that the ways in which issues were “framed” mattered (Prat and Strömberg 2011). Inasmuch as politics in Benin is characterized by clientelist provision of narrowly targeted benefits at the expense of broad public services (Wantchekon 2003), community radio broadcasts may not frame these service delivery issues in terms of government accountability, and citizens may not act upon information to demand greater government accountability (relying instead on private actions).

The Benin findings also contrast with an intervention in Uganda in which media were used more purposefully by higher-tier authorities. In this case, PETS measured discrepancies between budget allocations to schools and the amounts that actually reached the schools (Reinikka and Svensson 2004). After finding large-scale diversion of budgeted funds away from schools, the Ministry of Education began publishing information about school entitlements. This information campaign has been credited with reducing the “leakage” of funds (Reinikka and Svensson 2005). However, the lesson for the role of civil society in this case is unclear—did information provision succeed because it was led by the Ministry and served as a signal to district bureaucrats that any leakage would be punished?

Following the celebrated example of Uganda, higher-tier governments and external donors have supported a proliferation of PETS activities. They have sought to strengthen the capacity of civil society organizations to undertake surveys to uncover and publicize local leakage and thereby to stem it. Qualitative evaluations of these efforts suggest that they can be frequently successful in getting funds to reach the intended destinations, although to a lesser degree than the estimates in Uganda (McGee and Gaventa 2010). More rigorous impact evaluations are needed, especially to address the recurring questions in this paper: Is civil society more effective than strengthened institutions within the state? What role can civil society play in strengthening state institutions?

Other explanations of the Uganda experience suggest that it was driven by larger changes in political incentives; there was a push from higher-level government leaders to enforce the implementation of their school budget allocations (Hubbard 2007). In particular, free and universal primary education was a prominent theme in President Museveni’s 1996 election campaign (Uganda is considered a semiauthoritarian regime by Polity IV—see fig. 2 above). This explanation suggests the importance of the role of citizens or civil society as voters and the demands that they make of their political representatives.

African cross-country evidence shows that increasing political competition is associated with the abolition of school fees and greater access to education but is not necessarily associated with improvements in education quality (Harding and Stasavage 2011). A puzzle that remains unaddressed in PETS-type interventions is why other information about the wastage in public spending, such as large-scale
teacher absenteeism, does not lead to public action. The school grants covered by the PETS in Uganda are a much smaller proportion of government education spending than teacher salaries (Hubbard 2007). One reason may be the political constraints to improving teacher performance in the public sector. A body of evidence across states of India (reviewed in Khemani 2010), where more data are available, is consistent with the use of teacher hiring at election times to win support through patronage rather than by improving the quality of education. Can civil society interventions undercut such patronage politics and mobilize citizens to more effectively demand better quality education?

Interventions on the “Politics” Arm: Improving Political Accountability for Broad Public Goods

In the first round of the 2006 Presidential elections in Benin, a civil society group organized town hall meetings with political candidates to discuss specific policy proposals informed by empirical evidence. Where such town halls were held, voter turnout was higher and support for clientelist political platforms was lower (Wantchekon 2009). A campaign by a civil society organization in India to persuade voters not to vote on the basis of candidates’ caste identity was effective in increasing voter turnout and reducing the votes of caste-preferred candidates with criminal records (Banerjee et al. 2010a). Another civil society campaign in India to provide information on politicians’ performance in delivering benefits to their constituents led to poorly performing incumbents being voted out of office (Banerjee et al. 2010b). However, a similar experiment in Uganda with the African Leadership Initiative that provided information about the legislative activities of Members of Parliament and their efforts at spending their constituency development funds had little or no impact on election outcomes (Humphreys and Weinstein 2010).

In São Tomé and Príncipe and Mozambique, voter education campaigns were undertaken to persuade voters not to “sell” their vote. Previous work in São Tomé and Príncipe had documented that vote-buying practices increased dramatically in the late 1990s after the discovery of oil (Vicente 2010). As discussed in section 2, such practices enabled political leaders to extract large rents from public resources while providing low quality and a low quantity of public services. In Nigeria, community campaigns were undertaken by civil society groups to encourage voters to oppose political intimidation or violence (Collier and Vicente 2010).

These are important examples of the engagement of civil society in strengthening politics, but this body of work is not designed to examine the ultimate impact on policies and development outcomes. The work focuses on measuring the impact on specific political outcomes, such as voter turnout and the use of different electoral strategies by incumbents and challengers. It is therefore not immediately clear...
whether even significant electoral effects of voter education campaigns effectively translate into different or better policy choices. For example, by reducing the efficacy of vote-buying tactics employed by political challengers, the campaigns may primarily strengthen the hands of incumbent politicians, protecting them from the risks of losing office and enabling them to continue rent-extraction policies. An example from the Philippines suggests that addressing the proximate symptoms of clientelist political competition may not be effective. There, voter education campaigns to reduce vote-buying practices had the unintended effect of offending target groups—the poor—who were convinced of the legitimacy of receiving benefits from politicians in exchange for their vote. These campaigns may also have intensified “class divides” between the poor and the middle or upper classes who were the sponsors of the campaigns (Schaffer 2005).

**Interventions Addressing All Three Accountability Relationships**

Although we have attempted to assign specific types of interventions to one or the other of the three arms of accountability in figure 1, many interventions aim to address all three. An example is the institution of “participatory budgeting” pioneered in the 1990s in Porto Alegre, Brazil, which seeks to mobilize civil society to participate actively in the formulation and implementation of municipal budgets. Together with the citizen report cards of Bangalore, participatory budgeting in Porto Alegre has served as an inspiration for the trend toward civil society engagement.

Participatory budgeting has been credited in qualitative and descriptive analysis with substantial improvements in local governance and responsiveness to the needs of the poor, including significant changes in budget allocations. There is, however, a particular problem in identifying causation. The introduction of participatory budgeting in Porto Alegre went hand in hand with the election of a particular political party (Partido dos Trabalhadores or Workers Party) that catered to a particular constituency of citizen activists (and therefore adopted a particular mode of participatory institutions to implement its compact) and that had strong political incentives to serve the poor.

Baiocchi et al. (2011) use a regression discontinuity empirical design to address the possible conflation of the political effects of voter support for the Partido dos Trabalhadores with the use of the institution of participatory budgeting. They compare outcomes across municipalities where the party won or lost the elections by a narrow margin. Therefore, although these municipalities are similar in voter support for the party, participatory budgeting was adopted by the municipality in only one set, where the party won narrowly. They report significant differences in the process of citizen engagement with the local government, but they do not examine the impact on outcomes of service delivery or poverty reduction.
In a comparative study of three cities in Latin America—Porto Alegre, Montevideo, and Caracas—Goldfrank (2011) reports that Porto Alegre experienced the largest and most significant rise in civic engagement, Caracas experienced neither substantial citizen participation nor improved governance, and Montevideo fell somewhere in between. Cornwall et al. (2007) argue that a combination of innovative institutional design and the presence of effective preconditions for participatory governance explain why some participatory experiments in Brazil have been more successful than others in enhancing citizen engagement.

There is no rigorous evidence on whether or how “participatory budgeting” might be exported to contexts outside Brazil or Latin America, especially those as different as Africa, and whether this would reduce government failures. Combined with the lack of evidence on whether Bangalore-style Citizen Report Cards can be effective in reducing government failures in Africa, this review shows that any future support to civil society in Africa needs to be based on a systematic process of experimentation, impact evaluation, and learning. There are too many unanswered questions for policy to proceed as though “good practice” is already known. The potential for civil society action should be tapped through a program of rigorous, data-based impact evaluation rather than simply greater advocacy for supporting civil society or media independence and transparency. This analysis indicates that this evaluation would need to involve both an analysis of causal effects on development outcomes and a careful analysis of the mechanisms of impact, particularly how these mechanisms interact with the local political and social context.

Donor-driven creation or funding of local groups may change the nature of who participates most actively in these groups and for what purpose. Arcand and Fafchamps (2012) find that local organizations in two West African rural areas tend to be biased toward relatively privileged individuals—people with more ties to village authorities in Burkina Faso and those with greater landholdings in Senegal. Donor-supported community-based organizations were less elitist in some dimensions, at least in Senegal, but they still had elite biases. In a program in rural Kenya that provided leadership training and agricultural inputs to small self-help organizations with mainly women members, Gugerty and Kremer (2008) find that membership in the groups selected for the program changed as the program was rolled out, with greater representation of the more educated and those more likely to have formal-sector income. Group membership and leadership moved into the hands of younger and better educated women. At the same time, perhaps surprisingly, the program showed unimpressive results in terms of productivity gains.

In our review, we found no systematic or rigorous evidence on the organization and incentives of national civil society organizations. This is an obvious lacuna given recent policy directions for donor support to such organizations.
Emerging Efforts that Need Evaluation

There are some important recent examples of civil society interventions that may have an impact by tapping the potential from greater political contestation and citizen participation. These must be evaluated to assess whether and how they might overcome government failures. One such effort, called Uwezo,\textsuperscript{16} measures education service delivery and learning outcomes on a large scale, enabling performance measures to be computed at disaggregated jurisdictional levels, such as districts. Such disaggregated measurement might enable the attribution of performance in delivering education services to specific government agents and politicians, thereby strengthening political incentives.

The Uwezo effort was partly inspired by a similar effort in India, called the Annual Status of Education Report, which was run by the civil society organization Pratham.\textsuperscript{17} Although media stories suggest that such initiatives contribute to strengthening political accountability for education outcomes, there is no evidence of how they change politically motivated distortions in education allocations. In the case of India, eight years of Annual Status of Education Reports have played an important role in raising public awareness of education quality, but there has been no overall increase in test scores for rural children in the period to 2011 (ASER 2012). There remains substantial untapped potential for examining whether such disaggregated information can generate sufficient political pressure to improve performance, such as through yardstick competition across jurisdictions (Khemani 2007).

The conventional wisdom on the drivers of the recent Arab Spring indicates that collective citizen action at such a scale was facilitated by social media and ICT. Indeed, efforts to develop platforms for “crowd-sourced” information have their roots in Africa (specifically, the platform of Ushahidi\textsuperscript{18}), with much enthusiasm that these technological advances can have far-reaching accountability effects by spurring collective action among citizens. Again, however, there is little rigorous research on whether or how ICT enables greater public accountability. Most of the work on accountability and transparency from ICT remains at the level of assuming that greater use of and investment in such technology is both necessary and sufficient. Little or no work has investigated what types of accountability messages or interventions delivered through ICT achieve or fail to achieve development effectiveness. There is substantial scope for pursuing such a research program, which may have especially large returns in terms of reaching the pivotal “youth” group.

Conclusion: Implications for External Donor Support to Civil Society

This paper has evaluated whether civil society engagement in Africa can overcome government failures in facilitating growth and development. The evidence from

\textit{Devarajan et al.}
within and outside Africa suggests that there is little in the way of “best” or “good” practices. At the same time, there is growing political contestation and citizen participation in Africa, suggesting substantial scope and potential for civil society action, which could be explored through a program of rigorous learning by doing. That the evidence gathered within Africa is sometimes at odds with that from outside the continent reinforces the point that interventions should be structured so that there is maximal learning and the possibility of mid-course corrections.

Should external donors support civil society to address accountability problems in African states? The analysis of this paper indicates that such a policy, despite significant potential, is fraught with both difficulties and uncertainties—difficulties because of the intrinsic problem of an external actor seeking to change a sociopolitical system and uncertainties because of the weak information on what does and does not work. There is a prima facie case for greater action, but it is important that this be both organic (building on local forces of change in political and civil society institutions) and experimental (structured for careful monitoring and assessment of how interventions work in practice in relation to their political and social context).

Since the mid-1990s, the aid community has emphasized accountability, with activities supporting “governance” and community-driven development as well as a longer tradition of support to institutional development. The history with respect to actual results is cautionary. For example, an evaluation of public sector reforms by the World Bank’s independent evaluation group suggests mixed results at best (World Bank 2008). Although there have been some highly celebrated community-driven development programs (e.g., the self-help group movement in India), actual evidence of major change is weak, as indicated in the review of evidence above and other more extensive reviews (Mansuri and Rao 2013). In this concluding section, we extend our hypothesis that ignoring the underlying political economy drivers of accountability may have been a major factor in cases with results that were less than satisfactory to the case of aid. This is speculative, but we believe it is important for its potentially powerful implications for external aid strategies.

Aid that supports accountability improvements can get lost within a clientelistic or predatory state system, whether the aid attempts to improve public administration against the grain of underlying incentives and organizational culture or supports small local islands of client power. Furthermore, aid can exacerbate existing distortions within the system, especially when the aid is large in scale (as it is in many African countries). For instance, aid for civil society can become a basis for new breeds of rent-seeking NGOs, complete with mutually agreeable new narratives, just as African counterparts in the 1980s and 1990s learned the language of structural adjustment even while the aid was going into a highly clientelistic system.

The analysis of this paper therefore presents a difficult challenge: it develops the case for systemic change while suggesting the difficulty of external aid initiatives in doing this successfully (or at least not making things worse). This challenge is
further highlighted by the lack of clear evidence on what works. To resolve this challenge, we suggest some principles that external donors can apply to ensure that their interventions to support civil society and strengthen accountability will lead to improvements in development outcomes.

In general, aid should not be focused on “money.” This may be counterproductive, including creating incentives for rent-seeking NGOs that become adept at playing the latest game in the aid business. Rather, external partners can provide technical assistance in designing locally grown interventions, and they can play a role in financing information gathering by local NGOs (e.g., Pratham and Annual Status of Education Reports).

The most valuable area for external donors is likely to be support for a domestic process of innovation and learning through experimentation. Such a learning process can be costly—for state actors (who often have incentives to work within existing procedures and practices), for civil society actors, and for researchers. An important question concerns what techniques to use. Certainly, carefully structured experimentation is desirable; randomized control trials are one particularly powerful instrument for those interventions that are amenable to a design around treatment and control. However, the analysis suggests that it is also important to link these trials to analysis of the political and social context—the history of local political behavior, the nature of social networks and group mobilization, and so on.19 Some interventions will not be amenable to experimental methods, such as generalized legal changes in school accountability and in the responsibilities of mining investors to communities. Nevertheless, careful economic, social, and political analysis of change in different contexts can be undertaken.

Can aid ever lead to transformational changes in accountability relations? It almost certainly cannot, if designs are hatched and brought in from outside. However, aid can potentially provide a supporting role if it is aligned with the flow of internal initiatives, is consistent with domestic political strategy, and supports greater accountability at the margins of major projects. This paper has made the case that overcoming government failure requires addressing the underlying political incentives in the system. An aspiration to strengthen civil society to affect those political incentives is admirable. However, for donors, this aspiration needs to be blended with humility regarding the limits and unintended consequences of external action and a central focus on helping domestic actors learn by doing.

Notes

Shantayanan Devarajan is Chief Economist of the Middle East and North Africa Region of the World Bank. Stuti Khemani is Senior Economist in the Development Research Group of the World Bank. Michael Walton is Lecturer in Public Policy at the Harvard Kennedy School of Government and Senior
Visiting Fellow at the Centre for Policy Research, New Delhi. We are grateful to Avnish Gungadurdoss for excellent research assistance and the preparation of the literature review that is appended to this paper. Thanks to Maria Amelina, Ghazia Aslam, Kathy Bain, Punam Chuhan-Pole, Asli Gurkan, Hélène Grandvoisinnet, Brian Levy, and anonymous referees for their helpful comments. Corresponding author: Stuti Khemani, Development Research Group, The World Bank, 1818 H Street NW, Washington DC 20433, skhemani@worldbank.org.

1. Disclaimer: The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors and do not necessarily represent the views of the World Bank, its Executive Directors, or the countries they represent.

2. It is important to note that these are not problems that can be solved by rolling back the state. For a wide range of goods and policies, the state has an important role to play in facilitating markets, growth, poverty reduction, and human development. The concern here is how to make the state more accountable for better development policies in domains in which state action is desirable.

3. Many political scientists of Africa use the term “neo-patrimonial” to describe this particular mix.

4. See Polity IV (2011) for a description of the project and the data used here.

5. Eritrea and Somalia are not included in the figure because data were not available for the earlier year.


7. This was quite visible during field visits to Tanzanian villages by one of the authors (Khemani), with posters listing village groups and their membership plastered over the walls of the village government office.

8. A review of civil society theories is outside the scope of this short paper. Habermas (1984) is a seminal contribution; Baiocchi et al. (2011) Chapter 1, provides a useful interpretive survey.

9. This section makes use of selected studies drawn from more comprehensive reviews as well as an academic literature review, undertaken specifically for this paper by Avnish Gungadurdoss of the Harvard Kennedy School, that focused on evidence generated through randomized control trials and other econometric methods for the rigorous identification of impact. This literature review is appended to the working paper available at http://www.wds.worldbank.org/external/default/WDSContentServer/IW3P/IB/2011/07/25/000158349_20110725162228/Rendered/PDF/WPS5733.pdf.

10. While this and some of the other examples in the paper refer to contexts outside Africa, they are nevertheless relevant to our lessons and recommendations about intervention design and impact evaluation in Africa. Care should be taken, however, in expecting any one study to have the same impact in another context, be it within or across continents.

11. The authors use a measure of ethnolinguistic fractionalization and a household asset-based proxy for incomes and compare impacts for the 25th and 75th percentile of these measures.

12. We do not review the extensive literature on “community driven development” for this reason: most of the evidence consists of evaluating the impact of state devolution to local institutions created by the state, with local politicians and state-appointed bureaucrats at the helm (as in the case of local governments).

13. Despite this, other work, including work from the same author (Altschuler and Corrales 2012), suggests that parents who participated on school committees in Honduras and Guatemala developed skills to participate in other group activities and to join other civic organizations. Such spillover effects suggest that participatory interventions may have impacts outside the immediate domain of a project.

14. Even when governments (perhaps driven by donor or NGO pressure) elect to undertake projects that engage civil society, they may continue to let their bureaucrats and providers get away with high salaries and low effort in servicing citizens. This situation relates to a general concern with community participation initiatives, as expressed in Banerjee and Duflo (2008), that the responsibility for managing projects and services is being placed on largely poor communities and can be viewed as a
tax on their time and efforts, especially in comparison with other ways of making government agents work harder and better for the poor.

15. The Olken study was conducted within areas that had already experienced substantial efforts to improve local government and community involvement under the Kecamatan Development Program. There is also evidence from Brazil of the impact of state-led audit reforms in reducing local corruption (Ferraz and Finan 2008). However, there is no clear description available of the role of civil society for this reform.

19. Joshi and Houtzager (2012) make a similar methodological point in favor of analysis on local political processes, although we have argued that such an analysis can be linked to experimentation with interventions to identify specific features of what works, under what conditions, in effecting change.

References


Business training programs are a popular policy option to improve the performance of enterprises around the world, and the number of rigorous impact evaluations of these programs is growing. A critical review reveals that many evaluations suffer from small sample sizes, measure impacts only within a year of training, and experience problems with survey attrition and measurement that limit the conclusions one can draw. Over these short time horizons, there are relatively modest effects of training on the survivorship of existing firms. However, there is stronger evidence that training programs help prospective owners launch new businesses more quickly. Most studies find that existing firm owners implement some of the practices taught in training, but the magnitudes of the improvement to practices is often modest. Few studies find significant impacts on profits or sales, although some studies with greater statistical power have done so. There is little evidence to guide policymakers regarding whether any identified effects are due to trained firms drawing sales from competing businesses rather than through productivity improvements or to guide the development of the provision of training at market prices. We conclude by summarizing some directions and key questions for future studies.

JEL codes: O12, J16, L26, M53

Walk into a typical micro or small business in a developing country and spend a few minutes talking with the owner, and it often becomes clear that owners are not implementing many of the business practices that are standard in most small businesses in developed countries. Formal records are not kept, and household and business
finances are combined. Marketing efforts are sporadic and rudimentary. Some inventory sits on shelves for years at a time, whereas more popular items are frequently out of stock. Few owners have financial targets or goals that they regularly monitor and act to achieve. The picture is not much better in some medium and large firms: few firms use quality control systems, reward workers with performance-based pay, or adopt many other practices that are typical of well-managed firms in developed countries.

It is small wonder, then, that business training is one of the most common forms of active support provided to small firms around the world. There are a number of programs offered by governments, microfinance organizations, and NGOs in many countries around the world. Perhaps the mostly widely implemented training program is the International Labor Organization’s Start and Improve Your Business program. Established in 1977, the program claims more than 4.5 million trainees with implementation in more than 100 countries.¹ Other widely used programs include the GTZ/CEFE program, the UNCTAD/EMPRETEC program, business plan competitions and training run by Technoserve, content for microfinance clients developed by Freedom from Hunger, and the IFC’s Business Edge and SME Toolkit programs.

Until recently, however, there has been very little rigorous evidence on the impacts of these programs. Overviews of evidence from mostly nonexperimental evaluations of programs that focus on training for the unemployed in developed countries (Dar and Tzannatos 1999) and developing and transition countries (Betcherman et al. 2004) have found the existing evidence to be mixed, at best. A 2009 overview of impact evaluations in finance and private sector development found very little work on business training (McKenzie 2010). The last three years have seen a rapid increase in attention to the idea that “managerial capital” or poor management is a constraint to production in developing countries (Bruhn et al. 2010; Bloom and Van Reenen 2010) as well as the emergence of a number of impact evaluations of business training programs. This paper provides a critical overview of lessons from these evaluations for both policy and the next generation of research.

We use a variety of methods to identify all published studies and recent working papers that examine the impacts of business training in developing countries. These include an Econlit search for published studies, Google Scholar searches of papers that cite these published studies or other working papers, our contacts with scholars working in this field, input from recent training program inventory exercises, and knowledge of papers presented in recent seminars or conferences. We restrict our attention to papers with a clear impact evaluation design that address the selection of both observable and unobservable characteristics of business owners and that focus on enterprise management rather than solely on technical or vocational training.
We begin by assessing the comparability of these programs in terms of their course content and participants. We find considerable variation across studies in terms of the participants and the length and content of the training provided, although a number of core topics are covered in most training sessions. Next, we discuss a number of challenges faced by researchers when measuring impact. Critically, most of the existing studies measure impacts on relatively small samples of very heterogeneous firms. In addition, many existing studies only consider impacts within a year of training, a period that is too short to detect some changes. Many studies also experience problems with attrition, selective survival and start-up, and nonresponses for sensitive outcomes such as profits and revenues. A final concern is that training may change the measurement of outcomes even if it does not change the outcomes themselves. We discuss several studies’ attempts to show that their results are robust to reporting issues.

With these issues in mind, we assess what we have learned about the impacts of different programs on business survivorship and start-up, business practices, and profitability and enterprise growth. Among the minority of studies that have examined the effects of these programs on the survivorship of existing businesses, there is some weak evidence for a positive effect for male-owned businesses. However, for female-owned businesses, training is found to have either no effect or a slightly negative effect on survivorship. Stronger results have been found with respect to the impacts of training programs on new business start-ups. All of the studied training programs that include content specifically intended to help people start new businesses have found that training helps in starting firms, although there is some evidence that training merely hastens the entry of firms that would enter anyway and potentially changes the selection of which firms enter.

Almost all training programs find that treated firms implement some of the business practices taught in the training. However, the magnitude of the impact is small in many cases; a typical change is 0.1 or 0.2 standard deviations, or 5 to 10 percentage points. The combination of relatively small changes in business practices and low statistical power means that few studies find effects of training on sales or profitability, although a few studies find some positive short-term effects. Studies of microfinance clients find some evidence that training changes the rates of client retention and the characteristics of loan applicants. Finally, the three studies that examine the impact of individualized consulting provided to larger firms find evidence that consulting services can improve the performance of firms, including those with multiple plants and more than 200 workers.

Before concluding, we discuss several important issues for which existing studies provide very little evidence but which are crucial for the development of policy recommendations. These issues include whether gains from training are long lasting and whether these gains result from competing away sales from untrained firms or through other channels. We also discuss the need to address the heterogeneity of
training content and participants and to identify the market failures that may prevent firms from investing in training that may be beneficial. We conclude with recommendations for future work in this area.

What Does a Typical Business Training Program Involve?

Attempts to measure the impact of “business training” face multiple challenges that complicate comparisons across studies. The first challenge is that business training varies in what is offered and how it is offered across different locations and organizations. These differences in content are likely to be important, and they induce much more variation into the treatment of business training than exists in other firm interventions, such as access to capital through credit or grants. A second challenge (common to most evaluations) is that the impact of training is likely to differ depending on who receives the training. Thus, even if we compare the same training content in different locations, differences in the characteristics of the individuals receiving the training may result in different measured impacts. Therefore, it is important to carefully examine who participates and what is offered before making comparisons among studies.

Who Participates in Business Training Experiments?

Table 1 summarizes the key characteristics of the participants in recent business training evaluation studies. Classroom-based training offered by microfinance organizations or banks to their clients is the most common modality among these studies. This approach is particularly common for training offered to female microenterprise owners because the majority of microfinance clients are women. A second strategy is to offer training to firms in a particular industry or industrial cluster (Mano et al. 2011; Sonobe et al. 2011). A third strategy is for individuals to apply to participate in training as part of a competition, as Technoserve does (Klinger and Schündeln 2011), to be screened for interest in participating (Valdivia 2012), or for students to apply to participate in an entrepreneurship course (Premand et al. 2012). All of these approaches result in a selected sample of firms, which may differ from the general population, making it difficult to generalize their findings to an average firm. A final approach, used only by de Mel et al. (2012) and Calderon et al. (2012), is to draw a representative sample of the microenterprise population of interest and then offer the training to a random subsample of this population.

Most evaluations focus on existing businesses. Exceptions include studies in which many of the microfinance clients are borrowing or saving for household purposes but do not necessarily have an enterprise (Field et al. 2010) and studies based
Table 1. Who Are the Participants in Business Training Evaluations?

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Existing Businesses?</th>
<th>All microfinance/ bank clients?</th>
<th>Rural or Urban</th>
<th>Business Sector</th>
<th>Selected on interest in training?</th>
<th>Mean Age</th>
<th>% Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berge et al. (2011)</td>
<td>Tanzania</td>
<td>Existing</td>
<td>Yes</td>
<td>Urban</td>
<td>Many</td>
<td>No</td>
<td>38</td>
<td>65</td>
</tr>
<tr>
<td>Bruhn and Zia (2012)</td>
<td>Bosnia-Herzegovina</td>
<td>67% existing</td>
<td>Yes</td>
<td>Urban</td>
<td>Many</td>
<td>Yes</td>
<td>28</td>
<td>35</td>
</tr>
<tr>
<td>Calderon et al. (2012)</td>
<td>Mexico</td>
<td>Existing</td>
<td>No</td>
<td>Rural</td>
<td>Many</td>
<td>No</td>
<td>46</td>
<td>100</td>
</tr>
<tr>
<td>De Mel et al. (2012)</td>
<td>Sri Lanka</td>
<td>50% existing</td>
<td>No</td>
<td>Urban</td>
<td>Many</td>
<td>No</td>
<td>34-36</td>
<td>100</td>
</tr>
<tr>
<td>Drexler et al. (2012)</td>
<td>Dominican Republic</td>
<td>Existing (a)</td>
<td>Yes</td>
<td>Urban</td>
<td>Many</td>
<td>No</td>
<td>40</td>
<td>90</td>
</tr>
<tr>
<td>Field et al. (2010)</td>
<td>India</td>
<td>24% existing</td>
<td>Yes</td>
<td>Urban</td>
<td>Many</td>
<td>No</td>
<td>32.4</td>
<td>100</td>
</tr>
<tr>
<td>Giné and Mansuri (2011)</td>
<td>Pakistan</td>
<td>61% existing</td>
<td>Yes</td>
<td>Rural</td>
<td>Many</td>
<td>No</td>
<td>37.6</td>
<td>49</td>
</tr>
<tr>
<td>Glaub et al. (2012)</td>
<td>Uganda</td>
<td>Existing</td>
<td>No</td>
<td>Urban</td>
<td>Many</td>
<td>Yes</td>
<td>39</td>
<td>49</td>
</tr>
<tr>
<td>Karlan and Valdivia (2011)</td>
<td>Peru</td>
<td>Existing</td>
<td>Yes</td>
<td>Both</td>
<td>Many</td>
<td>No</td>
<td>n.r.</td>
<td>96</td>
</tr>
<tr>
<td>Klinger and Schündeln (2011)</td>
<td>El Salvador, Guatemala, Nicaragua</td>
<td>39% existing</td>
<td>No</td>
<td>n.r.</td>
<td>Many</td>
<td>Yes</td>
<td>36</td>
<td>28</td>
</tr>
<tr>
<td>Mano et al. (2012)</td>
<td>Ghana</td>
<td>Existing</td>
<td>No</td>
<td>Urban</td>
<td>Metalwork</td>
<td>No</td>
<td>45</td>
<td>0</td>
</tr>
<tr>
<td>Premand et al. (2012)</td>
<td>Tunisia</td>
<td>No</td>
<td>No</td>
<td>Urban</td>
<td>Many</td>
<td>Yes</td>
<td>23</td>
<td>67</td>
</tr>
<tr>
<td>Sonobe et al. (2011)</td>
<td>Tanzania</td>
<td>Existing</td>
<td>No</td>
<td>Urban</td>
<td>Garments</td>
<td>No</td>
<td>45</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>Ethiopia</td>
<td>Existing</td>
<td>No</td>
<td>Urban</td>
<td>Metalwork</td>
<td>No</td>
<td>44</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Vietnam</td>
<td>Existing</td>
<td>No</td>
<td>Urban</td>
<td>Rolled Steel</td>
<td>No</td>
<td>40</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Vietnam</td>
<td>Existing</td>
<td>No</td>
<td>Urban</td>
<td>Knitwear</td>
<td>No</td>
<td>41</td>
<td>66</td>
</tr>
<tr>
<td>Valdivia (2012)</td>
<td>Peru</td>
<td>Existing</td>
<td>No</td>
<td>Urban</td>
<td>Many</td>
<td>Yes</td>
<td>43</td>
<td>100</td>
</tr>
</tbody>
</table>

Note: n.r. denotes not reported.

(a) 78 percent of sample is existing businesses, and study does not look at business outcomes for those who were not existing at baseline.
on competitions or training of new businesses (Klinger and Schündeln 2011; Premand et al. 2012; de Mel et al. 2012). The majority of the evaluations to date have focused on urban clients, which likely reflects the greater density of businesses and training providers in urban areas.

The average age of a participant in a typical study is 35 to 45 years, although two studies focus on young entrepreneurs (Bruhn and Zia 2012; Premand et al. 2012). Some studies focus entirely on female business owners, and others focus on male owners. Relatively few studies have sufficient numbers of both genders to compare impacts separately. Finally, there is substantial heterogeneity in the education levels of participants, with averages as low as 2.5 years of schooling for females and 5.7 years for males in the study of rural Pakistan by Giné and Mansuri (2011) and as high as university level in the study by Premand et al. (2012).

Table 2 shows the degree of heterogeneity in firm size at baseline among studies that include existing firms. At the low end are subsistence firms run by women in Giné and Mansuri (2011) and de Mel et al. (2012), where 95 percent of the firms

<table>
<thead>
<tr>
<th>Study</th>
<th>% with zero employees</th>
<th>Mean Employees</th>
<th>Monthly Profits (USD)</th>
<th>Monthly Revenues (USD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berge et al. (2011)</td>
<td>n.r.</td>
<td>1.08</td>
<td>480</td>
<td>2102</td>
</tr>
<tr>
<td>Males</td>
<td>n.r.</td>
<td>1.18</td>
<td>528</td>
<td>2586</td>
</tr>
<tr>
<td>Females</td>
<td>n.r.</td>
<td>1.03</td>
<td>455</td>
<td>1847</td>
</tr>
<tr>
<td>Bruhn and Zia (2012)</td>
<td>n.r.</td>
<td>2.08</td>
<td>700</td>
<td>n.r.</td>
</tr>
<tr>
<td>Calderon et al. (2012)</td>
<td>60</td>
<td>1.6</td>
<td>121</td>
<td>398</td>
</tr>
<tr>
<td>De Mel et al. (2012)</td>
<td>95</td>
<td>0.06</td>
<td>35</td>
<td>109</td>
</tr>
<tr>
<td>Drexler et al. (2012)</td>
<td>60</td>
<td>n.r.</td>
<td>n.r.</td>
<td>747</td>
</tr>
<tr>
<td>Giné and Mansuri (2011)</td>
<td>90</td>
<td>2.43</td>
<td>n.r.</td>
<td>n.r.</td>
</tr>
<tr>
<td>Males</td>
<td>86</td>
<td>2.51</td>
<td>n.r.</td>
<td>380</td>
</tr>
<tr>
<td>Females</td>
<td>95</td>
<td>2.34</td>
<td>n.r.</td>
<td>80</td>
</tr>
<tr>
<td>Glaub et al. (2012)</td>
<td>n.r.</td>
<td>1.5</td>
<td>n.r.</td>
<td>100</td>
</tr>
<tr>
<td>Karlan and Valdivia (2011)</td>
<td>n.r.</td>
<td>0.22</td>
<td>-165</td>
<td>534</td>
</tr>
<tr>
<td>Klinger and Schündeln (2011)</td>
<td>n.r.</td>
<td>8</td>
<td>n.r.</td>
<td>6916</td>
</tr>
<tr>
<td>Mano et al. (2012)</td>
<td>n.r.</td>
<td>n.r.</td>
<td>2200</td>
<td>4717</td>
</tr>
<tr>
<td>Sonobe et al. (2011)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanzania</td>
<td>n.r.</td>
<td>5</td>
<td>530</td>
<td>866</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>n.r.</td>
<td>33</td>
<td>19599</td>
<td>142311</td>
</tr>
<tr>
<td>Vietnam - Steel</td>
<td>n.r.</td>
<td>17</td>
<td>2627</td>
<td>105787</td>
</tr>
<tr>
<td>Vietnam - Knitwear</td>
<td>n.r.</td>
<td>20</td>
<td>-888</td>
<td>7055</td>
</tr>
<tr>
<td>Valdivia (2012)</td>
<td>n.r.</td>
<td>0.23</td>
<td>n.r.</td>
<td>740</td>
</tr>
</tbody>
</table>

Note: n.r. denotes not reported.
have no paid employees, average monthly revenues are only $80–100 at market exchange rates, and profits are approximately $1 per day. Most of the rest of the studies focus on microenterprises, albeit ones with slightly larger revenues and potentially one or two employees. The main exceptions are the firms chosen from industrial clusters (Mano et al. 2011 and Sonobe et al. 2012), in which the firms are SMEs with five to 50 workers and monthly revenues of $5,000 or more (and in some cases, more than $100,000).

Training Delivery and Costs

All of the training courses reviewed here are classroom-based courses delivered to groups of individuals, although several of the programs provide additional one-on-one follow-up training, which we will discuss later. Table 3 provides key characteristics of the training delivery in the different studies. A first point is that many of the studies test content that is modified or developed specifically for the study of interest rather than content that has been taught for years. This situation may be significant if it takes time to adapt particular content to a local context or for instructors to become familiar with new material.

The length of the training course also varies substantially across studies. The shortest courses are two days or two half-days (Bruhn and Zia 2012; Field et al. 2010), whereas other courses are full time and last one week or more (de Mel et al. 2012; Sonobe et al. 2011). In most cases, the training is concentrated in a relatively short period, but in some of the cases, especially where training takes place in microfinance group meetings, it is spread over many months in blocks as short as half an hour (Karlan and Valdivia 2011). Longer full-time courses allow more content to be taught, but they are more costly and require business owners to be away from their businesses for longer.

In all of these experiments, training is offered for free. In addition, some studies have provided small supplements for travel or food or have offered the prospect of grants as an additional incentive. The training costs per person range from as little as $21 in Drexler et al. (2011), where training was conducted by local instructors once per week over five or six weeks in local schools, to more than $400 per firm in Sonobe et al. (2011), where instructor costs and venue rental costs per person for 15 days were relatively high. One argument for subsidizing costs is that many business owners have little perception of how badly managed their firms are. To these owners, training is a new and unproven concept with uncertain payoffs. Even those who are not liquidity constrained may be reluctant to pay, and training providers may find it costly and difficult to credibly signal quality. Two studies find evidence to support the idea that individuals who are the least interested potentially have the most to gain from training (Bjorvatn and Tungodden 2010; Karlan and Valdivia
## Table 3. Key Characteristics of Training Delivery

<table>
<thead>
<tr>
<th>Study</th>
<th>Training Provider</th>
<th>Training content</th>
<th>Course Length (hours)</th>
<th>Participant Cost (USD)</th>
<th>Actual Cost (USD)</th>
<th>Attendance Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berge et al. (2011)</td>
<td>Training professionals</td>
<td>New</td>
<td>15.75</td>
<td>0</td>
<td>$70</td>
<td>83%</td>
</tr>
<tr>
<td>Bruhn and Zia (2012)</td>
<td>Training organization</td>
<td>New</td>
<td>6</td>
<td>0</td>
<td>$245</td>
<td>39%</td>
</tr>
<tr>
<td>Calderon et al. (2012)</td>
<td>Professors &amp; Students</td>
<td>New</td>
<td>48</td>
<td>0</td>
<td>n.r.</td>
<td>65%</td>
</tr>
<tr>
<td>De Mel et al. (2012)</td>
<td>Training organization</td>
<td>Established (ILO)</td>
<td>49-63</td>
<td>0</td>
<td>$126-140</td>
<td>70-71%</td>
</tr>
<tr>
<td>Drexler et al. (2012)</td>
<td>“Standard”</td>
<td>Local instructors</td>
<td>New</td>
<td>18</td>
<td>0 or $6</td>
<td>$21</td>
</tr>
<tr>
<td></td>
<td>“Rule-of-thumb”</td>
<td>Local instructors</td>
<td>New</td>
<td>15</td>
<td>0 or $6</td>
<td>$21</td>
</tr>
<tr>
<td>Field et al. (2010)</td>
<td>Microfinance credit officers</td>
<td>New (a)</td>
<td>2 days</td>
<td>0</td>
<td>$3</td>
<td>71%</td>
</tr>
<tr>
<td>Giné and Mansuri (2011)</td>
<td>Microfinance credit officers</td>
<td>New (b)</td>
<td>46</td>
<td>0</td>
<td>n.r.</td>
<td>50%</td>
</tr>
<tr>
<td>Glaub et al. (2012)</td>
<td>Professor</td>
<td>New</td>
<td>3 days</td>
<td>0</td>
<td>$60</td>
<td>84%</td>
</tr>
<tr>
<td>Klinger and Schündeln (2011)</td>
<td>Microfinance credit officers</td>
<td>Established (FFH)</td>
<td>8.5-22 (c)</td>
<td>0</td>
<td>n.r.</td>
<td>76-88%</td>
</tr>
<tr>
<td>Mano et al. (2012)</td>
<td>Local instructors</td>
<td>New (d)</td>
<td>37.5</td>
<td>0</td>
<td>$740</td>
<td>87%</td>
</tr>
<tr>
<td>Premand et al. (2012)</td>
<td>Govt. office staff</td>
<td>New</td>
<td>20 days +</td>
<td>0</td>
<td>n.r.</td>
<td>59-67%</td>
</tr>
<tr>
<td>Sonobe et al. (2011)</td>
<td>Training professionals</td>
<td>New (d)</td>
<td>20 days</td>
<td>0</td>
<td>&gt;$400</td>
<td>92%</td>
</tr>
<tr>
<td>Tanzania</td>
<td>Training professionals</td>
<td>New (d)</td>
<td>20 days</td>
<td>0</td>
<td>n.r.</td>
<td>n.r.</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>Training professionals</td>
<td>New (d)</td>
<td>15 days</td>
<td>0</td>
<td>75%</td>
<td></td>
</tr>
<tr>
<td>Vietnam - Steel</td>
<td>Training professionals</td>
<td>New (d)</td>
<td>15 days</td>
<td>0</td>
<td>39%</td>
<td></td>
</tr>
<tr>
<td>Vietnam - Knitwear</td>
<td>Training professionals</td>
<td>New (d)</td>
<td>15 days</td>
<td>0</td>
<td>59%</td>
<td></td>
</tr>
<tr>
<td>Valdivia (2012)</td>
<td>Training professionals</td>
<td>New</td>
<td>108 (e)</td>
<td>0</td>
<td>$337 (f)</td>
<td>51%</td>
</tr>
</tbody>
</table>

Note: FFH denotes Freedom from Hunger; ILO denotes the International Labor Organization.

(a) Shortened version of existing program + new content on aspirations added.
(b) Adapted from ILO’s Know About Business modules.
(c) Training sessions were each 30 minutes to 1 hours, and up to 22 sessions occurred, but only half had done 17 sessions over 24 months.
(d) Based in part on ILO content + Japanese Kaizen content.
(e) Although only 42 percent of those attending completed at least 20/36 sessions, and only 28 percent attended 30 sessions or more.
(f) The basic training cost $337, while the technical assistance plus basic training cost $674.
2011). We will return to a discussion of market failures and subsidies later in the paper.

Although training is offered for free, the average participation rate across the different studies for individuals who are offered training is only about 65 percent. Low take-up rates make it difficult to measure impacts; decreasing the take-up rate from 100 percent to 65 percent increases the required sample size by 2.4 times. One would expect take-up rates to be highest when training occurs in the context of regular group meetings organized by microfinance organizations, but even in the “mandatory” treatment of Karlan and Valdivia (2011), attendance rates are only 88 percent. Screening for initial interest in training does not guarantee high take-up rates either. Bruhn and Zia (2012) and Valdivia (2012) focus on samples that had initially expressed interest in attending a training course, but they still only obtain attendance rates of 39 percent and 51 percent, respectively. In most short courses, there is very little drop out conditional on attending the first session of the course, but longer courses experience more drop out over time.

**Training Content**

Table 4 summarizes the key topics taught in the different courses. All of the studies focus on general business skills that should be broadly applicable to most businesses rather than technical knowledge or sector-specific content. However, there is significant variation in the depth and breadth of topics. The most common set of topics focuses on maintaining business records and encouraging small business owners to separate household and business finances. Many courses, especially those targeted to potential rather than existing business owners, focus on generating a product idea and the steps needed to take the product to market. A core set of topics for attempting to grow existing businesses includes marketing, pricing and costing, inventory management, customer service, and financial planning. Because few microenterprises have employees, employee management is not a significant part of most courses. Courses that focus on larger firms include content on quality management, lean production, or Kaizen and 5S techniques[^3] for continuous production improvement. Finally, in addition to targeting improvements in business practices, some courses attempt to change entrepreneurial attitudes or aspirations. The amount of time devoted to attitudes has been relatively low in the courses studied by economists, but Glaub and Frese (2011) review a number of nonexperimental studies of training programs in developing countries that focus on strengthening psychological factors. Glaub et al. (2012) provide an example of a three-day course focused on personal initiative training, a psychological intervention aimed at making business owners more proactive and self-starting with respect to new ideas and opportunities and more persistent in overcoming barriers.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Berge et al. (2011)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Bruhn and Zia (2012)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Calderon et al. (2012)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>De Mel et al. (2012)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Drexler et al. (2012)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X X X X X X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>&quot;Standard&quot;</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X X X X X X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>&quot;Rule-of-thumb&quot;</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X X X X X X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Field et al. (2010)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X X X X X X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Gine and Mansuri (2011)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X X X X X X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Glaub et al. (2012)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X X X X X X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Karlan and Valdivia (2011)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X X X X X X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Continued
Table 4. Continued

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Klinger and Schündeln (2011)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mano et al. (2012)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Premand et al. (2012)</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sonobe et al. (2011)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanzania</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Valdivia (2012)</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Note: Based on training descriptions provided in research studies.
The different types of content may affect business performance in different ways. Simple accounting practices and financial literacy training may give business owners a better understanding of the profitability of their business but may have little immediate effect on sales or profit levels. However, in the longer term, better accounting practices may enable owners to reinvest more in their firms because of higher savings or to put more effort on product lines that are more profitable. In contrast, some other practices may show impacts more quickly. For example, better marketing and customer service may directly increase sales, whereas costing and quality control practices may lead to reduced costs and increased profits. The development of a new product idea may have rapid and long-lasting benefits even if no other additional practices are introduced. Changes in entrepreneurial attitudes may affect how hard the owner works and the way the owner thinks about various business decisions. However, because all of the available training experiments contain a mixture of different content, existing studies are unable to determine which components of training are most important.

Challenges in Measuring Impact

Impact evaluations that measure the effects of business training programs on business performance rely primarily on survey data to measure outcomes. To obtain credible and useful estimates, studies must have sufficient statistical power, measure impacts over an appropriate time horizon, address survey attrition and the selective survival and start-up of firms, and address the possibility that training changes how firms report business outcomes even if it does not change those outcomes. We discuss each of these challenges and assess how well existing studies have met them.

Power

The power of a statistical test is the probability that it will reject a null hypothesis given that the null hypothesis is false. A starting point for most business training evaluations is to test the null hypothesis that the intervention had no effect, so the power of the experiment is a measure of the ability to detect an effect of training if such an effect does exist. The key determinants of the power of a study are the size of the sample, the amount of heterogeneity in the sample (the more diverse the set of firms, the more difficult it is to measure change in them), whether the intervention occurs at an individual or group level (power is lower for a given sample size when treatments are allocated at the group level), and the size of the treatment effect. Low take-up rates dilute the treatment effect, reducing power.

Table 5 compares studies in terms of these components of power. A typical study involves approximately 200 to 400 individuals or groups in each of the treatment
Table 5. Power of Studies to Detect Increases in Profits or Sales

<table>
<thead>
<tr>
<th>Study</th>
<th>Group or Individual Randomization</th>
<th>Sample Sizes in Treatment (T) and Control (C) Groups</th>
<th>C.V. Profits</th>
<th>C.V. Revenues</th>
<th>Attendance Rate</th>
<th>25% in Profits</th>
<th>50% in Profits</th>
<th>25% in Revenues</th>
<th>50% in Revenues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berge et al. (2011)</td>
<td>Group</td>
<td>119 (T), 116 (C) groups (a)</td>
<td>0.80</td>
<td>1.47</td>
<td>83%</td>
<td>0.631-0.842</td>
<td>0.996-1.000</td>
<td>0.239-0.365</td>
<td>0.705-0.897</td>
</tr>
<tr>
<td>Bruhn and Zia (2012)</td>
<td>Individual</td>
<td>297 (T), 148 (C)</td>
<td>2.69</td>
<td>n.a.</td>
<td>39%</td>
<td>0.070</td>
<td>0.132</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Calderon et al. (2012)</td>
<td>Two-stage</td>
<td>164 (T), 711 (C)</td>
<td>1.51</td>
<td>1.53</td>
<td>65%</td>
<td>0.263 (b)</td>
<td>0.754 (b)</td>
<td>0.257 (b)</td>
<td>0.743 (b)</td>
</tr>
<tr>
<td>De Mel et al. (2012)</td>
<td>Individual</td>
<td>200 (T1), 200 (T2), 228 (C)</td>
<td>0.49</td>
<td>0.91</td>
<td>70%</td>
<td>0.990</td>
<td>1.000</td>
<td>0.632</td>
<td>0.994</td>
</tr>
<tr>
<td>Drexler et al. (2012)</td>
<td>Individual</td>
<td>402 (T1), 404 (T2), 387 (C)</td>
<td>n.a.</td>
<td>1.63</td>
<td>49%</td>
<td>n.a.</td>
<td>n.a.</td>
<td>0.231</td>
<td>0.686</td>
</tr>
<tr>
<td>Giné and Mansuri (2011)</td>
<td>Group</td>
<td>373 (T), 374 (C) groups</td>
<td>n.a.</td>
<td>n.a.</td>
<td>50%</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Glaub et al. (2012)</td>
<td>Individual</td>
<td>56 (T), 53 (C)</td>
<td>n.a.</td>
<td>n.a.</td>
<td>84%</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Karlan and Valdivia (2011)</td>
<td>Group</td>
<td>138 (T), 101 (C) groups</td>
<td>-24.96</td>
<td>2.30</td>
<td>80%</td>
<td>0.057 (b)</td>
<td>0.078 (b)</td>
<td>0.120-0.757</td>
<td>0.335-1.000</td>
</tr>
<tr>
<td>Klinger and Schündeln (2011)</td>
<td>Individual RD</td>
<td>377 (T), 278 (C)</td>
<td>n.a.</td>
<td>2.51</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>0.259 (d)</td>
<td>0.746 (d)</td>
</tr>
<tr>
<td>Mano et al. (2012)</td>
<td>Individual</td>
<td>47 (T), 66 (C) (b)</td>
<td>1.23</td>
<td>1.20</td>
<td>87%</td>
<td>0.188</td>
<td>0.571</td>
<td>0.195</td>
<td>0.592</td>
</tr>
<tr>
<td>Sonobe et al. (2011)</td>
<td>Tanzania</td>
<td>53 (T), 59 (C)</td>
<td>1.99</td>
<td>1.61</td>
<td>92%</td>
<td>0.109</td>
<td>0.292</td>
<td>0.141</td>
<td>0.414</td>
</tr>
<tr>
<td></td>
<td>Ethiopia</td>
<td>56 (T), 47 (C)</td>
<td>1.94</td>
<td>2.49</td>
<td>75%</td>
<td>0.087</td>
<td>0.204</td>
<td>0.072</td>
<td>0.142</td>
</tr>
<tr>
<td></td>
<td>Vietnam - Steel</td>
<td>110 (T), 70 (C)</td>
<td>1.59</td>
<td>0.93</td>
<td>39%</td>
<td>0.075</td>
<td>0.153</td>
<td>0.124</td>
<td>0.353</td>
</tr>
<tr>
<td></td>
<td>Vietnam - Knitwear</td>
<td>91 (T), 70 (C)</td>
<td>-8.15</td>
<td>2.34</td>
<td>59%</td>
<td>0.052</td>
<td>0.058</td>
<td>0.074</td>
<td>0.150</td>
</tr>
<tr>
<td></td>
<td>Valdivia (2012)</td>
<td>709 (T1), 709 (T2), 565 (C)</td>
<td>n.a.</td>
<td>2.29</td>
<td>51%</td>
<td>n.a.</td>
<td>n.a.</td>
<td>0.207</td>
<td>0.626</td>
</tr>
</tbody>
</table>

Notes: n.a. denotes not available, either because the study did not report this outcome, or because it didn’t report the coefficient of variation (C.V.).

Personal correspondence with authors used to obtain C.V.s from studies which only report sample means and not standard deviations.

Where range is shown, first number is power if intra-cluster correlation is one, second is power if intra-cluster correlation is zero.

(a) Numbers in control and training only groups - the study also includes groups with grants. Power calculations based on random assignment to groups, which is the working assumption of the paper, although in practice true random assignment only occurred at the branch-day of the week level, in which case power is zero.

(b) Power calculation assuming randomization was at the individual level. Actual power will be lower once group-level randomization is accounted for.

(c) Assignment first at the village level to 7 treated villages and 10 control villages, then assignment within village to treatment and control.

(d) Study does not examine revenue as an outcome, since some data is collected retrospectively. Power calculations ignore survey attrition, which would further lower power. They also assume entire sample are existing enterprises.

Attendance rate for Klinger and Schündeln (2011) assumed to be 90 percent for purpose of power calculations.

Power calculations assume one baseline and one post-treatment survey, with an autocorrelation in the outcome variable of 0.5, and ANCOVA estimation.
and control groups, although sample sizes have been smaller for studies based on specific industrial clusters (Mano et al. 2012; Sonobe et al. 2011). A useful summary statistic of the cross-sectional heterogeneity in baseline firms is the coefficient of variation of profits or revenues, which is the ratio of the standard deviation and the mean. The two studies with the lowest coefficients of variation are both studies that restrict the heterogeneity in firms eligible for the study. De Mel et al. (2009) required firms to have baseline profits below Rs 5,000 per month ($43), whereas Berge et al. (2011) restricted training to firms with loan sizes in a narrowly defined range. In contrast, most studies contain a much wider mix of firms, resulting in coefficients of variation exceeding two or more. The more heterogeneous the firms are, the more difficult it is to detect changes in their average outcomes arising from treatment.

Many funding agencies consider 80 percent to 90 percent power an appropriate target (Duflo et al. 2008), and power of 80 percent or more is the standard in medical trials (Schultz and Grimes 2005). Table 5 shows that many—indeed, most—business training experiments fall well below these levels in terms of power to detect a 25 percent or even 50 percent increase in profits or revenues. For a microenterprise earning $25 per month (about $1 per day), a 25 percent increase in profits would be $75 per year, or about 75 percent of the direct costs of a typical microenterprise training program. Therefore, a reasonable assessment of impact should have the power to measure returns at least at this level. However, in fact, none of the studies achieves 80 percent power to detect a 25 percent increase in revenues, and only de Mel et al. (2012) and possibly Berge et al. (2011) exceed 80 percent power for a 25 percent increase in profits. Valdivia (2012) demonstrates the importance of heterogeneity and take-up. Although that study has the largest sample size of any individual experiment, high heterogeneity and a low 51 percent take-up rate (requiring four times the sample size to achieve a given power compared with a 100 percent take-up rate) yield very low power.

We should also note that power is generally much higher for detecting binary outcomes, such as whether a new business is started, whether a firm applies for a loan, or whether a firm implements a particular business practice. Therefore, studies with low power to inform about the impact of training on ultimate business outcomes may still be informative about other training impacts.

Timing of Effects

The short- and long-term impacts of many policies may differ substantially, so a key challenge for impact evaluation is determining when to measure outcomes (King and Behrman 2009). For business training, one might expect firms to make some changes relatively quickly after training. However, the full impact of training may take some time. Impacts on business survival may also take time to materialize.
However, firms may begin some practices and then drop them, so surveys that measure what occurs in the business only several years after training may miss the period of experimentation. Ideally, studies should trace the trajectories of impacts, measuring both short- and long-term effects.

Table 6 provides details on the number of follow-up surveys, their timing, and their attrition rate for the different studies. The majority of studies that we review use a single follow-up survey, providing a snapshot of information on the training impact but no details on the trajectory of impacts. Eight of the 13 studies are very short-term studies that examine impacts one year or less after training. De Mel et al. (2012) find that the impacts differ in the short and medium term in their study. For example, in their study, examining impacts within the first year shows that business training for women out of the labor force led to large increases in business entry, whereas surveys 16 and 25 months after training shows that the control group had caught up in terms of business ownership rates.

**Survey Attrition and Selective Survival or Start-up**

Survey attrition is another problem that complicates inference, especially if the reasons for attrition are business failure, refusal because of disappointment with the training effects, or successful business owners moving out of the area. Attrition
rates range from as low as 5.3 percent in Field et al. (2010) and 6 percent to 8 percent in de Mel et al. (2012) to 24 percent in Karlan and Valdivia (2011) 26 percent in Calderon et al. (2012), and 28 percent in Klinger and Schündeln (2011).

Attempts to examine the impacts of training on business outcomes face additional difficulties when training influences the rate of business survivorship or the likelihood of business start-up. If training leads to the survival of relatively unsuccessful firms that would otherwise have closed, then a straight comparison of profits or sales by treatment status will understate the impact of training. Note that even if training has no impact on the rate of business survivorship or start-up, it may still affect the characteristics of which firms survive, requiring authors to use nonexperimental methods to address this selectivity. For example, de Mel et al. (2012) find that training (and grants) leads to changes in the characteristics of who opens businesses, even though the rates of ownership do not differ in the treatment and control groups. They therefore use a generalized propensity score to reweight their regression estimates to correct for the selectivity they find on observables such as ability and wealth.

**Measurement Changed by Training**

A final challenge in measuring the impact of business training on business outcomes is measuring those outcomes. Start-up and survivorship are objective measures that can be verified, whereas business practices, profitability, and revenues are difficult to measure for most firms. Business practices (for example, keeping accounts, separating business and household expenses, advertising in the past month) are normally relatively easy concepts for firms to understand and are questions that firm owners are usually willing to answer. However, Drexler et al. (2012) note that treated individuals may report performing certain behaviors (for example, separating personal and business accounts) because the training told them this was important rather than because they actually perform the behavior.

Measuring profits and revenues poses further problems. Owners of the smallest businesses typically do not keep written records of these items, and owners of larger firms who do keep records may be reluctant to share them. De Mel et al. (2009a) study several approaches to obtaining profits from microenterprises and conclude that, in their context at least, a simple, direct question is more accurate and much less noisy than calculating profits from revenues and expenses. However, collecting profits has proved difficult for many studies, and several studies have not collected profit data at all (Valdivia 2012; Klinger and Schündeln 2011), have collected it but not used it because of too much noise (Drexler et al. 2012), or have collected only profit margins on the main product rather than overall profits (Karlan and Valdivia 2011). Most studies have collected revenue data, but some have struggled with
much lower response rates for revenues than for nonfinancial business questions (for example, Drexler et al. (2012) have a 46 percent attrition rate on revenues compared to 13 percent for their questionnaire as a whole).

Even when studies are able to obtain data on profits and sales, business training may change the reporting of this data irrespective of whether it actually changes profits and sales. This may occur because the practices taught in the training course lead to more accurate accounting or because training recipients are less likely to underreport profit and sales levels because, for example, they trust the enumerators more after being given the training. Few studies to date attempt to address this issue. Exceptions are Drexler et al. (2011), who examine reporting errors (for example, reporting profits higher than sales or bad week sales higher than average sales) to determine whether treatment reduces these reporting errors and the difference between self-reported profits and profits calculated as the difference between revenue and expenses; Berge et al. (2011), who compare self-reported profits to revenue minus expenses for treatment versus control groups; and de Mel et al. (2012), who do the same and who control for detailed measures of accounting practices as a further robustness check. De Mel et al. (2012) find little evidence that training has changed reporting, whereas Drexler et al. (2012) find that their rule-of-thumb training reduces the number of errors in reporting, and Berge et al. (2011) find that training increases the gap between self-reported profits and revenue minus expenses.

Impacts of Business Training Interventions

The previous section highlights issues with statistical power, timing of follow ups, attrition, and measurement that present challenges for interpreting the impacts identified in the different studies. With these caveats in mind, we examine the extent to which business training is found to impact business start-up and survivorship, business practices, business outcomes, and outcomes for microfinance lenders. Because studies of other microenterprise interventions (De Mel et al. 2009b) often find differences by gender, we separate results by gender to the extent possible.

Impacts on Start-up and Survivorship

Table 7 summarizes the impacts of different studies on business survivorship and new business start-ups. The coefficients are marginal effects on the probability of either outcome occurring, so a coefficient of 0.06 can be interpreted as a 6 percentage point increase. Consider first the impact on business survival. Survivorship is difficult to examine when attrition rates are high because closing is often a cause of
<table>
<thead>
<tr>
<th>Study</th>
<th>Gender</th>
<th>Impact on Survival</th>
<th>95% CI</th>
<th>Impact on Start-up</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bruhn and Zia (2012)</td>
<td>Mixed</td>
<td>0.013</td>
<td>(-0.09, +0.10)</td>
<td>0</td>
<td>n.r.</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>-0.125</td>
<td>n.r., not sig.</td>
<td>0</td>
<td>n.r.</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.072</td>
<td>(-0.07, 0.21)</td>
<td>0</td>
<td>n.r.</td>
</tr>
<tr>
<td>Calderon et al. (2012)</td>
<td>Female</td>
<td>-0.034</td>
<td>(-0.13, +0.06)</td>
<td>n.r.</td>
<td>n.r.</td>
</tr>
<tr>
<td>De Mel et al. (2012)</td>
<td>Female</td>
<td>-0.026</td>
<td>(-0.102, +0.051)</td>
<td>n.r.</td>
<td>n.r.</td>
</tr>
<tr>
<td></td>
<td>n.r.</td>
<td>n.r.</td>
<td></td>
<td>0.09 (4 months)</td>
<td>(0, 0.18)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.02 (25 months)</td>
<td>(-0.11, 0.07)</td>
</tr>
<tr>
<td>Giné and Mansuri (2011)</td>
<td>Mixed</td>
<td>0.034</td>
<td>(-0.021, 0.089)</td>
<td>-0.006</td>
<td>(-0.02, +0.01)</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td><strong>0.061</strong></td>
<td><strong>(-0.012, 0.133)</strong></td>
<td>-0.011</td>
<td>(-0.04, +0.01)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>0.001</td>
<td>n.r., not sig.</td>
<td>0.002</td>
<td>n.r., not sig.</td>
</tr>
<tr>
<td>Glaub et al. (2012)</td>
<td>Mixed</td>
<td>0.05</td>
<td>n.r.</td>
<td>n.r.</td>
<td>n.r.</td>
</tr>
<tr>
<td>Karlan and Valdivia (2011)</td>
<td>Female</td>
<td>n.r.</td>
<td>n.r.</td>
<td>-0.019</td>
<td>(-0.05, +0.01)</td>
</tr>
<tr>
<td>Klinger and Schündeln (2011)</td>
<td>Mixed</td>
<td>n.r.</td>
<td>n.r.</td>
<td>0.044</td>
<td>(-0.12, 0.21)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>n.r.</td>
<td>n.r.</td>
<td>-0.019</td>
<td>(-0.31, +0.27)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>0.465</strong></td>
<td><strong>(0.10, 0.82)</strong></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>n.r.</td>
<td>n.r.</td>
<td><strong>0.572</strong></td>
<td><strong>(0.04, 1.10)</strong></td>
</tr>
<tr>
<td>Mano et al. (2012)</td>
<td>Male</td>
<td><strong>0.095</strong></td>
<td><strong>(0.022, 0.167)</strong></td>
<td>n.r.</td>
<td>n.r.</td>
</tr>
<tr>
<td>Premand et al. (2012)</td>
<td>Mixed</td>
<td>n.r.</td>
<td>n.r.</td>
<td><strong>0.04</strong></td>
<td><strong>(0.02, 0.06)</strong></td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>n.r.</td>
<td>n.r.</td>
<td><strong>0.06</strong></td>
<td><strong>(0.04, 0.08)</strong></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>n.r.</td>
<td>n.r.</td>
<td><strong>0.03</strong></td>
<td><strong>(0.01, 0.05)</strong></td>
</tr>
<tr>
<td>Valdivia (2012)</td>
<td>General training</td>
<td>Female</td>
<td><strong>-0.045</strong></td>
<td><strong>(-0.094, +0.004)</strong></td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>Training + technical assistance</td>
<td>Female</td>
<td>0.021</td>
<td>(-0.014, +0.056)</td>
<td>-0.006</td>
</tr>
</tbody>
</table>

Notes: 95% CI denotes 95 percent confidence interval. Impacts significant at the 10 percent level or more reported in bold. n.r. denotes not reported. Not sig. denotes point estimate is not significantly different from zero. Berge et al. (2011) and Drexler et al. (2012) do not report impacts on either survivorship or start-up. Note Valdivia (2012) survival is based on whether they stopped any business in the past two years, while start-up is based on whether they started a new business in the last year.
attrition, and bounds that allow for attrition can be very wide. Because many studies examine impacts over only a short time, rates of business failure are often low. However, there are exceptions. Bruhn and Zia (2012) find that 36 percent of businesses close during their study period in Bosnia, a rate that is due in part to the downturn caused by the global economic crisis, while Calderon et al. (2012) find that 50 percent of the nonattriting businesses close by the time of their second follow-up survey 28 months after training.

The only study with a survival effect significant at the 5 percent level is Mano et al. (2012), which finds a 9 percentage point increase in the likelihood of survival 12 months after training. These authors do not provide bounds for this effect that control for survey attrition, but they note that none of the training participants had closed. Giné and Mansuri find a 6 percent increase in the likelihood of survival 18 to 22 months after training for the male owners in their sample, an effect that is significant at the 10 percent level, but no change for female owners, whereas Valdivia (2012) finds that training leads to a marginally significant reduction in the likelihood of survival for female firm owners. He attributes this phenomenon to the possibility that training teaches owners to close losing firms. The remaining studies that report survivorship find insignificant impacts but with confidence intervals that are wide enough to include at least a 5 percentage point increase or decrease.

Studies that focus on existing firm owners sometimes consider the start-up of a second business, but none has found significant impacts. However, studies that focus on training specifically tailored for starting new businesses have found some impacts. Klinger and Schündeln (2011) find very large point estimates for entry one year after participation in the second phase of Technoserve’s business plan competition in which training occurs, although the confidence intervals are very wide, and this impact includes the joint impact of grants given to the winners. Premand et al. (2012) examine a sample of 1,500 youths and find that participation in an entrepreneurship track rather than an academic track in the final year of university leads to an increase in self-employment rates of 6 percent for males and 3 percent for females one year later. Four months after training, Field et al. (2010) examine whether women reported business income over the preceding week, which reflects a combination of an effect on business start-up and an effect on survival. They find that upper-caste Hindu women who took the training were 19 percentage points more likely to report income, whereas the training had no effect on lower-caste Hindu women or on Muslim women. They attribute the lack of impact on these groups to social restrictions, arguing that training helped women whose businesses had been limited by social restrictions, but women who faced more extreme restrictions could not respond to training.

Training therefore appears to generate some short-run impacts on business start-up. However, this effect does not necessarily increase employment among trainees, who may simply switch from wage work. Premand et al. (2012) and de Mel et al.
(2012) both find that short-run increases in self-employment from training are coupled with reductions in the likelihood of wage work, so net employment effects on trained individuals are insignificant. Moreover, it is unclear whether training merely speeds up the rate of entry or permanently increases it. De Mel et al. (2012) find that training alone increases the rate of business ownership among a group of women out of the labor force by 9 percentage points within four months of the training, and giving these women grants increases this effect to 20 percentage points. However, by 16 and 25 months after training, the control group catches up. Given the short time horizon of the other studies that have found start-up impacts, it is unclear whether they too would show these effects dissipating over longer time horizons.

**Impacts on Business Practices**

A first link in the causal chain from business training to business profitability and growth is that business training improves the knowledge and implementation of business practices by business owners. There may be other potential mechanisms through which training affects business outcomes (for example, changing attitudes or work hours). However, failure to find any change in practices should cast doubt on the ability of the training to improve firm outcomes.

Table 8 summarizes the impacts identified by various studies on business practices. Almost all studies find a positive effect of business training on business practices, although the effect is often not significant once the sample is divided by gender. Studies differ in what specific practices they measure, how comprehensively they measure them, and how (if at all) they aggregate them. Several studies measure only one to three basic practices, such as Calderon et al. (2012), who examine whether the firm uses formal accounting, and Mano et al. (2011) who record whether the firm keeps records, whether it analyzes them, and whether it visits customers. Others record a broader range of practices, including different types of record keeping, different marketing activities, and other specific practices taught in the training.

One common approach to aggregating different practices is to normalize each practice as a z score (subtracting the mean and dividing by the standard deviation) and then to average these z scores. A coefficient of 0.03, as in Karlan and Valdivia (2011), is interpreted as an impact of 3 percent of a standard deviation. This is useful for considering the magnitude of the increase in relative terms, but it does not provide much guidance regarding the absolute size of the effect. Alternatively, one can examine the percentage point increase in the likelihood that a particular practice will be implemented or the change in the number of practices implemented out of some total, both of which provide more guidance on the absolute magnitude of the increase.
Many studies find baseline levels of business practices that are relatively low. For example, Giné and Mansuri (2011) report that only 18 percent of firms record money taken from the business, and only 18 percent record sales. Even among larger metalwork firms, Mano et al. (2012) report that only 27 percent of their sample keep business records, and only 20 percent visit customers at baseline. Although most studies find significant increases in the use of business practices taught during the training, the magnitude of these effects, although sometimes large in relative terms, is often small in absolute terms. For example, Drexler et al. (2012) find that rule-of-thumb training leads to an increase in individuals reporting that they separate personal and business expenses, keep accounting records, and calculate revenues formally, with each of these measures increasing 6 to 12 percentage points relative to the control group. In Giné and Mansuri (2011), treatment

### Table 8. Impact of Business Training on Business Practices

<table>
<thead>
<tr>
<th>Study</th>
<th>Gender</th>
<th>Units</th>
<th>Number of Practices</th>
<th>Point estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Berge et al. (2011)</td>
<td>Male p.p.</td>
<td>4</td>
<td>0.03-0.08</td>
<td>n.a.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female p.p.</td>
<td>4</td>
<td>-0.02-0.00</td>
<td>n.a.</td>
<td></td>
</tr>
<tr>
<td>Bruhn and Zia (2012)</td>
<td>Mixed s.d.</td>
<td>3</td>
<td>0.272 (0.03, +0.51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Male s.d.</td>
<td>3</td>
<td>0.290 (0.01, 0.57)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female s.d.</td>
<td>3</td>
<td>0.214 n.r.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calderon et al. (2012)</td>
<td>Female p.p.</td>
<td>1</td>
<td>0.062 (-0.02, +0.14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>De Mel et al. (2012)</td>
<td>Current Enterprises Female num</td>
<td>29</td>
<td>2.03 (1.27, 3.30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Potential Enterprises Female num</td>
<td>29</td>
<td>0.87 (-0.23, +1.97)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drexler et al. (2012)</td>
<td>Mostly Female s.d.</td>
<td>12</td>
<td>0.14 (0.06, 0.22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Rule-of-thumb&quot;</td>
<td>Mostly Female s.d.</td>
<td>12</td>
<td>0.07 (-0.03, 0.17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Standard&quot;</td>
<td>Mixed s.d.</td>
<td>3</td>
<td>0.131 (0.01, 0.25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giné and Mansuri (2011)</td>
<td>Male s.d.</td>
<td>3</td>
<td>0.114 (-0.05, 0.28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female s.d.</td>
<td>3</td>
<td>0.140 n.r.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Karlan and Valdivia (2011)</td>
<td>Mostly Female s.d.</td>
<td>14</td>
<td>0.03 (0.00, 0.06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mano et al. (2012)</td>
<td>Male p.p.</td>
<td>3</td>
<td>0.24-0.42 n.a.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valdivia (2012)</td>
<td>General training Female s.d.</td>
<td>11</td>
<td>0.01 (-0.02, 0.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Training + technical assistance Female s.d.</td>
<td>11</td>
<td>0.05 (0.02, 0.08)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** 95% CI denotes 95 percent confidence interval. Impacts significant at the 10 percent level or more reported in bold. Units for measuring practices are either standard deviations of a normalized aggregate (s.d.), percentage points (p.p.), or number of distinct practices improved (num).

Number of practices is the total number of practices measured.

When no aggregate measure is reported, the range of point estimates for individual practices is given. n.r. denotes not reported. n.a. denotes not applicable since range of estimates given.

(a) we include here their index of three marketing practices, plus their result on record-keeping. No aggregate measure is provided.
impacts include a 6.6 percentage point increase in recording sales and a 7.6 percentage point increase in recording money taken for household needs. In de Mel et al. (2012), existing enterprises implement an additional two practices out of 29. Mano et al. (2012) are an exception in this regard: they find a 30 percentage point increase in the percentage of firms keeping records in the treatment versus the control group. However, in general, given that the magnitude of the changes in business practices is relatively small, we might expect it to be difficult to detect impacts of these changes on business outcomes.

**Impacts on Business Profits and Sales**

Ultimately, from the viewpoint of an individual firm owner, an investment in training is justified only if there is an increase in profits. However, as noted previously, many studies struggle to measure profits, so not all studies consider this as an outcome. Table 9 summarizes those studies that do, converting, where necessary, point estimates of profit or sales levels to percentage increases relative to the control group mean to enhance comparability across studies. Several studies examine gender heterogeneity by reporting a point estimate for males and then an interaction effect for females, but they do not test the overall impact on females. Therefore, the table sometimes shows confidence intervals for males but not for females. Often, studies have more than one specification for profits or revenues, with variation in whether they include different controls and whether they truncate or trim the data or take a log transformation. We report impacts on the measure that corresponds most closely to profits or sales in the previous month. The data shown in the table do not account for differential attrition, though some studies report bounds that adjust for attrition.

The table shows that few studies detect significant impacts of business training on business profits or sales, although the confidence intervals are very wide in many cases. The wide confidence intervals reflect the issue of statistical power discussed earlier. The studies that have the most power according to the calculations in table 5 are the ones that are most likely to show significant effects. Berge et al. (2011) find that training increases profits by 24 percent and sales by 29 percent for males in the short run (five to seven months posttraining), but the point estimate of the impact on profits drops to 5 percent and is statistically insignificant in their longer-term follow up (30 months posttraining). There is a continued and marginally significant impact on sales. Their point estimates are much closer to zero and statistically insignificant for women in both the short and medium term. De Mel et al.’s (2012) study also has enough power to detect reasonable changes in profits. They find no impact of training alone on profits of existing firms over either the short or medium run, but they do find significant impacts of the combination of training and a grant on short-run profits, with these gains dissipating over time. In
a separate sample of women who were out of the labor force at baseline, training increased the profits and sales of start-up businesses by a statistically significant 40 percent, although the confidence intervals around this level are wide.

Calderon et al. (2012) find a 24 percent increase in weekly profits and a 20 percent increase in weekly revenues, both significant at the 10 percent level. However, given that attrition is 26 percent by the second round survey and that 50 percent of the nonattritors have closed, there is reason to be cautious in interpreting this estimate of the impact on surviving nonattriting firms. The only other study to find significant impacts on revenues, Valdivia (2012), finds a 20 percent increase for the group that received both training and intensive one-on-one technical

<table>
<thead>
<tr>
<th>Study</th>
<th>Gender</th>
<th>% increase</th>
<th>95% CI</th>
<th>% increase</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berge et al. (2011)</td>
<td>Male</td>
<td>5.4%</td>
<td>(-20%, +38%)</td>
<td>31.0%</td>
<td>(-4%, +79%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>-3.0%</td>
<td>(-23%, +22%)</td>
<td>4.4%</td>
<td>(-23%, +22%)</td>
</tr>
<tr>
<td>Bruhn and Zia (2012)</td>
<td>Mixed</td>
<td>-15%</td>
<td>(-62%, +32%)</td>
<td>n.r.</td>
<td>n.r.</td>
</tr>
<tr>
<td>Calderon et al. (2012)</td>
<td>Female</td>
<td>24.4%</td>
<td>(-1%, 56%)</td>
<td>20.0%</td>
<td>(-2%, +47%)</td>
</tr>
<tr>
<td>De Mel et al. (2012)</td>
<td>Current Enterprises</td>
<td>Female</td>
<td>-5.4%</td>
<td>(-44%, +33%)</td>
<td>-14.1%</td>
</tr>
<tr>
<td></td>
<td>Potential Enterprises</td>
<td>Female</td>
<td>43%</td>
<td>(+6%, +80%)</td>
<td>40.9%</td>
</tr>
<tr>
<td>Drexler et al. (2012)</td>
<td>&quot;Standard&quot;</td>
<td>Mostly Female</td>
<td>n.r.</td>
<td>n.r.</td>
<td>-6.7%</td>
</tr>
<tr>
<td></td>
<td>&quot;Rule-of-thumb&quot;</td>
<td>Mostly Female</td>
<td>n.r.</td>
<td>n.r.</td>
<td>6.5%</td>
</tr>
<tr>
<td>Giné and Mansuri (2011)</td>
<td>Mixed</td>
<td>-11.4%</td>
<td>(-33%, +17%)</td>
<td>-2.3%</td>
<td>(-15%, +13%)</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>-4.3%</td>
<td>(-34%, +38%)</td>
<td>4.8%</td>
<td>(-14%, +27%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>n.r. (a)</td>
<td>n.r.</td>
<td>n.r. (a)</td>
<td>n.r.</td>
</tr>
<tr>
<td>Glaub et al. (2012)</td>
<td>Mixed</td>
<td>n.r.</td>
<td>n.r.</td>
<td>57.4% (c)</td>
<td>n.r.</td>
</tr>
<tr>
<td>Karlan and Valdiva (2011)</td>
<td>Mostly Female</td>
<td>17% (b)</td>
<td>(-25%, +59%)</td>
<td>1.9%</td>
<td>(-9.8%, +15.1%)</td>
</tr>
<tr>
<td>Mano et al. (2012)</td>
<td>Male</td>
<td>54%</td>
<td>(-47%, +82%)</td>
<td>22.7%</td>
<td>(-31%, +76%)</td>
</tr>
<tr>
<td>Valdivia (2012)</td>
<td>General training</td>
<td>Female</td>
<td>n.r.</td>
<td>n.r.</td>
<td>9%</td>
</tr>
<tr>
<td></td>
<td>Training + technical assistance</td>
<td>Female</td>
<td>n.r.</td>
<td>n.r.</td>
<td>20.4%</td>
</tr>
</tbody>
</table>

Notes: 95% CI denotes 95 percent confidence interval. Impacts significant at the 10 percent level or more reported in bold. n.r. denotes not reported.

(a) They look at an aggregate sales and profitability measure and find no significant impact for either gender.

(b) Impact on profit from main product.

(c) Calculated as difference-in-difference calculation. Study reports difference in log sales is significant at the 1 percent level. Profit increases are scaled as a percentage of the control group mean to enable comparability. When multiple rounds are used, longest-term impacts available are reported.
assistance but no significant increase for training alone. Finally, Glaub et al. (2012) find a positive effect of personal initiative training on sales one year later, although they do not survey the noncompliers (individuals selected for training who do not attend), which is problematic if there is selective participation.

Several studies have emphasized the possibility that business training may have its strongest impact on sales during a bad month. The working paper version of Karlan and Valdivia (2011) stressed this avenue, noting that training might help clients identify strategies to reduce downward fluctuations in sales by considering diversifying the products that they offer and by being more proactive about alternative activities during slow months. The working paper estimate, which has gained some policy attention, showed a 30 percent increase in sales during a bad month. However, the published version of the paper deemphasizes this impact, noting that when an alternative (and now preferred) specification is used, the impact falls to an insignificant 5 percent to 7 percent increase. The possibility that training may be particularly valuable during bad times is also emphasized by Drexler et al. (2012), who find that their rule-of-thumb training leads to an increase in sales during bad weeks that is significant at the 10 percent level. However, Drexler et al. also ask firm owners to report sales in a bad month and find a very small and insignificant impact of training on this measure. Giné and Mansuri (2011), de Mel et al. (2012), and Valdivia (2012) find no significant impacts of training alone on sales during bad months. Viewing these studies together leads us to conclude that the evidence that training has particularly strong effects during bad periods is weak.

A microenterprise earning $1 per day would need to see only a 13.7 percent increase in profits to recoup the cost of $100 of training over two years. The confidence intervals for the studies that consider profits are almost all wide enough to include this level of return. For larger firms, the percentage increase in profits required to repay training costs is likely lower because the costs of training often increase more slowly than the size of the firm undertaking the training. For example, a firm with $500 in monthly profits would only need a 2 percent increase in monthly profits to recoup $250 worth of training costs over two years. The result is that training costs may be justified by increases in profits that are far too small for existing studies to detect.

**Impacts on Employment**

A further justification by policymakers for subsidizing business training is that business growth may have broader benefits for others in the community by increasing employment opportunities. For programs working with microenterprises, the most direct employment impacts are likely to be for the owner himself or herself, increasing employment by increasing the likelihood of starting a new business or reducing the chance of business failure.
The few studies using samples of microenterprises that report impacts on employment of other workers robustly show very small and statistically insignificant effects. Karlan and Valdivia (2011) find an increase of 0.02 workers, Valdivia (2012) finds a decrease of 0.06 workers from straight training and a similar decrease from training plus technical assistance, and Drexler et al. (2012) find an increase of 0.05 workers from standard training and a decrease of 0.02 workers from rule-of-thumb training. None of these impacts is statistically significant, but their point estimates suggest that no more than one in 20 microenterprises that take business training will hire an additional worker.

The one study to show a stronger employment effect is Glaub et al. (2012), which hints at the possibility of employment impacts when training larger firms. These authors find that employment in treated firms grows from 7.9 employees at baseline to 10.7 at follow up, whereas employment in control firms falls from 6.6 employees at baseline to 5.0 at follow up. This difference is significant at the 5 percent level. Their sample is small, and they drop noncompliers to their treatment, so this result is likely an overstatement of the effect. More studies with larger firms are needed.

Impacts on Microfinance Institution Outcomes

Because many of the studies work with microfinance clients, they also consider outcomes using administrative data from the microfinance organization. These data have the advantage of being available with less attrition and over longer periods, and they are useful for assessing whether offering training is cost effective for the microfinance organization. However, these data are less useful for explaining how such training affects firms. Karlan and Zinman (2011) find that training results in a 4 percentage point increase in client retention rates and a 2 to 3 percentage point increase in the likelihood of perfect repayment (although this is only marginally significant). However, they also note that some of the clients who leave cite the added length of the weekly meetings due to the training sessions as a factor in dropping out of the program. They note that these benefits appear to make the training profitable from the lender side. After their study, FINCA Peru implemented the mandatory version of their training in all village banks.

Gine and Mansuri (2011) find that training leads to a 16 percent increase in loan size for males, a reduction in loan size for females, and no change in repayment rates. They also find a change in the selection of who borrows; individuals with higher predicted probabilities of default are less likely to borrow after training. Field et al. (2010) find that upper-caste Hindu women are 13 percentage points more likely to borrow after training. In contrast, Drexler et al. (2012) and Bruhn and Zia (2012) find no significant impacts of training on the likelihood of taking loans or loan size, although Bruhn and Zia find an increase in loan duration and
more refinancing of loans. They attribute this finding to trained individuals making longer-term investments and being more aware of available interest rates.

Boosting the Intensity and Working with Larger Firms

Many of the training sessions are relatively brief, and the increase in business practices has been relatively small in a number of studies. One response to this phenomenon is that more in-depth and individualized follow ups on the training are needed, whereas another response is to focus on larger firms in which management practices may be of greater importance. We discuss the results of studies that have pursued these two approaches.

Individualized Follow Ups

Three of the business training evaluations had a treatment group that added individualized follow ups to the classroom training. In Drexler et al. (2012), trainers visited eight times over five months to answer queries, verify and encourage the use of accounting books, and correct any mistakes in completing books. These authors find no significant effects of this additional follow up. Gine and Mansuri (2011) added “hand-holding sessions” in half of the community organizations, with firms receiving visits one to two times per month for four months to discuss topics learned, answer questions, and suggest solutions to potential problems. They find that this hand holding had no effect on any of the aggregate outcomes for either men or women.

In both of these cases, the follow ups mostly reinforced the general business skills taught in training rather than providing firm-specific individualized advice. Valdivia (2012) examines more intensive follow up, with trainers providing specific technical assistance tailored to the needs of women’s businesses. The follow ups combined individual visits with group sessions among small groups of similar businesses during a three-month period. This component included 22 three-hour group sessions and five to six hours of individual sessions or visits. Valdivia finds some evidence to suggest this technical assistance helped firms; women assigned to receive the assistance experienced a 20 percent increase in revenue relative to the control group (significant at the 1 percent level) and showed more improvement in business practices than women who were assigned to only the basic training. This additional attention cost twice as much as the basic training alone.

Individual Consulting

A related body of literature examines the impact of providing consulting services on a one-on-one basis to firms to improve business and management practices. The
closest study to the business training experiments is the work of Karlan et al. (2012), who examine a mix of 160 male and female tailors in Ghana with five or fewer workers. Their study used local consultants from Ernst and Young in Ghana, who met with the tailors for 30 minutes to 1 hour several times a month over one year, with the average firm receiving 10 hours of consulting over a year at no cost to the firm. They find that some of the consultants’ recommendations were adapted for some months but had been abandoned one year after training stopped. There is no significant impact of either treatment on profits or revenues, with some specifications showing negative effects in the short run, although the power is very low and confidence intervals are wide.

Bruhn et al. (2012) evaluate a state government program in Puebla, Mexico, that paired small businesses with a consultant from one of several local consulting firms. Consultants spent approximately four hours per week over a year assisting the firm in overcoming constraints to growth. A total of 432 firms applied to the program, and 150 were chosen to receive heavily subsidized consulting services (at a cost that was approximately 10 percent of the commercial rate). The mean number of employees was 14, and 72 percent were male-owned firms. The training impact was assessed with a single follow-up survey one to three months after the consulting. The authors find large point estimates for the impacts on sales and profits, which are sometimes significant depending on the measure used and the extent of trimming. However, the study faces many of the same challenges as the business training studies reviewed above. First, the firms in the sample are very heterogeneous, with a baseline coefficient of variation in sales of 3.7, and 2.4 even after trimming the top 1 percent. Second, even though all firms signed a statement of interest, only 80 of the 150 firms (53 percent) assigned to treatment participated in the consulting. Third, attrition rates were reasonably high, and there was additional item nonresponse on profits and sales even among those who were interviewed, so only 288 firms (66.7 percent) provided data on profits in the follow-up survey. These challenges are likely to face any similar government program offering subsidized consulting or business services to firms, such as the matching grant programs used in many World Bank private sector loans.

The final individualized consulting study is Bloom et al. (2013), who focus on a much smaller sample of 17 large textile firms in India. The typical firm in their sample has 270 employees, two plants, and sales of $7.5 million per year. They provided 11 of these firms with five months of free intensive consulting from Accenture Consulting. The consultants averaged 781 hours per treated plant, working with the firms to implement 38 key management practices related to quality control, factory operations, inventory, human resource management, and sales and order management. They address the problem of small sample size by focusing on very homogeneous firms and collecting large amounts of data from them, including weekly data on quality, output, and inventories. They find that adopting these
management practices raised productivity by 17 percent in the first year through improved quality and efficiency and reduced inventory, and they find some evidence that within three years, adopting these practices led to the opening of more production plants. The results show that in large firms, at least, changing management practices can lead to substantial improvements in firm performance. However, the authors can only indirectly estimate the changes in profits from this effort.

What We Do Not Know

There are now a range of studies on a variety of business training programs that examine impacts on business practices, business outcomes, and (sometimes) outcomes for microfinance institutions. However, existing studies leave a number of open questions that are important in considerations of the case for policy action to support business training.

Who Does Training Help Most?

Our discussion above touches on heterogeneity in outcomes by the gender of the owner and, to some extent, across studies by firm size. Several studies have examined heterogeneity in other dimensions, such as the owner’s education and baseline business skill levels, business sector, and interest in training. However, the low power of most studies to find average effects for the full sample indicates low power for examining the heterogeneity of effects.

As a result, the question of who benefits most from training—or which types of training are most suitable for which types of firms—remains unanswered. On one hand, poor subsistence firms whose owners run the business only because they cannot find a wage job may have very low business skills. Thus, it should be relatively easy for them to make improvements. However, the owners may be less interested or able to implement the practices taught, or these practices may only have an effect when businesses reach a larger scale. There is much talk of targeting gazelles—firms that grow rapidly—but even if the characteristics to identify such firms in advance can be defined, it is unclear whether these firms need the help or would grow rapidly anyway. Theoretically, it would be preferable to target firms in which skills are the binding constraint on growth, but there is little evidence to date to determine which firms these are, especially among the smallest firms.
How Does Training Help Firms, and Do Gains Come at the Expense of Other Firms?

Most studies have not explored the channels through which training affects business outcomes. In part, this omission reflects the lack of power in detecting an impact on profits in the first place. Does training enable firm owners to use the same inputs more efficiently—thereby reducing costs and wastage—or is the main impact due to increasing revenues at the same cost ratios by new marketing and sales efforts? The policy implications differ depending on the channels. In particular, one possibility is that gains for the treated firms are due to these firms taking customers from other firms.

Such spillovers have implications for both internal and external validity. If the increased sales are mainly due to taking business from the control group firms, then the stable unit treatment value assumption, which assumes that the outcomes of each firm are not affected by the treatment statuses of other firms, is violated. As a result, the experimental estimate no longer provides the average impact of training for the sample population. If the increased sales are mainly due to other firms not in the sample, the results of the experiment could be misleading with respect to the gain to society from scaling up the training program. It should be noted that spillovers might instead be positive if control or nonsample firms copy some of the techniques or new products introduced by firms that have participated in training. Indeed, this possibility is often given as one of the main justifications for public subsidies of matching grant programs that subsidize the purchase of business development services by SMEs. These issues are part of the broader question of how competition responds to newly trained firms. We do not know whether this deters some new firms from entering the industry, causes others to exit, or causes the incumbents who remain in business to make other changes to the way they run their businesses.

To investigate this issue, a much larger sample is needed. Experimental variation in the intensity of the treatment within different geographical areas could be used to test for and measure these spillovers. An example in the context of labor programs for youth is found in Crepón et al. (2011). A first attempt in this direction for business training is found in Calderon et al. (2012), who randomly assigned 17 villages into seven treatment villages and 10 control villages, with half the individuals in the treated villages assigned to training. Their preliminary analysis surprisingly finds little evidence for spillovers despite working in remote villages with 1,500 or fewer households and with firms that mostly make or sell goods for local consumption. However, it is unclear how much power is available to examine these spillovers given the relatively small number of villages included in the study.
Do Larger Impacts Emerge over Time?

Most of the studies take a single snapshot of the impact of training a relatively short time after training has ended. Two studies that have traced the trajectories of impacts suggest that effects may vary considerably over time. In de Mel et al. (2012), the impacts on business start-up fade over time as control firms catch up. Bloom et al. (2013) find that introducing management practices in larger firms shows immediate effects on quality and then slowly leads to changes in inventory levels, output, and productivity. The impacts begin to appear in terms of employment generation (through new plants opening) only after several years of using these practices. Given the interest of many governments in employment creation, studies that consider only a year or so after treatment may miss effects that take some time to be realized—or, conversely, we may find that effects that seem promising in the short term dissipate over time.

If Training Is so Helpful, Why Do Firms not Purchase It?

It is notable that all of the business training studies reviewed here offer the training for free, as do two out of the three consulting experiments, with the other offering a 90 percent subsidy. In part, this approach is used for research purposes, to ensure sufficiently high take-up and to provide evidence on how training influences a range of firms. Even with this approach, we have seen limits to demand, with some studies struggling to encourage people to undertake training even when it is offered without cost.

As a result, we know very little about what types of firms would choose to purchase training at market prices and the effects of training on this subgroup of firms. Public intervention is typically motivated by the belief that market failures prevent firms who would benefit substantially from training from purchasing this training at market prices or a belief that there are positive externalities from training that lead firms to underinvest relative to what is socially desirable.

Even if market failures exist, the first-best solution would be to fix these market failures rather than to give away training for free or highly subsidized rates. However, given the difficulty of alleviating some of these market failures in many developing countries, subsidizing training may be seen as a feasible second-best solution. Several potential constraints or market failures are discussed in the literature. The first, and the one for which there is the most support (Karlan and Valdivia 2011; Bloom et al. 2013), is that of an information failure: entrepreneurs do not understand the value of business training. Those with the most to gain may understate the value the most because they do not realize how poorly their firms are run.

A second market failure is credit constraints. Firms may find it more difficult to borrow to finance training, an intangible asset, than to finance assets that could be...
seized by a bank in the event of nonrepayment. There is strong evidence that many microenterprises are credit constrained (de Mel et al. 2009b), but there is much weaker evidence to support the view that this is the key constraint to purchasing business training services.

A third possibility is the failure of insurance markets. Firm owners may be reluctant to take training even if they think it has a high expected payoff because they are unable to insure against the possibility that it will not work. There is some recent evidence to support the view that risk is a constraint to start-up and investment in small businesses (Bianchi and Bobba Forthcoming), but no evidence of which we are aware shows that alleviating this constraint leads to more purchases of training.

A fourth possibility is supply-side constraints. Consulting or training services simply may not exist in the market. Thus, even if a firm wants to purchase these services, it is unable to. This is likely to be true in some countries and areas, but in many others, such services do exist.

Even with market failures, public financing is not justified if the gains to training are realized entirely by the firms being trained unless the financing is provided with the goals of either poverty alleviation (raising the incomes of these particular firm owners) or productivity enhancement. More typically, public spending is justified by claims of positive spillovers, whereby the public gains from training are believed to greatly exceed the private gains, causing firms to underinvest. Such externalities have yet to be demonstrated empirically.

The optimal policy response differs depending on which constraint binds, so making progress on the issue of why more firms do not purchase training or consulting is likely to have useful implications for policy efforts.

*Practices or Personality?*

Business training courses have traditionally focused on teaching particular practices that firm owners can implement in their firms. However, another school of thought is that the attitudes and personalities that business owners bring to the business are equally, if not more, important. Premand et al. (2012) report that one of the main objectives of the trainers in their study was to change the students’ personalities to “turn them into entrepreneurs.” They find that their intervention led to measurable and significant changes in several domains of personality.

There is also a range of training courses studied by psychologists that focus more on the personality of an entrepreneur than on specific skills (Glaub and Frese 2011). Glaub et al. (2012) find some evidence to support a positive impact of such training in Uganda. Although several studies have incorporated some aspect of aspirations or entrepreneurial attitudes into their content, to date, no research tests the relative contribution of each type of training.
Conclusions and Suggestions for Future Work

The last few years have seen rapid growth in the number of randomized trials evaluating business training programs, providing a growing body of evidence in an area with large policy interest. However, a number of challenges have hampered how much we can learn from these studies. Methodological concerns and heterogeneity in both training content and the characteristics of who is trained complicate comparisons across studies. Many of the key questions needed to justify large-scale policy interventions in this area remain unanswered. Researchers continue to learn more about how to better conduct firm experiments, suggesting that these difficulties are not insurmountable.

To learn from the next generation of studies, we believe that the following elements are needed.

1. **Analyzing much larger samples or more homogeneous firms**: Rather than more studies with 100 to 500 individuals in each treatment or control group, we need studies to move to samples of several thousand or more. This would increase the power of the studies and allow more consideration of the types of people for whom training is most effective. An alternative to large cross-sectional samples is to reduce the heterogeneity of the sample by focusing on firms within one industry and size category and collecting much more frequent time series data on these firms (McKenzie 2011, 2012).

2. **Using better measurement of outcomes**: Measuring firm profits and revenues has proved to be a challenge for many studies, and little evidence is available on how training changes a firm’s production process. Further efforts to improve the measurement of financial information (and to ensure that there is not simply a measurement effect of training) are needed. Focus on a specific industry or sector may allow more detailed production-level monitoring of physical outputs and inputs.

3. **Designing experiments to measure spillovers**: These experiments could include greater use of global positioning system data to measure local spillovers (Gibson and McKenzie 2007) and randomization of the intensity of training at the local market level to determine whether effects differ when all firms competing in a local area are trained versus when only some of them are trained, building on the work of Calderon et al. (2012).

4. **Measuring trajectories of outcomes over longer periods**: The impacts of training may differ in the short and medium term. Measuring outcomes at multiple points in time would enable better understanding of whether effects take time to materialize or whether effects that emerge quickly persist.

5. **Testing which elements of content matter**: With larger samples, studies could build on the work of Drexler et al. (2012) and test different forms of training to determine which elements of business skills have the greatest impact and whether training should focus on entrepreneurial personality as well as processes.
However, researchers should avoid the temptation to perform this testing at the cost of insufficient power in each treatment arm.

6. **Understanding market failures and building market-based solutions:** Almost every study has given training away for free and experienced difficulties in take-up. There are many open questions concerning the development of a market for these business services and the types of policies that could overcome the market failures that prevent firms from using these markets.

**Funding**

This work was supported by the World Bank and the UN Foundation.

**Acknowledgments**

We thank the editor, three anonymous reviewers, Louise Fox, and Markus Goldstein for helpful comments and suggestions, and we thank the authors of the studies reviewed in this paper for providing us with additional details and clarifications on their studies.

**Notes**

David McKenzie is Lead Economist in the Finance and Private Sector Development group of the World Bank’s Research Development; email address: dmckenzie@worldbank.org. Christopher Woodruff is Professor of Economics at the University of Warwick.

2. All dollar amounts are U.S. dollars unless otherwise indicated.
3. Kaizen and 5S are Japanese systems for improving production efficiency based on a philosophy of continuous improvement and by improving workflow in a production process through standardized and efficient storage, set-up, and production.
4. See the working paper (McKenzie and Woodruff 2012) for a more technical discussion of the calculation of power in this table.
5. Note that Berge et al. (2011) take existing loan groups who meet on a given day in a given branch and randomly assign training to one of two days in each of the two branches. Thus, true randomization only involves choosing one of four possible allocations and has zero power according to permutation analysis. The authors claim that because loan groups are offered time on the basis of availability, this is as good as random, and so they proceed with analysis as if randomization was at the group level. Our table does the same, but this caveat should be noted.
6. A related concern is that people who take training may overreport profits or revenues after training to exaggerate how well their firms have benefited from training. The same robustness checks described in the text can help to rule out this sort of behavior, as can detailed probing and observation from the surveyors.
7. This effect includes the impact of seed money given to the top placed business plans, but the authors argue via various checks that the impact is not driven by these grants.
8. The sales impact is insignificant when covariates are dropped or clustering is used to attempt to address the fact that randomization did not occur at the loan group level.

9. If we account for discount rates, opportunity costs, and risk aversion, the desired returns would have to be higher. However, a 25 percent increase in profits would still likely provide a very reasonable return to microenterprise training, even after accounting for these factors.

References


This literature review focuses on the relationships between population, poverty, and climate change. Developed countries are largely responsible for global warming, but the brunt of the fallout will be borne by developing countries in forms such as lower agricultural output, poorer health, and more frequent natural disasters. Although carbon emissions per capita have leveled off in developed countries, they are projected to rise rapidly in developing countries because of economic growth and population growth. Unfortunately, the latter will rise most notably in the poorest countries, combining with climate change to slow poverty reduction. These countries have many incentives to lower fertility. Previous studies indicate that in high fertility settings, fertility decline facilitates economic growth and poverty reduction. It also reduces the pressure on livelihoods and frees resources that can be used to cope with climate change. Moreover, slowing population growth helps avert some of the projected global warming, which will benefit the poorest countries far more than it will benefit developed countries that lie at higher latitudes and/or have more resources to cope with climate change. Natural experiments indicate that family-planning programs are effective and highly pro-poor in their impact. While the rest of the world wrestles with the complexities of reducing emissions, the poorest countries will benefit from simple programs to lower fertility. Population, Poverty, Economic growth, Climate change, Global warming, Family planning, sub-Saharan Africa, Developed countries, Developing countries, Public policy, Sustainable development, Ecological injustice JEL codes: Q56, Q54, J13, J18

The relationships among population dynamics, poverty, and climate change are now recognized in the literature on sustainable development. This paper summarizes the evidence currently available on these relationships and their implications for the poorest developing countries. The paper begins with a review of the
literature on population growth and the depletion of natural resources. This relationship has been much debated, with some arguing that human innovation can overcome any natural resource constraint. The consensus now is that although this may apply to resources that are more fully priced, it is much more difficult to manage environmental common property resources. Efforts to price the use of such resources, notably by imposing a carbon tax, have so far met with very limited success.

The literature indicates that there is considerable “ecological injustice” between developed and developing countries. Although developed countries have generated most of the current stock of emissions that cause global warming, the brunt of the burden will be borne by developing countries. These burdens are reviewed in Section 2. Section 3 reviews the gains to the poorest countries from fertility decline, which facilitates economic growth and poverty reduction, helps mitigate the burdens of climate change that they face, and reduces future increases in global warming that will disproportionately affect these countries. Section 4 reviews the evidence on the effectiveness of family-planning programs in helping to reduce fertility. Natural experiments indicate that these programs are effective and have a strong pro-poor impact.

The paper concludes that while the rest of the world wrestles with the political and technological problems of reducing emissions, the poorest countries have available a simple and effective means of using family-planning programs to improve their circumstances.

**Population and Natural Resources**

The publication of the study *Limits to Growth* (Meadows et al. 1972) caused considerable controversy. It summarized the historical trajectory from 1900 to 1970 of non-renewable natural resources, pollution, population size, food production, and industrial output and simulated their trajectory from 1970 to 2100. The study concluded that sustainable development could not be achieved without curtailing population growth and the use of natural resources.

Others have argued that more rapid population growth may help drive economic growth by spurring technological innovation that can potentially stretch resources indefinitely. For example, Boserup (1965) argued that population growth helps induce agricultural innovation and agricultural intensification, allowing greater productivity per unit of land to feed the larger population. Similarly, Simon (1981, 1996) argued that people and markets innovate in response to potential resource shortages, and therefore the resource base is effectively infinite.¹

Simon’s arguments were supported by studies of the costs of some industrial resources, which were found to have fallen sharply between 1870 and 1957, a period
during which there was rapid growth in both population and industrial output (Potter and Christy 1962; Barnett and Morse 1963). There are strong private incentives to find innovative ways of managing the use of such clearly priced resources to keep prices down.

The concerns raised by studies forecasting resource depletion receded quickly as technological innovation rapidly increased agricultural productivity and kept the prices of some commodities down. However, these innovations have had a much smaller effect on reducing the depletion of environmental common property resources.

**The Complexities of Managing Environmental Common Property Resources**

Recently, widespread concern over environmental common property resources has again raised issues of sustainable development.

A driving need for continuing adaptation and innovation is generated by the world’s growing consumption needs, which are associated with increases in per-capita consumption levels and population growth. Technological progress has certainly increased production, but this has not been without negative ramifications. Common property resources are under pressure from activity to meet rising consumption requirements. For example, increasing agricultural production per acre through the higher use of chemicals and fertilizers has been very effective at raising food production, but it has also increased fertilizer runoff, thereby creating low-oxygen “dead zones” in coastal oceans (Map 1).

While market forces provide incentives to find ways to better manage the use of non-renewable resources that are clearly priced, it is proving more difficult to conserve resources that are unpriced or underpriced, such as oceans and the atmosphere. Even understanding the intricacies of environmental change is a challenging task for scientists, and organizing collective action to avert negative consequences is a challenging task for political leaders even at local levels, let alone at national and global levels.

These factors combine to create a daunting list of necessary adaptations and innovations, which are complex to develop and to implement. The World Development Report 2010 summarizes some of the measures needed for sustainable food production (World Bank 2010). To manage land and water resources to feed growing populations and protect natural systems, this report notes the need for politically daunting measures, such as the following:

- building flexible international agreements;
- pricing carbon, food, and energy;
- redirecting agricultural subsidies; and
- strengthening the policy environment for natural resource management.
Conventional estimates of growth in Gross Domestic Product (GDP) are misleading on the sustainability of production possibilities because they ignore the depreciation of natural capital (Arrow et al. 2004; Dasgupta 2010).

“Since GDP is the total value of the final goods and services an economy produces, it does not deduct the depreciation of capital that accompanies production—in particular, it does not deduct the depreciation of natural capital. In the quantitative models that appear in leading economics journals and textbooks, nature is taken to be a fixed, indestructible factor of production. The problem with the assumption is that it is wrong: nature consists of degradable resources. Agricultural land, forests, watersheds, fisheries, fresh water sources, river estuaries and the atmosphere are capital assets that are self-regenerative, but suffer from depletion or deterioration when they are over-used. . . . To assume away the physical depreciation of capital assets is to draw a wrong picture of future production and consumption possibilities that are open to a society.” (Dasgupta 2010, 6)
Moreover, “property rights to natural capital are frequently unprotected or ill-specified, . . ., (which) typically leads to their overexploitation, and so to waste and inequity” (Dasgupta 2010, 6).

Arrow et al. (2004, Table 2) estimate how much “genuine wealth per capita” (including natural capital, human capital, and manufactured capital) changed during 1970–2000. The estimates are necessarily approximate, but they have been made carefully, and the results are instructive. They find that although GDP per capita grew quite rapidly during 1970–2000 in all regions except sub-Saharan Africa, rates of growth in “genuine wealth per capita” were far lower. They declined sharply in sub-Saharan Africa and in the Middle East and North Africa (by -2.6% and -3.8% per year, respectively), and they grew very slowly (well below 1% per year) in South Asia and the United States. They grew rapidly only in China because of its low population growth and heavy investment in productivity. Revising the method to include more information on growth in human capital and institutional change, Dasgupta (2010, 9–10) derives far lower estimates of growth in genuine wealth per capita for China 1970–2000, and for South Asia, he estimates a decline of between -0.4% per year (India) and -1.4% (Pakistan).

Human ingenuity has faced an uphill task in devising ways of managing common property resources given the institutional and political challenges in aligning divergent interests. Markets are very poor at incentivizing people not to overuse resources that are unpriced or under-priced relative to social cost (Arrow 1969; Dasgupta 2001; Stern 2006), especially in the case of transnational common resources (Dasgupta et al. 1997). The consequent negative externalities need to be addressed through collective action, but in the absence of strong mechanisms for mutual coercion and cooperation, it is very difficult to align the interests of different stakeholders to this end. Ostrom (1990) has argued that common property can be successfully managed by user associations in small communities if eight “design principles” are met, including the ability to effectively exclude unentitled parties. Such conditions clearly do not apply to global common resources. As Lee (1990, 317) notes, “Each birth inflicts costs on all others by reducing the value of their environmental birthright”.2

The juxtaposition of these scientific, executive, and political challenges places high demands on the ability of national and global institutions to respond to these challenges, as evidenced by the slow progress made in decades of efforts to regulate carbon emissions. The original projections of the Limits to Growth study for the 1970–2000 period correspond broadly with the observed trends during this period (Turner 2008, 2012).

Managing Climate Change: Addressing Per-Capita Emissions and Population

Models of climate change take population size into account, but they typically treat it as a given (for example, Stern 2006; Nordhaus 2008, 2012). They tend to use
the United Nations medium variant population projections. Using this approach, the World Bank (2010, Figure 3.5) estimates the impact of climate change on the growth in agricultural productivity required to meet the world’s rising food demand. The model incorporates projected rises in food demand due to growth in incomes as well as in population size and shows how much more difficult it will be to meet that demand given anticipated climate change. What is needed is a huge increase in agricultural productivity backed by greatly intensified regulation to protect natural systems.

However, as we discuss below, population size is amenable to policy, and it makes a significant difference to the size of adjustments required on other fronts. Models vary, but the World Bank (2010) estimates that to meet the growing demand for food between 2005 and 2055, agricultural productivity will need to rise by 64% under the assumptions of the “business-as-usual” scenario and by a further 80% to offset the projected stresses arising from climate change (Figure 1). However, the model indicates that if population remained constant at the 2005 level, agricultural productivity would need to rise only 25% under the “business-as-usual” scenario; that is, more of the required productivity increase under the “business-as-usual” scenario is necessitated by population growth than by increases in consumption per capita.

The developed countries’ carbon emissions per capita are far higher than those of the developing countries, but the latter account for nearly all of the projected increase in emissions between now and 2050 (Stern 2006, Figure 7.3). Although emission rates in the developed countries seem to have peaked, they are growing rapidly in the developing countries due to both economic growth and population growth. Although China has had the steepest growth in carbon emissions with its high rate of economic growth, its estimated total emissions in 2008 were similar to those of other developing countries as a group (excluding India), partly because the latter had twice the population of China (UN 2013). GDP per capita is rising rapidly across developing countries, including sub-Saharan Africa in the 2000s (IMF 2010, 2011). Nearly all of the projected global population growth will occur in developing countries, whose population (excluding China and India) is projected to grow 2.7-fold between 2000 and 2100, driven largely by the six-fold increase projected for sub-Saharan Africa (UN 2013, medium variant). It is estimated that the effect of a 40% reduction in CO₂ emissions per capita in developed countries between 2000 and 2050 would be entirely offset by the increase in emissions attributable to expected population growth in poorer countries over this period, even if we assume no change in emissions per capita in these countries (Dyson 2005).

Managing global warming may require different policy approaches in different settings. Imposing a carbon tax is strongly recommended as the simplest way to reduce carbon emissions (Stern 2006; Nordhaus 2008, 2012). By putting a price on carbon emissions, such a tax creates incentives to conserve global common
property resources while providing incentives and fiscal resources for developing cleaner technologies. However, some major polluters among the developed countries have shown a limited political appetite for this, and developing countries argue that such taxes will unfairly constrain their economic growth, with restrictions that

Source: World Bank (2010) World Development Report 2010: Figure 3.5 (derived from Lotze-Campen et al. 2009). We thank Dr. Lotze-Campen for disaggregating the “business-as-usual” scenario into two estimates: (1) with population held constant at the 2005 level and (2) the WDR 2010’s “business-as-usual” scenario, which includes anticipated population increase to 9 billion by 2055.

Note: The original explanatory note said it was the required annual growth. Dr. Lotze-Campen corrected this by deleting the word “annual”.

Explanatory note from the original figure in WDR 2010: “The figure shows the required growth in an agricultural productivity index under two scenarios. In this index, 100 indicates productivity in 2005. The projections include all major food and feed crops. The straight line represents a scenario without climate change of global population increasing to 9 billion in 2055: total calorie consumption per capita and the dietary share of animal calories increasing in proportion to rising per capita income from economic growth; further trade liberalization (doubling the share of agricultural trade in total production over the next 50 years); cropland continuing to grow at historical rates of 0.8 percent a year; and no climate change impacts. The dotted line represents a scenario of climate change impacts and associated societal responses (IPCC SRES A2): no CO2 fertilization, and agricultural trade reduced to 1995 levels (about 7 percent of total production) on the assumption that climate change-related price volatility triggers protectionism and that mitigation policy curbs the expansion of cropland (because of forest conservation activities) and increases demand for bioenergy (reaching 100 EJ [10^{18} joules] globally in 2055)”. 

Das Gupta
the currently developed countries did not have to face when they were industrializing rapidly.

For countries that still have high levels of fertility, which typically are still poor and have very low per capita emission rates, the key approach seems to be to reduce fertility. Clearly, they do not owe it to the world to reduce fertility to help slow the pace of climate change. However, they have strong incentives of their own to reduce fertility, as discussed below.

The Unequal Impact of Climate Change

The developed countries are responsible for much of the accumulation of emissions that affect climate change because they began rapid industrialization by the end of the nineteenth century. However, in a twist of fate, the impact of climate change will be felt most sharply in the developing countries. Many of these countries lack the financial resources for adaptation/mitigation efforts, and for some, the capacity to act may also be hindered by poor governance (WDR 2008, 245).

“The impacts of climate change are not evenly distributed - the poorest countries and people will suffer earliest and most... First, developing regions are at a geographic disadvantage: they are already warmer, on average, than developed regions, and they also suffer from high rainfall variability... Second, developing countries - in particular the poorest - are heavily dependent on agriculture, the most climate-sensitive of all economic sectors, and suffer from inadequate health provision and low-quality public services. Third, their low incomes and vulnerabilities make adaptation to climate change particularly difficult... At a national level, climate change will cut revenues and raise spending needs, worsening public finances.” (Stern 2006, vii)

Modeling the effects of climate change is subject to considerable uncertainty, but there is consensus on its broad effects, some of which are summarized below.

(a) **Food and Water**: Global warming will reduce crop outputs at lower latitudes, undermining food security in the developing countries. If temperatures rise further, global food output will decrease, reducing the developing world’s access to imported food.

Approximately one-third of the world’s population lives in countries with moderate to high water stress (Stern 2006, 63), often exacerbated by poor management of water resources (World Bank 2007, 183). Rising demand for agriculture and other purposes will heighten water scarcity.

Climate change is also expected to disrupt rainfall patterns, threatening agricultural cycles and human lives with droughts and floods. These changes will
impact the poorest billion people in the world most heavily because 75% of these people live in rural areas and rely on agriculture for their livelihood (Stern 2006, 67).

(b) **Health and Human Capital:** Climate change is increasing morbidity and mortality from vector-borne and diarrheal diseases, as well as malnutrition, and children are the most affected (McMichael et al. 2004). This situation can have lasting consequences for human capital. A study in Zimbabwe found that young children who became stunted as a result of a drought faced long-term negative effects in school attainment and subsequent earnings (Alderman et al. 2006).

(c) **Natural Disasters:** The frequency and severity of natural disasters is expected to increase, affecting and even displacing large numbers of people. Low-lying coastal areas will become increasingly uninhabitable and subject to flooding and hurricanes. Sea-level rise will bring salinization, salt-water intrusion in groundwater aquifers, and, in some areas, complete inundation (WDR 2008, 200).

Many countries in the most affected regions have poor preventive health systems, with a low capacity for averting and controlling disease outbreaks even during routine conditions. This capacity becomes especially critical in the face of natural disasters, with their attendant health threats (Das Gupta et al. 2009).

(d) **Conflicts:** The pace of internal and international migration will rise with the combined pressures of climate change, population growth, and environmental degradation (Laczko and Aghazarm 2009; World Bank 2010; Gememne 2011), and migrants may not always be welcomed by people who may themselves feel under pressure. Migration from Bangladesh into parts of Northeast India has led to low-level conflict for decades, which could be exacerbated if the densely populated megadeltas of the Bay of Bengal are inundated by sea-level rise. Land degradation and drought have already caused considerable movement of people in sub-Saharan Africa. Mamdani (2001) notes that one of the factors underlying the Rwanda genocide was local resentment of the heavy in-migration of people seeking richer land.

**How Do The Poorest Countries Gain From Fertility Reduction?**

Fertility remains high in several developing countries, typically in the poorest ones. The least developed countries have an estimated average of 4.2 children per woman in 2010–5 (UN 2013, medium variant). The estimate for Sub-Saharan Africa is 5.1 children per woman in 2010–5, and the region’s population is projected to rise from 0.64 to 3.82 billion between 2000 and 2100 (UN 2013, medium variant). Total fertility rates also remain high in a scattering of other developing countries.
They remain above 3 children per woman in some larger Asian countries, such as the Philippines and Pakistan (UN 2013), and some of the least developed states of Northern India (Haub 2011, Figure 11).

Reducing fertility can benefit these countries in many ways, facilitating economic growth and poverty reduction. A large body of literature since the 1990s has discussed the “demographic dividend” that is enabled when fertility declines in high fertility settings. The resultant low dependency ratios create a window of opportunity for savings, increased productivity, and investment (Higgins and Williamson 1997; Kelley and Schmidt 1996, 2005). Some of this dividend is automatic, arising simply from increasing the resources per capita for services, infrastructure, and livelihoods. However, with good policy management and investment in physical and human capital, this window of opportunity can be used to transform economies such that their growth potential remains high after the window has closed. This is evidenced especially in East Asia (Bloom and Williamson 1998; Lee 2009). The more rapid a region’s fertility decline, the wider the window of opportunity, although its duration will be shorter because the population will age more rapidly.7

This literature on the “demographic dividend” is sometimes interpreted as implying that fertility decline is "wasted" without strong policy settings such as those in East Asia. Yet, these studies emerged decades after vigorous family-planning programs were started in most Asian countries in the 1960s and 1970s. These programs were explicitly motivated by widespread poverty compounded by sharply rising population growth rates and were viewed as an integral part of the countries’ development strategy.8 Reducing fertility helps to reduce poverty, as evidenced in India, where it mitigates some of the negative fallout of weak economic policies and slow job growth.

Micro-studies find that lower fertility helps to reduce poverty at the household level in developing countries. It has been found to be associated with better child health and schooling (Rosenzweig and Wolpin 1980; Rosenzweig and Zhang 2009), improved maternal health, increased women’s labor force participation, and higher household earnings (Joshi and Schultz 2007). Young women have benefited especially from access to the family-planning program in Colombia, obtaining more schooling and increasing their likelihood of working in the formal sector (Miller 2010). Similar results have emerged from studies conducted in developed countries, as discussed below. Miller (2010, 711) concluded that family planning may be “among the most effective (and cost-effective) interventions to foster human capital accumulation”.9

These benefits are especially critical given the shortage of land and jobs in these countries, which leaves their growing populations ever more squeezed for livelihoods. Land scarcity is acute in most Asian countries, and in sub-Saharan Africa available cropland per agricultural person decreased by 40% between 1960 and 2003 (World Bank 2007, 63). Although some sub-Saharan African countries have
considerable room for land expansion, high rural population growth drives expansion into forest or grazing land. Large investments in infrastructure, disease control, and soil management are needed to convert these lands to productive agriculture (World Bank 2007, 63). Food production per capita changed little in sub-Saharan Africa between 1961 and 2005 (Figure 2).

There is also a shortage of jobs. Levels of unemployment are already high in many countries, but the World Bank (2012, 51) estimates that substantial job creation is required just to maintain the 2005 levels of employment of the working age population in 2020. For example, an additional million jobs a month will need to be generated in South Asia. Given the slow pace of job growth in this region, it is fortunate that the population aged 0-14 is projected to decline soon (UN 2013, medium variant), as also observed in India (Figure 3). The report also estimates that the number of jobs in sub-Saharan Africa would have to increase by about 50%, which translates into employment growth of 2.7% a year. Meanwhile, the population aged 0-14 is growing rapidly in sub-Saharan Africa, and the numbers of people entering working age will continue to rise sharply for decades (Figure 3). The rapid projected growth of the young population in the least developed countries contrasts sharply with that of other developing countries (Figure 3), so countries with weak economic growth face the highest increase in numbers entering the labor force.

Population growth imposes a direct burden of resource depletion upon developing countries. Arrow et al. (2004, 164–5) estimate that the rate of depletion of

---

Figure 2. Changes in Food Production Per Capita, 1961–2005

Source: The Royal Society 2009: Figure 1.4.
“genuine wealth per capita” in sub-Saharan Africa during 1970–2000 was such that it would be halved about every 25 years. Kelley and Schmidt (2005) conclude that sub-Saharan Africa has benefited far less than have other regions from the impact of reduced dependency ratios on output growth per capita, because of its high fertility.

Lower fertility can also help the poorest countries’ efforts to mitigate the effects of climate change so that the shocks affect fewer people, and more resources per capita are available for coping with them. These resources can be used for adaptation measures, such as efforts to slow the decline in food production. Systems for disaster management and preventive health services can be strengthened to minimize the spread of existing and emerging diseases. Such measures will make it easier to coordinate collective efforts to cope with climate change.

With less pressure on livelihoods, poor households will also be better positioned to cope with the consequences of climate change. Looking to the future, slower
population growth will reduce these countries’ projected contribution to future climate change, which will, as before, have the most devastating impact on these countries.

**Can Family-Planning Programs Help Lower Fertility?**

Government programs to promote the use of effective contraceptive methods are by no means a necessary condition of fertility decline; birth rates fell in Europe with no state encouragement. Nor are they the only policy levers to encourage lower fertility. Female education has been found to be associated with a higher age at first birth and lower fertility in settings as varied as Guatemala, Indonesia, and Nigeria. The key question is whether family-planning programs can advance the timing of reproductive change and accelerate it once it is underway.

In a highly influential paper, Pritchett (1994a) argued that family-planning programs have little impact on fertility: “Ninety percent of the differences across countries in total fertility rates are accounted for solely by differences in women’s reported desired fertility. . . . The results contradict theories that assert a large causal role for expansion of contraceptive use in the reductions of fertility”. Many have taken Pritchett’s study to indicate that effort on family-planning programs is ill spent. However, in a subsequent paper he concludes that his estimates imply that strengthening a family-planning program substantially (by 50 points on of a scale of 0-100) would reduce fertility by one birth (Pritchett 1994b, 626). Bongaarts (1997) estimates the corresponding fertility reduction at 1.4 births, but even Pritchett’s lower estimate amounts to a very large difference in population size and in the momentum of population growth. If one less birth per woman were sustained in sub-Saharan Africa through 2100, the region would have an estimated 2.6 billion fewer people. This would more than halve the estimated total population at the end of this century (Figure 4).

A crucial gap in Pritchett’s argument is that he assumes that family planning programs work only on the supply side and overlooks their important role in reducing desired family size. He conducts cross-country regressions of total fertility rates against contraceptive prevalence and against family-planning efforts, but in both cases, he controls for desired fertility (Pritchett 1994a, Table 3). However, mass-media outreach to reduce desired family size is a major component of family-planning programs. Studies have shown that the mass media are very effective at increasing contraceptive use and reducing fertility (see below).

**What Do Family-Planning Programs Seek to Do?**

Family-planning programs seek to boost the use of contraceptive methods by expanding their supply and accessibility and disseminating information on the methods available, thus enabling couples to postpone or limit childbearing. This is
especially important for the poor, who typically have higher numbers of unwanted children than the rich, except in settings with very effective programs, such as Indonesia (Figure 5).

Family-planning programs also typically seek to reduce desired family size by disseminating information on new opportunities for altering living standards through new strategies for bearing and investing in children. Parents – especially poorer parents – have imperfect information on these issues. Households also appear to face difficulties in making optimal choices that involve long-term planning horizons (see, for example, Cronqvist and Thaler 2004 on pension decisions). People’s access to information can be improved by offering simple messages through mass communication or more complex messages through radio or television soap operas that portray the lives of people with small families and how they access new opportunities helps reduce imperfect information.

Media outreach has been found to be effective at increasing contraceptive use and lowering fertility. This has been found in many studies using cross-sectional survey data on access to media (e.g., Bhat 1998). The few quasi-randomized evaluations of media outreach have found it effective at altering fertility and contraceptive

---

**Figure 4.** Population Projections for Sub-Saharan Africa Indicate that Maintaining One Less Birth per Woman Halves the Estimated Population in 2100

![Population Projections Graph](image)

*Source: United Nations (2013).*

*The UN “high” and “low” projections diverge during 2010-2020, reaching a difference of 1 birth by 2020.*
Use in Tanzania (Rogers et al. 1999) and reducing fertility in Brazil and India (La Ferrara et al. 2008; Jensen and Oster 2009).

To motivate their evaluation of the impact of Brazilian soap operas on fertility, La Ferrara et al. (2008, 9) report the results of an experimental focus-group discussion in which adult women of middle- and lower-class backgrounds were asked to portray the families that are frequently displayed on television as well as those of common people. “The results were clear: television families are small, rich and happy; the families portrayed as common people are poor, contain more children and the faces reveal unhappiness. . . .constant exposure to smaller, less- burdened television families may have created a preference for fewer children and greater sensitivity to the opportunity costs of raising children”.

This is exactly the approach used in many countries, such as South Korea and India, using simple billboards. Their family-planning programs catalyzed demand through media blitzes that conveyed images of glowing parents with one or two flourishing children, sometimes juxtaposed with images of overwhelmed parents surrounded by many children living in much poorer conditions. Brief jingles on the radio and television reinforced the message that “a small family is a happy family”.

Such media blitzes are especially important in settings where contraceptive use is not yet commonplace. By reaching entire communities, they help to change social norms and reduce barriers to the use of contraceptives. One barrier may be that women are more motivated than men to control childbearing. For example, a study

\[\text{Figure 5. Unwanted Fertility is Higher among the Poor, and Effective Family-Planning Programs Can Reduce this Gap}\]

Source: Gillespie et al. (2007), Table 1.
in Zambia found that women who were given contraceptive information and access without their husbands present were more likely to use contraception and less likely to give birth than a control group of women accompanied by their husbands (Ashraf et al. 2012). Another such barrier is suggested by a study in urban slums in Pakistan, which found that mothers-in-law influenced contraceptive decision-making (Fikree et al. 2001). Studies in several countries show that women resort to the covert use of contraception, when their use is not generally accepted. By helping shift social norms, media outreach helps empower women to use contraception.

Evaluations of Family-Planning Programs

Evaluating the impact of family-planning programs is challenging because these programs are rarely randomly placed and uniformly executed. However, studies using very different analytical approaches, including natural experiments, indicate that family-planning programs do affect fertility.

Schultz (2009, 4) notes that several careful evaluations of family-planning programs find a negative association between “the regional intensity of program treatment and the regional level of fertility” in a country. These include studies of programs in Taiwan (Schultz 1973, 1992), Colombia (Rosenzweig and Schultz 1982), and Indonesia (Molyneaux and Gertler 1994). Some studies are simple cross-sectional analyses, but others have gone further to analyze panel data and include fixed effects for regions and time. However, the estimated program impact may be biased by nonrandom placement.

Several studies use natural experiments or quasi-randomized trials. In the Matlab program in Bangladesh, half of the villages studied for the 1974–96 period received more intensive family planning and maternal and child health program inputs from 1977–8, whereas the other half received regular government program inputs. Note that the country was poor and largely illiterate for much of the study period. The first set of villages showed more rapid fertility decline after the program began and maintained 15% lower fertility in 1982–96 (Joshi and Schultz 2007, 30). Sinha (2005) found that 18 years after the Matlab program began, it accounted for a 14% decline in lifetime fertility (0.6 fewer births per woman) compared with women in the second set of villages. This difference is especially striking given that fertility was falling rapidly across the country. If sustained over time, this difference in fertility can considerably reduce the momentum of population growth, as the difference between the UN projection variants show (Figure 4).

Miller (2010) evaluated Colombia’s family-planning program, exploiting differences in the timing of the introduction of the family-planning program to estimate the impact of contraceptive availability on fertility. The program was found to have lowered fertility by about 10%, despite the fact that fertility was declining rapidly.
across the country. Households with lower fertility also showed improvements in schooling, health, and earnings. Miller noted that access to family planning helped young women obtain more schooling and increased their chances of working in the formal sector.

These evaluations may tend to underestimate the impact of family-planning programs insofar as their measures of program effort are more likely to pick up variation on the supply side. Mass communication efforts to reduce desired family size are likely to reach people regardless of whether they live in areas with higher or lower supply-side program effort.

Some recent studies have used natural experiments that were created by policy-related variations in access to family planning to examine the impact of access to contraceptives. These studies indicate that facilitating access to contraception is highly pro-poor, as indicated by Figure 5. Two studies examine the impact of shifts in the application of the United States’ “gag rule” (Mexico City Policy), which restricts foreign aid for family planning to any organization that may provide abortions using other funds. Jones (2011) estimated that the policy was associated with a 12% increase in pregnancies among rural women in Ghana, increasing both abortions and unintended births. The unintended births were concentrated among the poorest and least educated women, and these children had significantly lower height-for-age relative to their siblings. Bendavid et al. (2011) found that after the Mexico City Policy was reinstated in 2001, abortion rates rose in sub-Saharan African countries that receive high levels of foreign assistance from the United States for family planning and reproductive health. Salas (2013) found that policy-related disruptions in the public supply of free contraceptives in the Philippines was associated with elevated birth rates, especially among poor, less educated, and rural women.

Similar findings emerge from the analyses of natural experiments in the developed world. Kearney and Levine (2009) examined the impact of state-level Medicaid policy changes in the United States that expanded eligibility for family-planning services and found that it reduced births, particularly for teenagers and those with lower educational attainment. Bailey (2012) estimated that federally funded family planning in the United States reduced childbearing among poor women by 19% to 30% between 1964 and 1973. Reflecting the findings in Bangladesh and Colombia discussed above, analyses of natural experiments in the United States and Sweden have found significant female labor supply responses to differences in the provision of the birth control pill (Goldin and Katz 2002; Ragan 2013).

Jones’ (2011) finding that unintended children in Ghana were more likely to be stunted than their siblings is consistent with other studies that indicate that greater investments are made in planned children. For example, Do and Phung (2010) used the fact that in Vietnam, some years in the animal zodiac cycle are considered
especially auspicious to bear children. They found that larger cohorts of children are born in auspicious years and that these children have higher schooling attainment. They concluded that this is because parents are more likely to invest in planned children. Analyses of natural experiments in the United States and Sweden indicate that increasing women’s ability to plan their births was associated with substantial improvements in their children’s education and earning capacity (Rotz 2013, Madestam and Simeonova 2013).

Are Family-Planning Programs Likely to Work in the Poorest Countries?

The experience of countries such as the Republic of Korea in the 1960s and others such as Indonesia, Bangladesh, or Nepal shows that sustained fertility decline can occur in poor countries given political commitment to family-planning programs. This commitment is driven by poverty and sharply rising population growth rates. For example, India’s censuses showed decadal growth of 11%-14% from the 1920s, but this jumped to 22% in 1951–61 and 25% during 1961–71. Similar population growth rates were observed in the Asian region and sub-Saharan Africa (UN 2013). However, sub-Saharan Africa’s small population base in the 1950s meant low increases in numbers, a situation that is changing very quickly (Figure 4).

Political commitment to family planning has sometimes been low in many sub-Saharan African countries, which may have contributed to their slow fertility decline (Cleland et al. 2006, 2011; Bongaarts 2006; Machiyama 2010). However, this situation can change quickly, as evidenced by the success of the Rwandan government’s concerted push since the mid-2000s to reduce fertility. Until then, both Rwanda and its neighbor, Burundi, were poor, densely populated countries with high fertility and weak family-planning programs. Then, in contrast with Burundi, Rwandan government officials spoke out about the need to reduce fertility. A country-wide information dissemination program was implemented, along with sharply increased access to contraceptive methods. Between the DHS surveys of 2005 and 2010, the total fertility rate fell from 6.1 to 4.6 children per woman, and the use of modern methods of contraception among married women rose from 10% to 45% (National Institute of Statistics of Rwanda 2011). Meanwhile, total fertility in Burundi was 6.4 in 2010 (Institut de Statistiques et d’Études Économiques du Burundi 2010).

Most countries in sub-Saharan Africa show some fertility decline, indicating a desire to lower fertility. Family-planning programs can build on this desire and accelerate fertility decline. These programs are likely to be most effective when accompanied by other measures addressing basic government failures that help sustain poverty and high fertility, including efforts to improve health and schooling and to expand income-earning opportunities. Family-planning programs help by increasing access to contraception and by providing informational outreach to accelerate
perception of the benefits of shifting to a more secure equilibrium in which people have fewer children and are able to invest more in them.

Conclusions

The management of environmental common property resources is complex because these resources are unpriced, so people must agree to self-impose a price for using them. Their over-exploitation by countries that began industrializing early has led to global warming. The fallout of this will, in an ironic twist of fate, fall primarily on the developing countries, many of which are still poor and have low per-capita emissions. These countries will be the first to experience declines in agricultural output, poorer health outcomes, disruption of rainfall patterns, and more frequent and severe natural disasters, which render some areas uninhabitable. These changes make it harder for poor people to emerge from poverty, and they push others into poverty.

However, although per-capita emissions in developed countries remain much higher than those in developing countries, their growth seems to have peaked. Most of the projected growth in emissions derives from the developing countries, due to their economic growth and their population growth. Most future population growth is projected to take place in these countries, with the highest growth rates in sub-Saharan Africa and the least developed countries.

In this highly complex situation, analysts have focused on policies to reduce greenhouse gas emissions. A carbon tax is proposed as the simplest approach to reduce carbon emissions. By putting a price on carbon emissions, such a tax creates incentives to conserve their use while providing incentives and fiscal resources for developing cleaner technologies. Some have argued that pricing carbon use could be introduced more gradually in developing countries to impose less constraint on their potential economic growth.

Neglected in these policy debates is the fact that a substantial part of future growth in emissions derives from population growth, mostly in the poorest countries. Although population size is incorporated into models of climate change, the population projections are taken as a given. However, fertility is highly amenable to policy intervention. For countries that still have high fertility and that typically have very low per-capita emission levels at present because they are still poor, the more immediate approach might be to lower fertility.

Clearly, the poorest countries cannot be expected to reduce fertility to help the world as a whole, especially when they are suffering from the excesses perpetrated by the rest of the world. However, they have much to gain from lowering fertility. It will increase their available resources per capita, enabling them to invest more in the human and physical capital needed for economic growth. It will also increase
per-capita resources to strengthen systems for disaster management and for disease prevention and control, helping them to cope more effectively with the multiplicity of stresses associated with climate change. It will reduce growth in the demand for jobs and livelihoods. Moreover, studies from both developing and developed countries show that access to birth planning helps reduce poverty through increased female labor force participation and better child schooling and health outcomes. Furthermore, fertility decline in poor countries yields a substantial “demographic dividend” in reducing poverty and vulnerability, even without the large additional gains that can be obtained with strong economic policies.

Lower fertility will also benefit the poorest countries by reducing the pace of future global warming, the negative effects of which affect them far more than developed countries. The latter countries mostly lie at geographically higher latitudes that are less negatively affected by climate change. Furthermore, the developed countries have far greater resources to cope with climate change.

The means of lowering fertility are well documented. Family-planning programs help by increasing access to contraception and by catalyzing demand for contraception through media blitzes. Studies show that family-planning programs are effective at helping to lower fertility, and highly pro-poor in their impact. Easier access to family-planning most benefits women who are poor and uneducated. Family-planning programs are a simple, effective, and relatively inexpensive way to achieve a multiplicity of benefits for poor countries.

While the rest of the world wrestles with the political and technological complexities of reducing emissions, family-planning programs offer the poorest countries a simple and effective means to reduce poverty and mitigate the impact of climate change.

Notes

Research Professor, Department of Sociology, University of Maryland, College Park, MD and Visiting Fellow, Population Reference Bureau, Washington, DC. Email: mdasgupta@gmail.com. This work was supported by the William and Flora Hewlett Foundation [Trust Fund TF070424 to the World Bank and grant number 2012-7611 to the Population Reference Bureau]. Comments from Emmanuel Jimenez and an anonymous reviewer were very helpful for revising the draft, all errors remain mine.

1. Simon and Boserup both argued that higher population densities can increase the economies of scale in providing productivity-enhancing infrastructure and services, such as transport and extension services (Glover and Simon 1975; Boserup 1981).
2. See also Lee and Miller (1990).
4. Many have estimated that slowing population growth could substantially reduce carbon emissions (see, for example, Meadows et al. 1972; Bongaarts 1992; and O’Neill et al. 2010).
6. For example, PAHO has estimated that the incidence of dengue, another vector-borne disease, has risen in the temperate as well as the tropical zones of the Americas.

7. Other macro-studies indicate that rapid population growth can constrain economic growth (Galor and Weil 2000; Weil and Wilde 2009) and reduce growth in income per capita (Acemoglu and Johnson 2007). For reviews of studies of the relationship between population and economic growth, see Johnson and Lee (1986); Kelly (1988); and Das Gupta et al. (2011).


9. Some studies in the developed world also find high fertility is negatively associated with child schooling and female labor-force participation (Black et al. 2005; Caceres-Delpiano 2006; Angrist and Evans 1998; Conley and Glauber 2006). Other studies do not find evidence of a quantity-quality trade-off in childbearing (Angrist et al. 2010).


11. Zimbabwe offers an example of rapid fertility decline with strong political will, but it was not a poor country at the time.

References


Das Gupta


An orderly sovereign debt restructuring should place the debtor nation’s public debt on a sustainable trajectory while minimizing procrastination and contagion. However, the experiences with the debt crisis of the 1980s, Russia 1998, Argentina 2001, and Greece 2010 indicate that orderly debt restructurings remain elusive, even with high-powered official intervention. When solvency problems are present, the chances of success increase if official money is lent at the risk-free rate, reflecting its low risk, and if private creditors receive an upfront haircut. The paper examines the obstacles, which include moral hazard, difficulty in distinguishing between solvency and liquidity crises, and the “political economy” resistance to upfront haircuts. Orderly sovereign debt restructurings are likely to remain elusive notwithstanding recent evidence that the official mindset may be changing. Sovereign Debt, Debt Restructuring, Solvency, Liquidity, Seniority, Sector Board, Economic Policy (EPOL). JEL codes: E61, E65, F34

Introduction

This much we know from the experience of emerging market countries from the 1980s onwards: except for a few small cases, sovereign debt restructurings have tended to be costly and chaotic, with orderly sovereign debt restructurings seemingly impossible to achieve. This holds even when high-profile official intervention occurs. We explore why this happens and identify the impediments to an orderly debt restructuring (ODR).

The finding from emerging markets (EMs) about ODRs being elusive carries over to Greece 2010. In March 2010, it became clear that Greece’s fiscal fundamentals
were weak and that it would need official assistance to avoid a default. Discussions on a bailout began. A counterintuitive feature was that market sentiment worsened as negotiations proceeded between Greece and the European Union-European Central Bank-IMF troika. The spread on Greece’s bonds rose significantly, though the size of the bailout package was upped substantially. Remarkably, the two-year bond spread shot up, though Greece was effectively being “taken out of the market”, that is, the announced bailout funds were more than enough to pay off maturing short-term private creditors in full. This situation suggested deep-seated market skepticism about solvency and the feasibility of the fiscal program accompanying the bailout.\(^1\)

Notwithstanding these negative market signals about solvency, the authorities made it abundantly clear that any haircut for Greece’s private creditors was out of the question. The troika’s gamble was that structural and fiscal reforms would restore Greece to a sustainable debt path that would lower interest rates to non-default levels without a debt write down, which was believed would have costly contagion effects.

Barely a year later, in July 2011, the official position reversed dramatically. By that time, contagion from Greece, Ireland, and Portugal (the latter two countries had also received official bailouts by then) had begun to spread to the core of the European Union (EU). A July 21, 2011 Eurozone summit announced support for a haircut for Greece’s private creditors while also agreeing to a major softening of loan terms to bring official EU lending rates closer to the risk-free rate while lengthening maturities significantly. A subsequent summit in October 2011 announced that private Greek bondholders would receive a 50 percent write down on principal and that the European Financial Stability Facility (EFSF) would be leveraged to €1 trillion to support Italy and Spain. Stock markets reacted euphorically, but Greece announced and then withdrew a referendum on the bailout toward the end of 2011, while Italian 10-year bond yields approached the 7 percent threshold at which other countries had been bailed out. The ECB saved the day by injecting liquidity via two tranches of a Long-Term Refinancing Operation (LTRO). This lent commercial banks huge sums at 1 percent for 3 years, which they used to buy Spanish and Italian sovereign bonds, lowering their yields substantially and boosting confidence.\(^2\)

In February 2012, the EU approved a second bailout for Greece amounting to €130 billion but required a PSI (private sector involvement) debt exchange, which inflicted losses of some 70 percent in net present value (NPV) terms on €197 billion in privately-held debt, equivalent to approximately 97 percent of the projected 2012 Greek GDP. Even so, the government debt-to-GDP ratio under the program assumptions at the time was expected to fall to only 120 percent by 2020. The subsequent bond price movements indicated that the debt deal had done little to alter market perceptions about Greece’s credit standing. Figure 1 plots the 10-year Greek bond price
and its spread from January 1, 2010 to the end of May 2012, noting key events. Subsequent developments in the Eurozone are summarized in Section 3.

As the Greek crisis indicates, sovereign debt restructuring is complicated, official intervention notwithstanding. Does this mean that official intervention does more harm than good? It is difficult to answer this question conclusively because of the difficulty in developing a counterfactual. The major lesson from the debt overhang of the 1980s (Krugman 1988; Sachs 1986) is that such intervention is needed to solve coordination and free-rider problems among creditors.

Apart from the inherent complexity in sovereign debt restructuring, two reasons may explain why official intervention (OI) tends not to work well in prominent sovereign debt crises, such as the debt crisis of the 1980s in Latin America, Russia in 1998, Argentina in 2001, or Greece in 2010. The first is a seeming inability to distinguish between liquidity and solvency crises, that is, situations where a sovereign may be unable to rollover maturing debt versus one where the debt has simply become too large to be serviced. Although the catalytic effect of official finance may
work well in persuading short-term creditors to roll over their loans in countries with acceptable fundamentals (as in Morris and Shin 2006), these elegant results tend to break down once one acknowledges that official loans may be senior to private loans and that the country is facing a solvency instead of a liquidity problem (Kharas, Pinto, and Ulatov 2001; Chamley and Pinto 2011). The second reason is legal impediments to a smooth bankruptcy process for sovereigns. Our focus will be on the basic economics of and political obstacles to an ODR.

Section 2 presents the motivation for ODRs based on a survey of EM sovereign debt restructuring and the part played by official intervention. This is followed by a discussion of procrastination in sovereign debt restructuring in Section 3. Section 4 builds on Sections 2 and 3 to tease out the desirable attributes of an ODR. However, the track record inevitably raises a question about the feasibility of an ODR. Hence, Section 5 discusses the obstacles, which include political economy and the difficulty in distinguishing between liquidity and solvency problems for countries. Section 6 concludes.

Context and Motivation for ODRs

The sovereign debt literature concerns itself with fundamental questions, such as why sovereign debt exists in the first place, considering difficulties in enforcing contracts; why default by a sovereign does not mean permanent exclusion from future borrowing; and why countercyclical fiscal policy (saving during good times, depleting accumulated saving during bad times) and self-insurance against shocks cannot substitute for borrowing. An excellent survey is presented in chapter 2 of Sturzenegger and Zettelmeyer (2006), which also contains a concise account of the seminal papers.

Our goal is different. We want to review the empirical experience with sovereign debt restructuring since the landmark EM debt crisis of 1980s and use this as a forward-looking platform for discussing the desirable attributes of an ODR. The first point to note is that the vast bulk of EM sovereign debt restructurings since the 1980s have involved private creditors (Table 1). Some US$325 billion in principal owed to private creditors has been restructured, compared to just US$29 billion with official creditors via the Paris Club.

In contrast, official (including bilateral and multilateral, such as IMF, World Bank, International Development Association (IDA), African Development Fund) creditors have accounted for the lion’s share of sovereign debt restructurings for low-income countries, which typically have limited access to the international capital markets. As of February 15, 2012, the Paris Club had treated debt amounting to US$556 billion for 88 developing countries under 423 agreements. Multilateral creditors have provided debt relief through the Heavily Indebted Poor
Countries (HIPC) Initiative and the Multilateral Debt Relief Initiative (MDRI). However, this initiative is available only to low-income countries, and eligibility criteria are restrictive. Given the eligibility requirements for HIPC and MDRI, none of the EMs has benefitted from multilateral debt restructurings.

The rest of the review considers the origins of debt crises and goes on to the role of official intervention. The 1980s experience showed that it is difficult to design efficient official intervention. This was confirmed by the subsequent experience with Russia in 1998 and Argentina in 2001, which illuminated another important issue: why official intervention may not be catalytic in terms of persuading private creditors to roll over their loans.

Origins of Debt Crises

One set of constants has marked all serious EM debt crises since the 1980s: fixed exchange rates, open capital accounts, weak growth prospects, and concerns about fiscal solvency. Fiscal fundamentals play a crucial role, either at the outset or eventually, as a result of bailing out the domestic private sector. In addition, though the crisis itself typically involves an abrupt economic disruption, its seeds tend to be sown over long periods, reflecting policy and political economy.

Heavy external borrowing preceded the 1980s debt crisis. Such borrowing may have been motivated by the need to finance development, sometimes via ill-advised public investments; by social spending needs; and even by the desire to enrich well-connected groups. Money-center banks were happy to roll over maturing principal and even interest payments because the key creditworthiness indicator at that time was the external debt-to-exports ratio, and nominal export prices in dollars continued to rise faster than the nominal interest rate, keeping this ratio under control.

<table>
<thead>
<tr>
<th>Plan/Country</th>
<th>Amount restructured (in US$ billion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brady Plan (1989)</td>
<td>200</td>
</tr>
<tr>
<td>Russia London Club (2000)</td>
<td>32</td>
</tr>
<tr>
<td>Argentina (2005 &amp; 2010)</td>
<td>76</td>
</tr>
<tr>
<td>Ukraine (2000)</td>
<td>2.3</td>
</tr>
<tr>
<td>Uruguay (2003)</td>
<td>5.1</td>
</tr>
<tr>
<td>Others</td>
<td>7.2</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>322.6</strong></td>
</tr>
</tbody>
</table>


Notes: For the 1980s, the 1985 Baker Plan is not included because the restructured debt amounts are subsumed under the Brady Plan. The Russian and Argentine pre-crisis swaps are not included but are discussed below. US$6 billion of defaulted debt owed to Argentina’s private creditors is still unresolved.
Sachs notes (1990, p 8), “During the heady days of the 1970s . . . countries and their banks had the illusion of an unending Ponzi game . . . ”. Eventually, with their terms-of-trade declining sharply in the early 1980s along with the record rise in interest rates in the US—a combination we shall refer to as the “twin shocks”—the bubble burst, and countries now had to service their debt the old-fashioned way: by generating current account surpluses to pay down their debt. This meant politically unpalatable fiscal austerity and cuts in real wages.

Three complications frequently arose. First, with fixed pegs to the dollar the norm, the private sector began speculating against their home currencies once they realized that the exchange rate was becoming overvalued. This led the government and central bank to borrow overseas in support of the peg. The acceleration of private capital flight exacerbated the eventual public debt burden while exerting ruinous effects on domestic banks and the financial system.

Second, some central banks imposed restrictions on convertibility in an effort to prevent foreign exchange reserve depletion, leading to a high black market premium on foreign exchange. In this milieu, foreign banks were reluctant to keep rolling over loans, forcing governments to switch to monetary financing of the fiscal deficit. Furthermore, the rate of inflation to generate a given amount of seigniorage for financing the fiscal deficit went up as the population’s ability to shift into dollars raised the inflation elasticity of domestic money demand.

Third, inflation may have become entrenched as a result of the indexation of wages and asset prices, as in Brazil during the 1970s and 1980s, making extrication from high inflation all the more difficult. Not surprisingly, the major Latin American countries entered a rut of repeated failures in stabilization, ever higher public debt, and severe costs for growth and economic welfare, especially for vulnerable groups.

The link between stabilization programs and debt crises provides a natural bridge from the 1980s debt crises to those of Russia in 1998 and Argentina in 2001. Russia achieved single-digit inflation in early 1998 but suffered a devastating triple exchange rate-public debt-banking sector crisis less than six months later. This 1998 crisis had echoes in the 2001 crisis in Argentina. Both involved fixed (managed in the case of Russia, constitutionally mandated in the case of Argentina) pegs to the dollar, which had been chosen to squeeze inflation out; both eventually developed unsustainable debt dynamics (masked by real appreciation of the exchange rate in conjunction with a significant share of public debt denominated in dollars); and in both cases, banks became vulnerable to sovereign risk. In addition, Argentina’s banks became vulnerable to currency mismatches, having borrowed in US dollars but on-lent to companies with local currency revenues. The net result was a downgrading of growth prospects and a rise in interest rates, which eventually fueled a meltdown. We shall not go into the details of these crisis episodes, which
have been well-documented elsewhere, but we use these as a springboard for a dis-
cussion of the implications for sovereign debt restructuring later in the paper.\textsuperscript{12}

\textit{Official Intervention: Insights from the 1980s}

Following Mexico’s announcement in August 1982 that it could no longer service
its external commercial bank loans, 27 countries owing US$239 billion had either
rescheduled their bank loans or were engaged in doing so by October 1983. Sixteen
were from Latin America; of these, the four largest debtors, Mexico, Brazil, Venezuela,
and Argentina, owed US$176 billion, or 74 percent of the total outstanding EM
debt.\textsuperscript{13} Although it was evident by 1985 that the debtor countries were not recover-
ing, debt reduction remained politically unacceptable. Instead, the Baker Plan,
named after US Treasury James Baker, was launched in October 1985. It emphasized
new lending from commercial banks in exchange for market-based reforms. The 10
Baker Plan agreements rescheduled US$165 billion of debt. The World Bank was
expected to play a large role with “structural adjustment loans” in helping implement
the market-based reforms. However, the plan did not work, and with the decade
being inexorably lost, the US government finally threw its weight behind debt reduc-
tion. The Brady Plan was announced by US Treasury Secretary Nicholas Brady in
March 1989, with Mexico becoming the first major test. In total, US$60 billion of
debt was forgiven, and US$200 billion of bank claims were converted into US$154
billion of Brady bonds.\textsuperscript{14}

From the perspective of achieving an ODR, three questions stand out: (i) Is official
intervention needed? (ii) When is official intervention most likely? and (iii) does offi-
cial intervention help? We consider each question in turn.

Is official intervention needed? Krugman (1994, 710) noted with regard to the
Brady Plan that the idea of a voluntary approach was soon dropped, and a “combi-
nation of legal maneuvering and pressure on banks” left no option but to partici-
pate in a debt reduction program. The notion that a voluntary approach would not
work is intuitively plausible: no creditor would willingly write down their claims,
and this could scuttle what is a collectively superior outcome. This understanding is
reinforced by the events during the Baker Plan intervention. Though commercial
banks were supposed to come up with new money, Krugman (1994) and Dooley
(1994) note that the main outcome was that they managed to substantially reduce
their exposure to the debtor countries over the course of the 1980s and that loans
from official creditors rose sharply. Therefore, official intervention is needed for an
ODR because otherwise an impasse would result.

When is official intervention most likely? The economic self-interest of the more
advanced and influential countries plays a powerful role. The immediate response
after Mexico’s default announcement in 1982 was to provide government-to-
government bridge loans so that debtor countries could remain current on their
interest payments and thereby avoid imperiling the US banking system. The four largest debtors, Mexico, Brazil, Venezuela, and Argentina, owed $37 billion to the eight largest US banks, which amounted to 147 percent of these banks’ capital and reserves. At the time, the US was in a deepening recession, and there were widespread fears that a debt default could trigger another Great Depression. The bridge loans eventually led to the Baker Plan. The need for concerted lending stemmed from the “free-rider” problem; each bank on its own would have preferred to reduce its exposure, but every bank doing so might have forced a default and, eventually, an international financial meltdown. Solving this collective action problem provided the rationale for intervention by the US.

Once the need for debt reduction was recognized, a beneficial solution for both debtors and creditors would have been for the commercial banks to voluntarily accept a haircut in the hope of lowering the probability of default and raising the market value of the remaining debt as per the debt overhang argument (Sachs 1986; Krugman 1988). In other words, indebtedness had reached a crippling level which, in the event of a default, would lead to such a low recovery value for creditors that it would be better for them to forgive part of the debt and recover a larger amount as the country resumed growing and restored solvency—an outcome that would be facilitated by the partial debt forgiveness. However, such forgiveness was unlikely to happen spontaneously because individual creditors would be tempted to hold out under a voluntary scheme to gain on their entire holding of the country’s debt. Solving this free-rider problem provided the rationale for the Brady Plan. Thus, the economic arguments for official intervention—solving the collective action problem and lowering transactions costs—are clear. However, the question remains whether such intervention will materialize without the interests of influential countries being at stake.

Does official intervention help? Dooley (1994) conjectures, “It is difficult to rule out the possibility that all the direct benefits of the Brady deals to date went to the banks. Moreover, it is generally agreed that the direct benefits of Brady restructurings have been too small to account for much of the increase in the secondary market prices since 1990”. The specific case of Mexico is insightful. While acknowledging the official arm-twisting needed for debt reduction, Krugman nevertheless notes (1994, 702), “Mexico achieved a reduction in the present value of its debt of approximately $14 billion or 14 percent. This was clearly insufficient . . .”.15 Krugman (1994) describes the outcome of Baker and Brady as exhibiting two features: financial stability was maintained, but the debtor countries did badly. He cites Cline (1990) as arguing that the debtor countries were going to do badly on growth anyway, so saving the financial system was a signal success. However, Eichengreen and Portes (1989) showed that countries willing to default early and massively during the 1930s crisis did better than those not willing to do so. This eventuality was pre-empted by the Baker and Brady Plans during the 1980s.
Turning to the last round of EM crises during 1997–2001, the experiences of Russia in 1998 and Argentina in 2001 clearly show that the intervention of the IFIs in situations of low foreign exchange reserves and unsustainable debt dynamics carries a serious risk of prolonging crisis, eventually requiring the country to address a much larger debt problem. This is what Rogoff (2003), former chief economist of the IMF, had to say about the Russian rescue package of 1998:

As a result, the official lending community, typically led by the IMF, is often unwilling to force the issue and sometimes finds itself trying to keep a country afloat far beyond the point of no return. In Russia in 1998, for example, the official community threw money behind a fixed exchange-rate regime that was patently doomed. Eventually, the Fund cut the cord and allowed a default, proving wrong those many private investors who thought Russia was “too nuclear to fail.” But if the Fund had allowed the default to take place at an earlier stage, Russia might well have come out of its subsequent downturn at least as quickly and with less official debt.

Interestingly, Rogoff (2003) attributes procrastination in Russia 1998 to the fact that “. . . current international law makes bankruptcies by sovereign states extraordinarily messy and chaotic”. However, the analysis in Kharas, Pinto, and Ulatov (2001) points to mistakes in diagnosis that resulted in a rescue package that emphasized liquidity over solvency and eventually led to a much larger problem.16

In its post-mortem of Argentina 2001, the IMF’s Independent Evaluation Office noted that official rescue packages are unlikely to be catalytic in insolvency situations, that financial engineering in the form of voluntary debt swaps is ineffective, and that procrastination is costly.17 These were very much the lessons from Russia 1998. The first two lessons, on why rescue packages may not be catalytic and the inefficacy of sovereign debt swaps, are discussed below. The third, on procrastination, is discussed in Section 3.

Official intervention that is catalytic could mean one or all of three things: (i) private holders of government bonds are persuaded not merely to rollover maturing loans but also to increase their exposure to the debtor government; (ii) the government implements fiscal and structural reform as part of the official rescue package that places government debt on a sustainable trajectory and improves growth prospects; and (iii) interest rates come down because risk spreads relative to the benchmark country (for example, the US or Germany) decline.

It should be obvious that all three would be much easier to achieve if the government faced a liquidity but not a solvency problem. In fact, one could argue that in the case of a pure liquidity problem, only (i) would be needed as part of the catalytic effect (as in Morris and Shin 2006). However, in a solvency problem, an official rescue loan package could worsen the situation due to the seniority of official loans in conjunction with certain design aspects of the package, which we
discuss later. This brings us to why sovereign debt swaps tend not to work when fiscal fundamentals are weak.

Neither the 1998 Russian swap out of GKOs (ruble T-bills) into Eurobonds nor Argentina’s 2001 mega-swap, which were designed to lower borrowing costs and lengthen maturities, worked. Even worse, they backfired. The reason is that in a market-based voluntary debt swap (the case both for Russia 1998 and Argentina 2001), investors work to protect the value of their assets. For debt swaps to work, they have to reduce the debt burden of countries. Creditors are unlikely to let this happen in a voluntary fashion—a result that the reader will recognize as a variant of the Modigliani-Miller Theorem from corporate finance. Indeed, creditors could demand additional compensation that would worsen the fiscal situation. For example, Argentina’s swap was concluded at a spread of 1,100 basis points, whereas according to Mussa (2002), calculations showed that at spreads of over 1,000 basis points, Argentina’s debt dynamics were “virtually hopeless”. After the swap, meltdown proceeded as tax collections continued to flag, bond spreads rose further, and bank runs intensified because of concerns about the viability of the hard peg. Six months later, Argentina defaulted on its debt, including the bonds restructured as part of the mega-swap.

Procrastination in Sovereign Debt Restructurings

Procrastination is a major impediment to ODRs. We discuss two points: first, why procrastination is costly; second, we attempt to understand why procrastination occurs.

Why Procrastination Is Costly

With adverse debt dynamics and diminished chances of a positive catalytic effect of official bailout funds due to official seniority and debtor country solvency concerns—illustrated most vividly by Russia 1998 and the common lessons from this crisis and Argentina 2001 that were discussed above—procrastination becomes costly. This is because the ratio of debt-to-GDP continues rising until a default or debt restructuring becomes unavoidable. The official bailout funds only enable short-term creditors to exit at 100 cents on the dollar. The costs to the debtor nation and creditors can then rise substantially, except for the short-term creditors. The long-term prospects for the debtor country and for the remaining private creditors are likely to worsen. To the extent that private creditors hold both short- and long-term claims, they are liable to lose on the latter whatever they gain on the former. Greece is the latest illustration of this point and is discussed further below.

118 The World Bank Research Observer, vol. 29, no. 1 (February 2014)
Returning to the 1980s, an FDIC study (FDIC 1997) noted that the four largest debtors, Mexico, Brazil, Venezuela, and Argentina, owed $37 billion to the eight largest US banks, which amounted to 147 percent of the banks’ capital and reserves. Suppose the banks had to take a 25 percent haircut on these loans. This would have amounted to $9.25 billion and wiped out some 37 percent of these banks’ capital. However, even their total exposure, $37 billion, was less than 1 percent of the US GDP in 1985. Different and less costly strategies were conceivable that would have benefited millions of poor people in Latin America, who eventually bore the brunt of its lost decade even as official intervention kept the banks going. Borrowing access by the debtor countries would most likely have been disrupted. However, such a hardening of their budget constraint was needed to address the underlying fiscal and governance problems, as the country studies by Sachs (1990) indicate.

Similarly, in Russia’s case, it became obvious by mid-May 1998 that government debt was on an unsustainable course and that the ruble was hugely overvalued. Suppose the official community had persuaded Russia at that point to float the ruble, backed it in seeking a haircut for private creditors, and given it an official rescue package at the risk-free rate (reflecting its seniority and risk status). The situation would still have been difficult, and the U.S. most likely would still have had to bail out the systemic hedge fund operated by Long Term Capital Management (LTCM), which fell victim to contagion from the Russian default (Dungey et al. 2006). However, the problem would have been smaller, as noted by Rogoff. Dollar-denominated public debt had increased by $16 billion, or 8 percent of the post-crisis GDP, over the 10 weeks preceding the date of the crisis.20

In the case of Greece, the October 2011 EU Summit announcements on the need for haircuts for private creditors were in line with the conclusion in Chamley and Pinto (2011). They were eventually implemented in March 2012, two years after negotiations on a bailout began. In the meanwhile, a debt problem amounting to 3-4 percent of the euro area GDP in Greece had mushroomed by the summer of 2011 into a situation in which “nearly half of the €6.5 trillion stock of government debt issued by euro area governments . . .” was at risk, with the crisis having spread to Italy and Spain (IMF 2011, p. 16). As a result, the euro area has become locked into an interacting vulnerability linking sovereign debt and exposed banks.

This situation raises a tantalizing question: had the haircut for Greece’s private creditors been imposed upfront in March 2010 when the solvency problem was first detected and the vulnerable foreign banks ring-fenced, would the wider contagion have been avoided? A speculative attack on the debt of other vulnerable sovereigns might well have followed, but this could have also spurred a more decisive response by the official community, including the major central banks. The back-to-the-wall effects of crises in concentrating the mind and a two-year head start in implementing fiscal, banking, and structural reforms in the vulnerable Eurozone countries

Canuto et al. 119
should not be dismissed. As it turned out, contagion spread to the core of the Eurozone. Notwithstanding the palliative effects of ECB’s LTRO interventions (see Introduction), Italian 10-year bond yields were once again approaching the 7 percent “bailout” threshold by June 2012.

Two subsequent events have kept a lid on the spreading sovereign debt crisis in the Eurozone. The first was a now-famous announcement at the end of July 2012 by Mario Draghi, President of the ECB, that the ECB would do “whatever it takes” to support the euro and keep the Eurozone together. The second was an announcement in September 2012 by the ECB about its government bond buying program, Outright Monetary Transactions (OMT), provided the country concerned agreed to a corrective program. At the same time, attention turned to creating a fuller fiscal and banking union to save the euro. As of November 2013, the fiscal and banking unions as well as the OMT program remain works in progress. In its April 2013 Fiscal Monitor, the IMF observed that of the ten advanced countries with a debt-to-GDP ratio over 90 percent and with adverse dynamics, seven are in the Eurozone. The sovereign debt situation has noticeably worsened in the euro periphery.

Why Is There Procrastination?

Let us return to the 1980s. It took seven years to accept that debt reduction was needed. It took a few years more to negotiate and implement the Brady debt deals based on the menu of options available. However, even if the Brady deals had been negotiated and implemented instantaneously, seven years would still have been lost. Thus, the first impediment to an ODR seems to be an inbuilt bias towards procrastination. Where does this procrastination come from? One might think that politicians in the debtor country have an incentive to procrastinate rather than to admit that mismanagement on their watch led to a default. However, Mexico was quick to admit in 1982 that it could no longer service its external debt. Similarly, Russia pulled the plug on its international rescue package by devaluing and defaulting on August 17, 1998, less than a month after it had been approved by the IMF.

Do private creditors have an incentive to procrastinate? If the prospects for the country are bad and the debt dynamics adverse, then individual creditors have an incentive to sell and exit before the others do to minimize their own losses in line with the prisoners’ dilemma. This would bring matters to a head, but might not happen for two reasons. First, if all creditors exit simultaneously and panic results, everyone loses much more, as in a fire sale. This might keep creditors from exiting. A second pivotal reason is the anticipation of an official bailout package. However, creditor reaction would also depend on the maturity of the debt held. If it is short term, there is a clear incentive to hang on if the probability of a large rescue package is high and exit at 100 cents on the dollar. However, this would be at the expense of long-term bondholders.
What about economists? Implicit in the preceding Rogoff quote on Russia 1998 is that economists should know when a currency is overvalued and when there is a solvency rather than a liquidity problem. In fact, Rogoff was arguing that economists knew it all in the case of Russia 1998 but were driven to continue with the (unsustainable) status quo because there was no easy bankruptcy process for sovereigns. We sympathize with the view that economists should be able to assess whether a currency is overvalued and whether the public finances are salvageable without a debt write down.

What about the official community, including the IFIs? One would have to admit that the record is mixed, with Russia 1998 and Argentina 2001 as examples of procrastination and the flawed design of rescue packages. The latter, in particular, means one of two things: either (a) that the economists involved were not sufficiently astute in assessing the sustainability of the fixed exchange rates or of the public finances or (b) that debt reduction is anathema and that the official community will do whatever it takes to bail out private creditors and avoid setting a moral hazard-inducing precedent for debtor countries. Although one might have believed that (b) was true, the experience from Russia 1998, Argentina 2001, and Greece 2010 suggests that it must be revisited.

ODRs – What Should They Look Like?

Successful ODRs have been few and far between, typically involving tiny amounts of debt. One was Ukraine’s debt exchange offer of February 2000 involving $2.6 billion, which achieved an NPV reduction of 22 to 35 percent (Table 5.4, Sturzenegger and Zettelmeyer 2006) and elicited a high participation rate. Ukraine was then under the IMF’s three-year US$2.2 billion Extended Fund Facility (signed in September 1998), and the IMF made it clear that Ukraine could not use its low reserves to service maturing debt and that the IMF program depended upon a satisfactory debt restructuring. This unambiguous signal of “no bailout” persuaded private creditors to agree rapidly to a deal. In addition, Pakistan’s debt restructuring of 1999 involving $610 million (Table 6.3, Sturzenegger and Zettelmeyer 2006) was in large part due to a comparability requirement imposed by the Paris Club, which had rescheduled Pakistan’s loans in January of that year. The swap offer attracted a participation rate of close to 99 percent, partly because of default concerns with the original bond, and it achieved a reduction of 30 percent in NPV terms.

However, lingering dissatisfaction with the process and outcome of debt restructurings in more complicated cases has prompted a few corrective proposals. Sachs (1995) proposed an international bankruptcy mechanism to achieve ODRs that would entail a payment moratorium by the debtor country during debt

Camuto et al. 121
renegotiations. The Sovereign Debt Restructuring Mechanism (SDRM) was proposed by the IMF in 2001 to facilitate creditor coordination in the event of debt restructurings for bond debt, the holdings of which are much more dispersed than the concentrated syndicated bank loans that featured in the debt crisis of the 1980s. In addition, a voluntary code of conduct was proposed by Jean-Claude Trichet in 2001 that spelled out nine principles governing creditor-debtor relations during debt restructurings. However, none of these proposals has gained traction.

The only mechanism that has been widely accepted by the market has been Collective Action Clauses (CACs). These are part of the terms and conditions governing a bond issue and can be invoked by the debtor government. The most frequently used CAC is one that entails a modification of payment terms requiring a favorable vote by a majority of the outstanding bond holders (75 percent, typically; 85 percent in some cases, but it could be lower). Empirical evidence on the impact of CACs on bond pricing has been inconclusive, and their usefulness in achieving an ODR is questionable.

How should the way forward look? Based on Sections 2 and 3, we posit three conditions that an ODR should fulfill at a minimum, with which we believe most economists would be comfortable:

- Restore the debtor country’s government debt to a sustainable trajectory;
- Minimize procrastination and costs for both the debtor country and its creditors; and
- Minimize any harmful contagion effects in our interconnected world.

However, the cumulative EM experience augmented with Greece 2010 demonstrates that the preceding conditions are seriously incomplete when there is official intervention in an insolvency setting, that is, when the present value of primary surpluses is less than the present value of outstanding debt obligations. In this case, either primary surpluses will need to be raised (“fiscal reform”) or debt will have to be written down (“haircuts”) to restore solvency. Suppose the market does not believe that primary surpluses can be raised to restore solvency. In this case, a numerical example based on Chamley and Pinto (2011) in Annex 1 shows that two additional conditions are needed: first, there should be an upfront haircut for private creditors to help restore debt sustainability; second, official funds should be lent at the risk-free rate, reflecting their seniority. These two conditions will lead to a less onerous and therefore more credible fiscal program to restore solvency because there is less debt to address and a smaller official loan will be required. Moreover, an upfront haircut imposed on all private creditors is more equitable than a situation in which short-term creditors gain at the expense of long bondholders. This might induce long bondholders to hang on instead of selling off. It could also have political economy benefits: with private creditors receiving an
upfront haircut, the less severe fiscal austerity program becomes easier to sell to the public.

Obstacles to an ODR

The most controversial aspect of an ODR discussed above is likely to be the idea of an upfront haircut for private creditors in the event of a solvency problem for the debtor country. Three objections could be raised: moral hazard, the difficulty in distinguishing solvency from liquidity problems, and political economy considerations.

Moral Hazard

For governments, moral hazard implies that countries deliberately and irresponsibly run up debt to precipitate a solvency problem in which private debt will be written down. Although it is conceivable that countries have behaved in this manner in the past and could do so again in the future, such behavior is unlikely to be the norm. Three points are worth noting in the specific context where the IFIs (international financial institutions, such as the IMF and World Bank) are brought in to orchestrate a rescue aiming to restore the government to solvency.

First, consider who is really being bailed out. It cannot be the country because any official funds received have to be paid back in full, and such debt is difficult to renegotiate. Therefore, engineering a situation in which official loans are obtained to pay off maturing private debts at 100 cents on the dollar does not “subsidize” the country’s “bad” behavior, although one cannot rule out unfair redistributions within the country itself as well-connected people benefit from the external loans that are then serviced by the taxpayers, as frequently happened in Latin America during the 1970s and 1980s.

Second, moral hazard implies a proclivity by countries to default strategically, that is, to default based on unwillingness to pay rather than an inability to do so. Once again, there is little evidence to support such a position. Countries typically default only as a last resort.

Perhaps the most compelling argument against moral hazard by debtor countries is the unambiguous trend toward self-insurance by EMs, documented in Aizenman and Pinto (2011, 2013) and Pinto (2014, chapter 9). By definition, a country prone to moral hazard will not self-insure because this would be contradictory to the idea that someone else is insuring the country’s behavior. However, EMs moved aggressively to self-insure at three levels after their 1997–2001 crises, taking steps to (a) restore sustainable public debt dynamics by raising primary surpluses and strengthening fiscal institutions; (b) insure against shifts in market sentiment and sudden stops by building up foreign exchange reserves and restricting
currency mismatches on government and private balance sheets; and (c) lower contingent liabilities from the private sector by shifting to flexible exchange rates, monitoring private external borrowing, and strengthening financial institutions.

What about moral hazard for private creditors? First, these creditors price risk and are diversified. Nevertheless, as documented in Kharas, Pinto, and Ulatov (2001, Box 2), private creditors are often in the forefront of the drumbeat for big bailout packages. What would be better than to price government debt at default levels and exit at 100 cents on the dollar? In other words, they are not innocent bystanders. Second, the economic benefits of external financial integration for developing countries are seriously questioned (Aizenman and Pinto 2011, 2013), and one should take threats that haircuts for private creditors will have disastrous effects for EMs (by shutting off market access) with a grain of salt. If anything, experience shows that disruptions in market access force countries to finally address the fundamental fiscal problem at the root of sovereign debt crises. The self-insurance by EMs after 1997–2001 discussed above is an extreme manifestation of precisely such behavior. Third, where official funds are used to bailout private creditors, the primary fiscal surplus targets needed to assuage default fears on the remaining private debt might simply be out of reach, as argued earlier.

Ultimately, the prospect of an upfront haircut for private creditors is a matter of pragmatism because it increases the chances of a successful and credible fiscal program and implies equal treatment for short- and long-term creditors. Otherwise, it makes little sense to inject senior official funds into an insolvency situation. The knowledge that they would be subject to an upfront haircut in the event of insolvency would also make private creditors exercise greater caution ex ante in lending to sovereigns, reducing an important source of moral hazard, which is in line with the caveat emptor principle.

**Solvency versus Liquidity**

The seriousness of insolvency can be gauged from two vantage points. First, what are the market signals? If bond spreads indicate a high probability of default and keep rising even as official bailout discussions continue (the case with Russia 1998, Argentina 2001, and Greece 2010), then this should be considered a red flag. Second, what are the country’s fiscal and growth prospects as conveyed by an assessment of its fundamentals by economists at the IFIs? How likely is the country to generate the needed primary surpluses to stabilize and even lower debt? If the chances are slim, accompanying official loans at the risk-free rate with an equal upfront haircut for all investors will enable more credible fiscal targets and lead to lower long-bond spreads, minimizing reputation costs.

However, we admit that it is not always easy to judge whether a country is dealing with multiple equilibria (liquidity and confidence) or a fundamental
(insolvency) problem. A case in point is Brazil in the summer of 2002, just before Luiz Inacio Lula da Silva, the candidate of the Brazilian Workers’ Party, was elected president. Bond spreads reached 2000 basis points that July as presidential election polls “indicated that Lula would win the presidential election . . .can [investors] be certain that a Brazil run by a president with a past record of sympathizing with default will not take the easy way out?” In the article from which this quote was taken, Williamson (2002) argued that fundamentals were sound: primary fiscal surpluses had been raised substantially, and budget constraints hardened for the state governments. Additionally, the real had been floated in 1999.37

In his classic on multiple equilibria, Obstfeld (1994) recognizes, “Ultimately...crises based on limited foreign reserves [liquidity] must also be based on overall fiscal weakness: [otherwise]...it would be...feasible to borrow sufficient reserves to...fend off any attack [on the fixed exchange rate]” (material in square brackets added). Brazil was different in that it had taken clear steps to strengthen its fiscal fundamentals. Not only was it running significant primary surpluses, it had raised these substantially compared to the period prior to the float of the real in January 1999. Adopting a float while moving to address currency mismatches (the government was simultaneously shifting toward local currency debt) would substantially alleviate the international liquidity problem because the central bank would no longer be in the position of having to defend a fixed peg. Therefore, at the time that the bond vigilantes went after it in the summer of 2002, Brazil’s problem was one of political risk and confidence, which was boosted, albeit feebly, by the announcement of a $30 billion loan from the IMF. In the case of Greece, bond spreads continued rising with each successive bailout augmentation.

Hence, if a country is taking steps to self-insure along the comprehensive lines discussed above (including the adoption of flexible exchange rates and hardening budgets), then one might be a bit more cautious about confusing a liquidity with a solvency problem. However, this was not the case with Russia 1998, Argentina 2001, or Greece 2010, which were much more clear-cut ex ante on both market signals and fundamentals.

Political Economy

We interpret “political economy” as anything that would lead to procrastination. The greatest resistance is likely to come from the creditors themselves. For example, the large commercial banks holding Greek debt were in the forefront of warning against any Greek debt restructuring because of the contagion risks. It was in their interest to allow official creditors to replace exiting private creditors at 100 cents on the dollar while fiscal and structural reforms were implemented.

However, three caveats are in order. First, creditors are not a homogeneous bunch. They are distinguished by the maturity of the debt they hold, with
short-term creditors benefiting most from the strategy of taking the sovereign out of the market for a few years with official creditors replacing exiting private creditors; by the size of their exposure; and by whether they are covered by insurance, for example, in the form of credit default swaps, especially if such insurance was purchased before the solvency problem was detected. Second, replacing exiting private lenders with official lenders does not lower the country’s debt burden in a present value sense and results in more ambitious fiscal targets to restore solvency, which are by definition unattainable; otherwise, we would be dealing with a liquidity and not a solvency problem. Therefore, there is a risk that the country could abandon the program for political reasons before short-term creditors have exited. This was definitely the case in Russia 1998 and, judging by the extraordinarily high two-year bond spreads (see Introduction), appeared to be the fear in the case of Greece 2010 as well. Third, creditors may hold both short and long maturity debt. In this case, provided the official loans are priced in accordance with their risk (and the interest rate should equal the risk-free rate in the case where official loans are first in the queue and small enough to be paid in full), creditors could lose on their holdings of long-term debt what they gain on their short-term debt holdings. This is because in an insolvency case, a default and debt write down become inevitable, with all the burden of the restructuring falling on the remaining private debt.

Therefore, an upfront haircut is simply a way of distributing restructuring costs more fairly across creditors in the event of insolvency. It is the analog of the “concerted lending” approach applied to the money center banks during the 1980s crisis to pre-empt the free-rider problem—that is, some banks reducing their exposure as others roll over their loans—once a debt overhang develops.

Conclusions

The record on official intervention in sovereign debt crises is not flattering, whether it be the 1980s, Russia 1998, Argentina 2001, or Greece 2010. An important reason is the tendency to procrastinate, treating solvency problems as liquidity problems even when the distinction between the two is clear. If the goal of official intervention in such circumstances is to teach debtor countries a lesson, nothing needs to change. However, if the goal is to increase the likelihood of an orderly debt restructuring, then pricing official loans at the risk-free rate (in line with their more-or-less zero risk) and insisting on an upfront haircut—when the bargaining power of the official sector is the greatest—for private creditors will help. It will also share the burden more equitably between short- and long-term creditors. Although this may appear to be a recipe for moral hazard, the aggressive self-insuring behavior of emerging market countries after their crises of 1997-2001 suggests behavior diametrically opposed to what one might expect from countries confident...
of being bailed out should they run up debt irresponsibly. Additionally, private creditors are hardly innocent bystanders; they are sophisticated investors who price risk and are diversified. Therefore, an upfront haircut in the event of a solvency problem should not come as a total surprise to them and could make their \textit{ex ante} lending behavior more diligent.

Ultimately, there is a stark choice between two strategies: gambling for redemption as in Conesa and Kehoe (2011), in which an immediate haircut is avoided, and insisting on an upfront haircut for private creditors while keeping in mind that the cost of default is large and that the chances of a default increase with procrastination, as in Chamley and Pinto (2013). The approach to sovereign debt restructuring favored by official agencies has been to gamble for redemption, reflected in the use of official funds to take the country out of the market while implementing fiscal and structural reform to raise primary fiscal surpluses and spur growth. It has tended not to work—Latin America in the 1980s, Russia 1998, Argentina 2001, and Greece 2010 are examples.

However, the official mindset may be changing in favor of inflicting upfront haircuts on private creditors when a country is obviously insolvent. The official approach to the Cypriot banking crisis, which came to a head in March 2013, was to insist on losses for bank depositors because the size of the banking system (800 percent of GDP) precluded a government bailout. Indeed, in a May 2013 draft law on bank resolution, the EU embraced the idea that in future banking crises, burden sharing might be needed with shareholders, bondholders, and uninsured depositors. On June 5, 2013, the IMF published its post-mortem of the Greek bailout, noting that (IMF 2013, 28; words in square brackets added) "not tackling the public debt problem decisively at the outset or early in the program created uncertainty about the euro area’s capacity to resolve the crisis and likely aggravated the contraction in output. An upfront debt restructuring would have been better for Greece . . . . A delayed debt restructuring also provided a window for private creditors to reduce exposures and shift debt into official hands . . . [which] occurred on a significant scale and limited the bail-in of creditors when PSI eventually took place, leaving taxpayers and the official sector on the hook."

In conclusion, we have laid out the desirable attributes of an orderly sovereign debt restructuring (ODR) when a solvency problem is involved. In addition to ensuring that government debt attains a sustainable trajectory and that procrastination and contagion costs are minimized, pricing official funds in line with their risk and using official bargaining power to insist on an upfront haircut for private creditors would be desirable. However, there is no perfect formula for distinguishing between liquidity and solvency problems and upfront haircuts are going to encounter stiff political resistance. Therefore, ODRs are likely to remain elusive, recent signs of a change in the official mindset notwithstanding.
Notes

* The authors are Senior Adviser, Development Economics Department of the World Bank; Chief Economist, Emerging Markets, GLG Partners LP; and Senior Economist, Poverty Reduction and Economic Management Department in the Europe and Central Asia region of the World Bank, respectively. They acknowledge valuable comments from three anonymous referees. The views herein are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations or those of the Executive Directors of the World Bank or the governments they represent.


2. The ECB lent €498 billion to 523 banks in December, with another €53.3 billion to 800 banks at the end of February 2012.

3. The New York Federal District Court ruling in November 2012 on the Argentine default is a case in point. The ruling required Argentina to honor the pari passu clause, which means equal treatment of all bondholders. Hence, it required the Argentine government to make payments to holdout creditors if the restructured bonds of 2005 and 2010 were honored. The U.S. Federal Appeals Court upheld this decision in August 2013. However, as noted by Pinto (2014) and in contrast to the discussion on self-insurance later in this paper, Argentina is an exception to the rule that EM government behavior changed for the better following the crises of 1997–2001.

4. A detailed account of specific country episodes involving sovereign debt restructuring is contained in the work of Sturzenegger and Zettelmeyer (2006). Pinto and Tanaka (2005) describe the various instruments and options availed of during these restructuring episodes, starting with the 1980s debt crisis.


7. This statement applies to the period under review. This does not rule out macroeconomic crises with flexible exchange rates. For a theoretical example of the latter, see Kumhof, Li, and Yan (2007).


9. An exacerbating factor was that although they were of long maturity, these external hard currency debts had floating rates, with the interest rate adjusted every six months based on the market index. Therefore, once the U.S. started raising interest rates, the interest burden of the developing countries quickly increased.


11. See, for example, the case study on Argentina by Dornbusch and de Pablo (1990). Brazil had several stabilization programs during the 1980s and until 1994. In July 1994, after six failed price stabilization plans over the previous ten years, Brazil finally initiated a successful stabilization effort embedded in the Real plan. It lowered consumer inflation from 228.7 percent in 1994 to 71.9 percent in 1995, 18.2 percent in 1996, and finally to 7.7 percent in 1997. See Blanco et al. (2011).


16. The arguments that follow were made prior to the crisis in real time by the economics unit of the World Bank office in Moscow, which the second author of this paper then headed.

18. For a description of the swaps and a formal argument for why they backfired, see Aizenman, Kletzer, and Pinto (2005). In a similar vein, the Mexican Government began rolling over its short-term peso-denominated debt (Cetes) into short-term dollar indexed debt (Tesobonos) after March 1994 to avoid raising interest rates to deal with rising devaluation risk. This became a major source of financial vulnerability. See Sachs et al. (1996).

19. In the sovereign case, the market value of the debt is determined by the present value of the future primary fiscal surpluses. As long as this is fixed (when discounted at the risk-free rate), shuffling the mix of debt instruments through market-based swaps will not change the present value of the debt burden.

20. Details may be found in Kharas, Pinto, and Ulatov (2001) and Pinto and Ulatov (2012).

21. Although Rogoff’s is an excellent point, the absence of an easy bankruptcy procedure was not the only reason the doomed Russian rescue package of July 1998 proceeded. There was a strong belief in influential quarters that with Russia having achieved single-digit inflation in February 1998, these hard-won stabilization gains had to be preserved, and that with Russian government debt much less than the Maastricht criterion of 60 percent of GDP, the market was overreacting.


23. In contrast, the debt amount involved in Russia 1998 was $77 billion ($45 billion in ruble debt, $32 billion owed to the London Club). Pinto, Gurvich, and Ulatov (2005, 431).


25. See Pinto and Tanaka (2005).


29. For example, it was reported in the news on March 9, 2012 that Greece was able to secure a 95.7 percent participation rate among private-sector creditors in its bond exchange by invoking CACs to make the deal binding on holders of Greek-law bonds (until the bailout began, much of Greek debt was under Greek law). However, by then, severe damage had been done to the Greek economy and the wider euro area.

30. For technical details, see Burnside (2005).

31. See, for example, de Bolle, Rother, and Hakobyan (2006).

32. Sturzenegger and Zettelmeyer (2006, 4, 38) argue that most sovereign defaults since the 1970s were driven by interactions between domestic policies and economic shocks (including exogenous shocks), sometimes worsened by political shocks. In this sense, ability and willingness to pay are difficult to disentangle. However, Ecuador’s default of 2008 on its US$3.2 billion Eurobonds was a rare instance of a country that did not repay its debt even though it had the resources to do so. The Eurobonds were declared “illegitimate”, and the government bought back 91 percent of the defaulted debt in the secondary market at 35 percent of face value.

33. Tomz and Wright (2007) find a negative but weak relationship between economic output and default on external loans from private creditors. Eden, Kraay, and Qian (2012) also come to a similar conclusion that defaults are more likely to occur during growth slowdowns in countries with weak policy performance and that have seen rapid debt accumulation.

34. This is further corroborated by Aguiar and Amador (2011), who build on the debt overhang argument and find that countries that grow rapidly are those that accumulate net foreign assets because growth in capital requires a reduction in the stock of debt.

35. See Canuto, Pinto, and Prasad (2012, 22–3) for some pointed examples.

36. For example, Eichengreen and Ruhl (2000) argued, in the context of Ecuador, Pakistan, Romania, and Ukraine following the East Asian and Russian crises, that IFIs acted to avoid “a costly,
extended interruption to market access” and were therefore not credible when they sought to impose haircuts on private creditors.

37. Williamson (2002) was partly responding to an estimate by Morris Goldstein that there was a 70 percent chance that Brazil would be forced to restructure its debt by the end of 2003. See Goldstein (2003) and the excellent overview in Giavazzi, Goldfajn, and Herrera (2005).

38. With adverse debt dynamics and the growing risk of contagion, the official sector may find its bargaining power eroding as time passes.

39. These conclusions were anticipated in Chamley and Pinto (2011) and Canuto, Pinto, and Prasad (2012), the working paper version of this publication.

40. Even the official sector may find it difficult to assume a fully objective stance. For example, the interests of the US in the 1980s debt crisis and those of the ECB and EU in the ECB-EU-IMF troika in the Eurozone crisis may not always coincide with what a dispassionate body like the IMF is likely to recommend.

Annex 1: Numerical Example on Official Intervention in Insolvency Situations

Consider the 2-period situation in Annex Table 1. The debt service due in each period is shown in the second row of the table. The risk-free sovereign yield in a benchmark country such as the US or Germany rate is assumed to be 5 percent. In Scenario 1, the government faces a liquidity problem because the debt service payments falling due of 100 in period 0 exceed the primary surplus of 75. However, it is solvent in the sense that the present value of primary fiscal surpluses equals that of the debt to be repaid; both equal 250 when discounted at the risk-free rate of 5 percent. In this case, the government can borrow 25 at the risk-free rate of 5 percent either from the markets or the IMF to make up the difference. The total amount it must repay in period 1 is $25 \times 1.05 + 157.50 = 183.75$, which can be exactly met out of the primary surplus in period 1.

Now suppose an adverse shock occurs and the period 1 primary surplus falls to 175, as in Scenario 2. The government now has a solvency problem in the sense that the present value of the primary surpluses at the risk-free discount rate of 5 percent falls to 241.67. Equilibrium can be restored to the government’s intertemporal budget constraint if the price of the debt were to fall from 1 to 0.967 (the ratio of the present value of primary surpluses to that of debt service due at the risk-free

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Period 0</th>
<th>Period 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Debt service due</td>
<td>100</td>
<td>157.50</td>
</tr>
<tr>
<td>Primary surplus</td>
<td>75</td>
<td>183.75</td>
</tr>
<tr>
<td>Primary surplus</td>
<td>75</td>
<td>175</td>
</tr>
</tbody>
</table>

Annex Table 1. Two Hypothetical Fiscal Scenarios

The World Bank Research Observer, vol. 29, no. 1 (February 2014)
discount rate = 241.67/250). Even with this haircut, the government would still need to borrow an amount given by 96.7-75 = 21.7 in period 0 to pay off the maturing debt. In other words, it has both a liquidity and solvency problem.

Suppose it were to go to the market to borrow this amount of 21.7. What would it be charged? Anticipating the haircut, the market would charge an interest rate \( i \) given by the arbitrage condition: \( 0.967(1 + i) = 1.05 \). This can be solved to give \( i = 8.58 \) percent.\(^1\) The spread jumps from 0 to 358 basis points. The amount of 21.7 can also be sought from official sources, such as the IMF. In this case, the amount due in period 2 is \((21.7 \times 1.05 + 0.967 \times 157.5) = 175\), where the IMF’s seniority means it gets repaid in full; but in line with the Modigliani-Miller theorem, this effect is offset by its charging the risk-free rate. However, the amount due to period 1 private creditors is still subject to the same haircut imposed on period 0 creditors.

This brings us to the situation reflective of Greece 2010, as well as both Russia 1998 and Argentina 2001. Anxious to avoid a first period default, the government goes to official creditors and borrows 25 at the risk-free rate to pay off period 1 creditors \textit{in full}. The big difference is that the size of the official loan goes up from 21.7 to 25. In this case, the price of second-period debt falls from 0.967 to \((175-25 \times 1.05)/157.5 = 0.944\). This is equivalent to an interest rate of 11.2 percent, or a spread of 620 basis points. This is exactly what we have witnessed in practice, with long bond spreads rising substantially and persisting at elevated levels following the announcement and implementation of the bailout for Greece.

A crucial difference in the two responses to the insolvency situation is that in one case, an upfront haircut is imposed on all creditors, leading to a smaller official loan. In the second case, the one witnessed in practice, short-term creditors get paid in full, leaving less for long-term creditors. The point of bringing in official creditors is to engender positive catalytic effects, namely, persuading private creditors to roll over their loans instead of exiting; and putting pressure on the debtor country for fiscal reform and austerity to increase primary surpluses. In practice, short-term creditors have been exiting and bond spreads have continued to rise; in Greece’s case, the share of official loans (official creditors plus the European Central Bank, ECB, and Eurosystem) had risen to 58 percent by the end of April 2012. In effect, the official bailouts have taken the countries out of the debt market in the hope that in the meanwhile primary surpluses will increase to levels consistent with solvency. In the case of Greece, at least, the market does not seem to have ever believed this would happen, as conveyed by the evolution of the long bond spread.

Returning to our numerical example, suppose the country is better off the closer the price of period 2 debt is to 1, which would imply a lower spread and smaller reputation costs. Raising the price all the way back to 1 would require restoring the primary surplus in period 2 to 183.75; then we would be back to a liquidity problem, with the solvency problem solved. This raises the question of how serious
the solvency problem was in the first place, a point we return to later. In the meantime, suppose the interest rate on the official bailout package were higher than the risk-free rate in spite of the seniority of the official loans. Then, by continuity, the second period primary surplus would have to be even higher than 183.75 in order for the price of period 2 debt to return to 1, which is likely to severely strain credibility and derail any catalytic effects of official finance. This points to the importance of pricing official funds at the risk-free rate in view of their seniority as otherwise the credibility of the accompanying fiscal package would be lowered: it would require primary surplus targets that would be too onerous to be believable.

The numerical example shows that when the market believes that there is a solvency problem, an upfront haircut for all creditors combined with official funds priced at the risk-free rate will lead to more believable fiscal targets to restore the country’s reputation. The upfront haircut will require that the official loan be just 21.7 instead of 25; and the period 1 primary surplus target will need to be 180.3 instead of the pre-shock 183.75 to return bond prices to 1.

Notes
1. It can be cross-checked that 0.967 [21.7X1.0858 + 157.5] = 175, where the expression in square brackets equals the new amount payable in period 1.
2. For example, the IMF loan interest rate for Greece, while well below what the market may charge, involves significant spreads above the IMF’s own borrowing cost (a spread of 200 basis points for amounts in excess of 300 percent of quota, which goes up to 300 basis points after 3 years if the credit is still above 300 percent. Greece’s loan was 3,200 percent of quota).
3. The result here on official finance is diametrically opposed to that in Morris and Shin (2006), who treat official and private loans as strategic complements. Here, they become imperfect substitutes because of insolvency and official seniority.
4. Notice that if the official loans were also subject to default, they would be priced above the risk-free rate in anticipation of the haircut; the eventual expected payout would be the same. Therefore, so long as a haircut on official loans is ruled out, these loans should be priced at the risk-free rate.
5. 21.7X1.05 + 157.5 = 180.3.

References


Canuto et al. 133


Paris Club. www.clubdeparis.org


INSIDE

Knight on Inequality in China
Devarajan, Khemani, and Walton on Can Civil Society Overcome Government Failure in Africa?
McKenzie and Woodruff on Business and Entrepreneurship Evaluations
Das Gupta on Population, Poverty, and Climate Change
Canuto, Pinto, and Prasad on Orderly Sovereign Debt Restructuring