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

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


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

Forthcoming papers in

THE WORLD BANK ECONOMIC REVIEW

- Global Income Distribution: From the Fall of the Berlin Wall to the Great Recession
  Christoph Lakner and Branko Milanovic

  Philip Keefer and Stuti Khemani

- The Whole is Greater than the Sum of Its Parts: Complementary Reforms to Address Microeconomic Distortions
  Raphael Bergoeing, Norman V. Loayza, and Facundo Piguillem

- Risky Business: Political Instability and Sectoral Greenfield Foreign Direct Investment in the Arab World
  Martijn Burger, Elena Ianchovichina, and Bob Rijkers

MINI-SYMPOSIUM “DOING BUSINESS”

- The Impact of Business Environment Reforms on New Registrations of Limited Liability Companies
  Leora Klapper and Inessa Love

- Deals and Delays: Firm-level Evidence on Corruption and Policy Implementation Times
  Caroline Freund, Mary Hallward-Driemeier, and Bob Rijkers

- The Role of Regulation on Entry: Evidence from the Italian Provinces
  Francesco Bripi
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning Dynamics and Support for Economic Reforms: Why Good News Can Be Bad</td>
<td>1</td>
</tr>
<tr>
<td>Sweder J. G. van Wijnbergen and Tim Willems</td>
<td></td>
</tr>
<tr>
<td>A Helping Hand or the Long Arm of the Law? Experimental Evidence on What Governments Can Do to Formalize Firms</td>
<td>24</td>
</tr>
<tr>
<td>Gustavo Henrique de Andrade, Miriam Bruhn, and David McKenzie</td>
<td></td>
</tr>
<tr>
<td>Economic Shocks and Subjective Well-Being: Evidence from a Quasi-Experiment</td>
<td>55</td>
</tr>
<tr>
<td>Jacob Gerner Hariri, Christian Bjørnskov, and Mogens K. Justesen</td>
<td></td>
</tr>
<tr>
<td>Does Access to Foreign Markets Shape Internal Migration? Evidence from Brazil</td>
<td>78</td>
</tr>
<tr>
<td>Laura Hering and Rodrigo Paillacar</td>
<td></td>
</tr>
<tr>
<td>The Decision to Invest in Child Quality over Quantity: Household Size and Household Investment in Education in Vietnam</td>
<td>104</td>
</tr>
<tr>
<td>Hai-Anh H. Dang and F. Halsey Rogers</td>
<td></td>
</tr>
<tr>
<td>The Impact of Vocational Schooling on Human Capital Development in Developing Countries: Evidence from China</td>
<td>143</td>
</tr>
<tr>
<td>Prashant Loyalka, Xiaoting Huang, Linxiu Zhang, Jianguo Wei, Hongmei Yi, Yingquan Song, Yaojiang Shi, and James Chu</td>
<td></td>
</tr>
<tr>
<td>Financial Inclusion, Productivity Shocks, and Consumption Volatility in Emerging Economies</td>
<td>171</td>
</tr>
<tr>
<td>Rudrani Bhattacharya and Ila Patnaik</td>
<td></td>
</tr>
</tbody>
</table>
Learning Dynamics and Support for Economic Reforms: Why Good News Can Be Bad

Sweder J. G. van Wijnbergen and Tim Willems

Support for economic reforms has often shown puzzling dynamics: many reforms that began successfully lost public support. We show that learning dynamics can rationalize this paradox because the process of revealing reform outcomes is an example of sampling without replacement. We show that this concept challenges the conventional wisdom that one should begin by revealing reform winners. It may also lead to situations in which reforms that enjoy both ex ante and ex post majority support will still not come to completion. We use our framework to explain why gradual reforms worked well in China (where successes in Special Economic Zones facilitated further reform), whereas this was much less the case for Latin American and Central and Eastern European countries. JEL classification: D72, D83, P21

Why have gradual economic reforms worked out well for China, whereas this is much less the case for most Latin American and Central and Eastern European countries? How is it possible that so many of the reforms that began successfully while enjoying majority support subsequently lost this support, although there are also examples of reforms that did not begin well but nevertheless managed to maintain momentum among voters?

The most dramatic example of a reformist government that lost majority support in spite of strong economic performance is Slovakia in 2006. At that time, the Wall Street Journal Europe wrote,

"Imagine you’re the leader of a country where economic growth is running at 6.3%, your government has been praised by the World Bank as the best market reformer in the world [and] unemployment has fallen to a record low of 10.6% from around 20% in just four years. [. . .]"

* Sweder van Wijnbergen is a professor at the Department of Economics, University of Amsterdam, The Netherlands. He is also affiliated with the Tinbergen Institute. His e-mail address is s.j.g. vanwijnbergen@uva.nl. Tim Willems (corresponding author) is a research fellow at Nuffield College and the Department of Economics, University of Oxford, UK. He is also a member of the Centre for Macroeconomics. His e-mail address is tim.willems@economics.ox.ac.uk. The authors thank the editor (Andrew Foster), two anonymous referees, Philippe Aghion, Björn Brugemann, Tom Cunningham, Allan Drazen, Michal Horvath, Ruixue Jia, Matija Lozej, Torsten Persson, Dani Rodrik, Gérard Roland, and audiences at the EBRD and the 2013 SITE Conference in Stockholm for useful comments and discussions.

With this record in mind, now consider that you face parliamentary elections this Saturday at which, unless the opinion polls change dramatically, you risk annihilation by a leftist opposition party with no experience of government and a policy agenda filled with populist rhetoric.

Welcome to the world of Mikuláš Dzurinda, prime minister of Slovakia, who for the past eight years has led what can reasonably claim to have been the most successful neo-liberal government of the 21st century so far.”

Despite his impressive reform successes, Dzurinda lost the 2006 elections to Robert Fico of the SMER party (a breakaway party from the successor to the original Communist Party of Slovakia), who reversed many of Dzurinda’s reforms.

With important reforms currently being implemented in many African and Southern European states, it is important to understand why such puzzling reversals can occur. In this paper, we focus on the interaction between learning from reform outcomes and the dynamics of public support for gradual economic reforms.

We believe that learning processes play a key role in determining support for reforms. So far, however, the literature has remained relatively silent on this issue. Although there are many informal discussions of learning from reform outcomes, formal treatments are scarce.

Even if everyone gains from efficient reforms in the long run, there will almost inevitably be losers during the transitional phase (for example, certain generations or professions). The model we construct captures the fact that reforms typically generate reform winners and reform losers, but (as emphasized in the seminal paper by Fernandez and Rodrik (1991)) these winners and losers cannot always be identified in advance. That is, there is individual uncertainty, which causes the reform to have uncertain distributional consequences. As the reform progresses over time, voters update their beliefs about whether they will end up in the winners’ group or in the group of losers. Because the full version of our model not only features individual uncertainty but also allows for aggregate uncertainty (which implies that voters are unsure about the exact share of the population that will benefit from reform), this paper can be seen as augmenting the Fernandez-Rodrik setup with aggregate uncertainty and learning dynamics.

The fact that existing reform measures affect the distribution from which future sampling will occur plays a key role in our analysis. Specifically, the process of revealing reform outcomes is an example of sampling without replacement. This implies that the revelation of reform winners deteriorates the quality

2. With the possible exception of price decontrol, all reforms are gradual (as opposed to “big bang”), if only because of implementation delays. As noted, for example, by Gupta, Ham, and Svejnar (2008), even reforms that were supposed to be “big bang” (such as the Balcerowicz reforms in Poland) were not completed instantaneously, thereby giving voters an opportunity to update their beliefs about the effects of the reform. In this sense, all reforms are gradual, but some reforms are “more gradual” than others.

3. Some exceptions are van Wijnbergen (1992, where voters learn about the effects of price reform), Dewatripont and Roland (1995, where the public uses early reform outcomes to learn about the expected outcome of later ones), Veldkamp (2009, where laid-off workers learn about their re-employment chances), Strulovici (2010, analyzing determinants of collective experimentation), and Morrow and Carter (2013, studying the impact of learning dynamics on the support for redistributive policies).
of the remaining pool, thereby making unreformed agents less eager to continue the reform process. We derive a condition under which these dynamics are so strong that they lead to the counterintuitive situation in which reform successes make the median voter begin opposing a reform he used to support. In these circumstances, the reforming government suddenly loses majority support although (or because) the reform is progressing in such a successful way. Consequently, even reforms that enjoy both ex ante and ex post majority support may not come to completion.

We emphasize that this phenomenon results from rational economic thinking and that it arises as soon as a reform is believed to generate losers whose identity is ex ante unknown - a feature that we see as being in accordance with many economic reforms in reality (see also Fernandez and Rodrik 1991). In addition, sampling without replacement continues to play a role when one adds aggregate uncertainty. In that setup, the revelation of winners may also lead to an upward revision of the expected aggregate number of reform beneficiaries (enhancing support for reform). However, we show that it is possible for this strategy to sow the seeds for its own destruction.

Although this learning mechanism applies to many reform types (such as land reform, the gradual abolition of subsidies/price controls, or the reduction in trade barriers), we often link it to privatization. Privatization is a good example of a reform in which learning dynamics may be important, and the choice between starting with “good” or “bad” companies arises consistently. Because the government is the incumbent owner of the firms that are to be privatized, it often has inside information on the profitability of these firms and on future policies that may benefit or harm them. This situation brings up the sequencing question for the government and the issue of learning for other agents (see, e.g., Roland 2000, chapter 2).

On a more general level, this paper develops a theory of agents who are learning from realizations that are sampled without replacement. As a result, the sampling process itself affects the distribution from which future sampling will take place. From the so-called “Monty Hall problem”, we know that this situation can give rise to counterintuitive dynamics that are not easy to understand without a formal model. Building such a model may therefore make this paper of independent interest because it may have implications for many other economic problems (a topical example is the process of revealing the identity of good and bad banks in a financial crisis).

Our results question the political feasibility of the so-called “sectoral gradualist” way of privatization. This strategy has been advocated, for example, by Kornai (1990) through his plea for the “case-by-case” approach, and it has been applied to many countries in Latin America and Central and Eastern Europe as well as to the UK during its liberalization phase. This strategy implies that one sector (or firm) is reformed after the other (cf. Berg and Blanchard (1994, 53, 63)). However, as we argue below, following such a gradual, sequential approach triggers the “sampling without replacement” effect presented previously. This may explain why practitioners have experienced political difficulties with the case-by-case strategy. Lipton and Sachs (1990, 298) note that “in almost all
countries where privatizations have been attempted, there have been major political obstacles to the case-by-case approach”. Boycko, Shleifer, and Vishny (1993, 148) state that reforms that proceed at a rather slow pace are likely to reach a deadlock. As we argue later in this paper, “spatial gradualism” (reforming one region after another, as China did by installing Special Economic Zones) can avoid the “sampling without replacement” problem. Thus, the mechanisms explored in this paper may help explain why gradual reform strategies have been more successful in China than in Latin America and Central and Eastern Europe.

The outline of this paper is as follows. We first describe various examples of reforms for which support dynamics have been counterintuitive. Next, we construct a learning model that provides an explanation for these puzzling dynamics. We then consider the question of why gradualism worked quite well for China, although this was not the case for most Latin American and Central and Eastern European countries. Finally, we conclude.

Support Dynamics for Economic Reforms: A Short History

In addition to the case of Slovakia discussed in the introduction, there are many examples of economic reforms that lost support despite their initial success (and vice versa). Stokes (2001) provides a thorough analysis of support dynamics related to various reforms. In that volume, several authors examine the public’s reactions to reforms in Spain, East Germany, Poland, Mexico, Peru, and Argentina. In her summary of the study, Stokes (2001, 25) notes that “[their] most startling result is that in every country people sometimes reacted to economic deterioration by supporting the government and its economic program more strongly. Conversely, they sometimes reacted to economic improvement with pessimism and opposition”. Similar findings are reported by the econometric studies of Remmer (1991) and Tucker (2000), who analyzed data from 12 Latin American and five post-communist countries, respectively, and reported a negative causal effect of economic performance on support for incumbents.

Stokes (2001) provides various specific examples of these counterintuitive dynamics. For example, in all three Latin American countries studied (Mexico, Peru, and Argentina), economic expansion (measured by either wage or GDP growth) was followed by pessimism about the future and opposition to the reform program. Similarly, increased real wages in Poland did not generate support for the reforms but created agnosticism instead.

With respect to the latter case, Rodrik (1995, 404) expresses surprise as well. When discussing the return to power of the former Polish communist party in 1993, he writes, “Why this should be so is not so easy to understand. [. . .] By most standards, Poland must be judged a success case”.

4. Poland had a high unemployment rate at the time, but as Rodrik (1995, 405) notes, it is not clear whether that was to blame for the deadlock. The unemployed group is too small to be decisive in national elections, and it is not straightforward that their interests are best served by policies that slow reforms.
Regarding general experiences in Central and Eastern Europe, Fidrmuc (2000, 1491) notes that “the collapse of communism occurred amidst overwhelming popular support for fundamental economic and political reforms. However, only a few years later the pendulum swung back and the reformers were voted out”. For example, Slovenia faced great difficulties in its reform process though it already had quite a few positive experiences with market forces from the past (Pleskovic and Sachs 1994). Although the 1968 Hungarian reforms began successfully, they encountered difficulties in the mid-1970s when the country underwent periods of recentralization (Qian and Xu 1993). Similarly, after the second wave of reforms following the demise of communism, the reformist Hungarian government lost the 1994 elections, and the former communist party returned to power (as in Poland and later in Slovakia), a pattern that led Kornai (2000) to conclude that the gradual reform strategy may not be feasible from a political point of view.

Latin America offers examples of countries that have had similar experiences. Puzzled by this situation, Tommasi and Velasco (1996) ask, “Why did Venezuelans riot, twice attempt to overthrow and eventually impeach a president (Carlos Andrés Pérez) who in 1990–2 brought them an average growth rate of 7.8% (the highest in Latin America), while Peruvians massively re-elected Alberto Fujimori, under whose stewardship consumption dropped by 15.3% in 1990?” Similarly, Iglesias (1994, 497–8) notes, “In my country (Uruguay), which is growing by 11.5 percent, where unemployment and inflation are down, and where reserves are up, the popularity rating of the president is 12 percent. That’s why the administration lost its bid to privatize the telephone company.”

A similar story holds with respect to India: notwithstanding the successes of the Indian liberalization policies adopted in the 1990s, India is currently struggling to implement new reforms and has turned into “a place that has fallen out of love with reform” (as stated in The Economist, 24 March 2012, 14). More generally, Sachs and Warner (1995) have documented how many countries slowed down (or even reversed) their liberalization policies in the 1960s and 1970s, though the economic performance under the more liberal regime was impressive.

These examples suggest that a successful beginning of a reform is by no means a sufficient condition for the reform to maintain majority support along the way. This observation is at odds with the conventional wisdom that a favorable start facilitates continuation.

---

5. In 1994, Luis Alberto Lacalle (of the Partido Nacional) was president of Uruguay. After taking office in 1990, he began significant economic reforms (in the sphere of both taxation and liberalization), but his initiatives later lost support (despite successes; cf. Iglesias’ quote). Subsequently, he lost the 1995 elections and was replaced by Julio María Sanguinetti of the rival Partido Colorado, who reversed many of Lacalle’s reforms.
At the other end of the spectrum are the gradual economic reforms in China. There, the government established Special Economic Zones in 1980, after which the economies of those regions began booming. In contrast to the experiences of Central and Eastern Europe and Latin America, the initial successes of these Chinese reforms led to increased support for further reforms in China (Litwack and Qian 1998; Qian, Roland, and Xu 1999, 2006).

In the next section, we develop a model that is able to rationalize the confusing support dynamics in Latin America and Central and Eastern Europe while simultaneously shedding light on the question of why the initial Chinese reform successes did not invoke such a paradoxical public response.

Support Dynamics for Economic Reforms: A Model

In this section, we describe our model. Although we frame the model in the context of economic reforms, it can also be seen as a more general model of agents who are learning from realizations that are sampled without replacement.

Our model is dynamic and contains uncertainty at both the aggregate level and the individual level. The former makes agents in the model uncertain about the total number of reform winners, whereas the presence of the latter implies that the reform will have uncertain distributional consequences. Implementing reform takes time. Consequently, reforms are completed gradually (as in reality; recall footnote 2), and agents have the opportunity to update beliefs about their chances of benefiting from the reform as it progresses over time. This belief updating process lies at the heart of our paper. The model presented by Fernandez and Rodrik (1991) emerges as a special case of our framework without learning dynamics and without aggregate uncertainty.

To build intuition for the mechanisms at play and to set the stage for our full model, we first consider a setup in which there is only individual uncertainty but no uncertainty in the aggregate. That is, in the first subsection, individuals know what fraction of the population will gain from the reform, but ex ante they do not yet know who these winners will be.

Subsequently, we add aggregate uncertainty. Then, individuals are also not sure about what fraction of the population will benefit from reform. The existence of such uncertainty has often been named as a reason to reform gradually by starting with the revelation of reform winners (cf. Roland 1994, 1164). However, as we show, it is still possible that a successful beginning to a reform will sow the seeds for its own destruction in that setup.

Without Aggregate Uncertainty

To build intuition, we first illustrate our point in a setup without aggregate uncertainty. Time is discrete, the horizon is infinite, and there is a large number of risk-neutral voters aligned uniformly between 0 and 1, indexed by \( i \). Each voter \( i \)
can be thought of as representing individuals associated with a particular firm or sector. We assume that voters are rational and forward looking.\textsuperscript{6}

Voters are faced with a reform proposal $R_g$, which is to replace the status quo. This proposal is assumed to leave everyone with a net present value payoff of 0. Reform $R_g$, in contrast, is known to benefit a fraction $\gamma > 1/2$ of the population with certainty (yielding them a net present value payoff of $S > 0$). The losing fraction $(1 – \gamma)$ is assumed to receive a symmetric negative payoff of $-S$ (where the symmetry simplifies the algebra, without loss of generality).\textsuperscript{7}

This implies that there is no aggregate uncertainty and because $\gamma > 1/2$, the reform is efficiency enhancing (according to the Kaldor-Hicks criterion) and would always be welcomed by a majority ex post.

However, the electorate faces individual uncertainty. In response to this uncertainty, voters form (potentially heterogeneous) beliefs about the effects of the reform on their personal well-being. We allow for belief heterogeneity in a discrete way: there is a fraction $\alpha_t$ that believes (or already knows) ex ante that it will belong to the group of reform winners, whereas a fraction $\beta_t$ believes/knows that it will be among the losers (with $0 < \alpha_t, \beta_t < 1/2$). The remaining fraction $(1 – \alpha_t – \beta_t)$ (which we assume to share a common prior) does not know at time $t$ whether it will gain or lose from reform. The members of that fraction will base its decision upon the expected value of the reform for them. If we sort all individuals (indexed by $i$) such that the $\gamma$ ex post winners of the reform are located on the left of the interval and the $(1 – \gamma)$ losers are on the right, we obtain the configuration shown in Figure 1.\textsuperscript{8}

Voters with $i < \alpha$ know that they are among the reform winners, whereas voters with $i > 1 – \beta$ (where the “one minus” follows from the fact that $\beta$ is measured from the right) know that they are among the losers. At the beginning of the

$6$. Whether voters are forward looking or backward looking is somewhat debated. Although the early papers on this issue report that voters are myopic and backward looking (see, e.g., Kramer 1971), more recent studies tend to find that rational forward-looking behavior dominates (cf. MacKuen, Erikson, and Stimson 1992; Fidrmuc 2000). Introducing retrospective voting would produce goodwill for the reforming government after early reform successes, somewhat similar to what arises when there is aggregate uncertainty (see Section II). As in that setup, the pace at which the various forces operate will then determine which effect eventually dominates.

$7$. With asymmetric payoffs, such as when winners obtain $G$ and losers receive $-L$ with $G > L$, reforms with $\gamma < 1/2$ could also be welfare enhancing in the Kaldor-Hicks sense (and vice versa when $L > G$). This generalization would lead to a different set of reforms that can be implemented through majority voting but do not affect the mechanics of the “sampling without replacement” effect central to this paper because the “sampling without replacement” effect refers to the availability of winning places relative to the losing ones, not to the exact magnitude of the associated gain or loss. To see this, note that replacing $S$ with $G$ (respectively $L$) in equations (3) and (4) below does not change any of our results.

$8$. We rule out partial reform (e.g., reform the winners and keep the losers under the state wing) as a desirable outcome. Clearly, a formal analysis of such an outcome would require the introduction of interactions between reformed and unreformed parts of the economy and of costs associated with the use of public funds (keeping loss-making, government-owned firms in operation is costly to society). Incorporation of this point would detract from the clarity of our core message; thus, we do not introduce these obfuscating factors here. See Murphy, Shleifer, and Vishny (1992) for a model that addresses the problems related to partial reforms. In addition, Dewatripont and Roland (1992) show that partial reforms are typically time inconsistent and therefore unsustainable.
reform, $\alpha$ and $\beta$ can be equal to 0, but this does not necessarily have to be the case; it is perfectly possible that some agents already operate under the new regime before the reform has started (for example, as a remnant of uncompleted past reform attempts) or that their identity is obvious up front. Agents between $\alpha$ and $1 - \beta$ are uncertain about their identity and do not know whether they will be a reform winner or a reform loser. Because the identity of more and more individuals is revealed as the reform progresses over time, $\alpha$ and $\beta$ become time varying and thus obtain a time index. In contrast, $g$ is a time-invariant structural parameter characterizing the reform (with aggregate uncertainty, the public’s estimate of $g$ can become time varying, though $g$ itself is fixed, which is what we allow for in the next section).

The expected value of the reform for uncertain individuals (i.e., those with $i \in (\alpha_t, 1 - \beta_t)$) equals

$$E_t \{ R_y | i \in (\alpha_t, 1 - \beta_t) \} = (\gamma - \alpha_t) \cdot S_t + (1 - \gamma - \beta_t) \cdot (-S_t). \quad (1)$$

Individuals in that group follow the decision rule

$$\delta_t = \begin{cases} 1 & \text{if } E_t \{ R_y | i \in (\alpha_t, 1 - \beta_t) \} > 0 \\ 0 & \text{if } E_t \{ R_y | i \in (\alpha_t, 1 - \beta_t) \} \leq 0 \end{cases}, \quad (2)$$

where $\delta_t$ is a support indicator that takes the value 1 if the uncertain group votes in favor of the reform and zero otherwise. Because $\alpha_t$ and $\beta_t$ are both smaller than 1/2, the decisive median voter is located in this uncertain group.9

The expected value of this uncertain group (expressed by (1)) can be negative for a wide range of parameter combinations, thereby making all $(1 - \alpha_t - \beta_t)$ uncertain individuals oppose the reform package ex ante. Because $\alpha_t < 1/2$, this implies that the reform does not enjoy majority support up front, though it would be welcomed by a majority ex post (because $\gamma > 1/2$).10

9. We follow Fernandez and Rodrik (1991) and many others in assuming that a reform is more likely to be adopted if there is a larger number of individuals in favor of it, but we use the language of majority voting for concreteness. The model can, however, also be interpreted as describing support dynamics for reforms in non-democratic countries. Decision rule (2) could then be interpreted in terms of joining an anti-reform protest or not (in that case, it might be realistic to include a “protest cost”).

10. Here, one should note that the Fernandez-Rodrik model assumes that it is not possible to compensate the losers ex post. As noted by Messner and Polborn (2004, 118), this assumption is “standard (and often even implicit) in the literature”. Given the well-known difficulties that governments face in committing to future policies, this assumption may not be unrealistic.
As in Fernandez and Rodrik (1991), the presence of individual uncertainty can thus prevent efficiency-enhancing reforms from being implemented. In particular, there are currently ex post winners blocking the reform ex ante because they do not know that they will be among the ex post winners.

Because individual uncertainty lies at the core of the problem, one may think that reducing individual uncertainty by revealing winners (i.e., increasing $a_t$ to $a_{t+1} = a_t + \Delta a_{t+1}$, bringing it closer to $\gamma$) would make a yes vote more likely. This turns out not to be true, opening an interesting perspective on voter dynamics. To see this, consider how the expected value for uncertain individuals changes with $\alpha$:

$$\frac{\partial E_t\{R_y|i \in (\alpha_t, 1 - \beta_t)\}}{\partial \alpha_t} = -S < 0.$$  \hfill (3)

Therefore, a decrease in individual uncertainty brought about by the revelation of additional winners makes individuals who remain uncertain more negative about their chances of gaining from the reform. The reason is that in the absence of aggregate uncertainty (which is added in the next section), increasing $a_t$ to $a_{t+1}$ implies that there are $\Delta a_{t+1}$ fewer gaining places left for those who remain uncertain (because the revelation of reform outcomes is an example of sampling without replacement). This makes these uncertain individuals more pessimistic about their chances of ending up as reform winners (because the revelation of winners deteriorates the quality of the remaining pool). When the median voter is located within this uncertain group, he also becomes more pessimistic.

Revealing losers, in contrast, increases the expected value of the reform for those who remain uncertain:

$$\frac{\partial E_t\{R_y|i \in (\alpha_t, 1 - \beta_t)\}}{\partial \beta_t} = S > 0.$$  \hfill (4)

At this stage, one should note that there is a wide range of values for $\alpha$ and $\beta$ where changes in uncertainty will not change the outcome of the vote. If the vote is initially “no”, then increases in $\alpha$ will only make $E_t\{R_y|i \in (\alpha_t, 1 - \beta_t)\}$ more negative, and the median voter will continue to oppose the package.

There is an intriguing possibility if the median voter initially supports the reform package. To see this, hold $\beta_t$ constant at $\bar{\beta}$ for a moment\(^{11}\) and let us investigate what happens if the government tries to complete the reform gradually by increasing $\alpha_t$ (i.e., revealing winners). If the increase in $\alpha$ is small enough, $\delta$ will remain 1, and the median voter continues to vote “yes”, pushing the overall vote in favor. However, because of the effect captured by (3), one can define a

\(^{11}\) The same argument applies, mutatis mutandis, to changes in $\beta_t$ if the median voter initially opposes reform. However, in that case, the reform cannot be started along democratic lines.
critical value for $\alpha$ (call it $\alpha^*$) such that if $\alpha$ rises above $\alpha^*$, the median voter swings around, causing a rejection of the package.\textsuperscript{12}

The critical value $\alpha^*$ is thus the point at which the median voter begins to oppose a reform that he used to support. Crossing it from below implies that the reform process stalls. Mathematically, $\alpha^*$ is defined by $E_t\{R_t\} i \in (\alpha^*, 1 - \beta)$; thus, from (1), we can derive

$$\alpha^* = 2\gamma + \beta - 1.$$  \hspace{1cm} (5)

More formally, we can now see that if the median voter initially favors the reform (i.e., $\delta_t = 1$), the total supporting fraction (given by $\Psi_t = \alpha_t + (1 - \alpha_t - \beta)\delta_t$) will remain constant at $1 - \beta$ if $\alpha$ increases to an $\alpha_{t+1} < \alpha^*$. In this case, the revelation of $\Delta\alpha_{t+1}$ additional winners does not make $\alpha$ cross the critical value $\alpha^*$. When $\alpha$ is increased, there are more individuals supporting the reform (because they have now learned that they are reform winners), but the uncertain block (which also supports the reform in this case) shrinks one-for-one with the increase in $\alpha$. On balance, total support for reform remains unaffected.

However, as more winners are revealed, $\alpha$ will eventually exceed $\alpha^*$. If this happens when $\alpha^* < 1/2$ (a condition to which we return to in the next subsection), the median voter switches sides and begins opposing the reform package that he used to support. A sudden loss of majority support for the reforming government results.

This opens the possibility of a reform that starts well (individuals involved with reformed firms/sectors turn out to be better off), but as individual uncertainty continues to decrease, the “sampling without replacement” effect captured by (3) eventually causes the median voter to swing against the package. Thus, the model produces support dynamics that are very much like the practical experiences of many reformist governments (as noted above).

Hence, once one accounts for individual uncertainty and the “sampling without replacement” effect, the conventional sequencing wisdom that one should begin by reforming firms or sectors that are most likely to benefit from reform (to boost public support) is challenged.

This conventional wisdom, to which we adhered before analyzing a formal model, is expressed, for example, in Roland (1994, 1164), who writes that “if the best firms get privatized first [. . .] the likelihood of a successful economic performance will be higher. Initial economic successes for privatized firms will enhance support for privatization and build constituencies for further reforms”. Similarly, The Economist of March 24, 2012, writes about the opposition the Cuban reform process is currently experiencing and states that to increase public support for the reform process, “Raúl Castro urgently needs to create some winners” (p. 20). However, this line of reasoning only seems to consider aggregate uncertainty and overlooks individual uncertainty and the accompanying

\textsuperscript{12} The last part of this statement, of course, assumes that $\alpha^* < 1/2$, a condition to which we will return to in the next section.
“sampling without replacement” effect. By allowing for these elements, the present paper points out that Raúl Castro may very well decrease support for his reforms even further by revealing winners.

**With Aggregate Uncertainty**

To capture the reasoning underlying the aforementioned conventional wisdom, which relies upon the existence of aggregate uncertainty, we next investigate what happens when we add such uncertainty to the model. In that case, voters also do not know the true value of \( \gamma \) (the fraction of individuals who will benefit from the reform ex post) with certainty. Instead, the public has beliefs about \( \gamma \). Let us use \( \hat{\gamma}_t \) to indicate the beginning of period \( t \) estimate of \( \gamma \). Any valuable information that becomes available during period \( t \) will lead to an updated estimate, \( \hat{\gamma}_{t+1} \) (where updating occurs via the application of Bayes’ rule).

Voters hold a prior belief about \( \gamma \) that is given by a Beta\((a, b)\) distribution. This distribution is a natural choice because it is the conjugate prior of the binomial distribution underlying the present model. Assuming a symmetric loss function (for example, the traditional quadratic loss function underlying OLS, which also has the convenient property that the point estimate for gamma summarizes all relevant information and therefore becomes the sole object of interest) then implies that for \( a = a_t \) and \( b = b_t \), the time \( t \) estimate of \( \gamma \) equals

\[
\hat{\gamma}_t = \frac{\alpha_t}{\alpha_t + \beta_t}.
\]  

Expression (6) is intuitive: \((\alpha_t + \beta_t)\) represents the total sample of outcomes we have gathered so far, whereas \( \alpha_t \) is the fraction of winners in this sample relative to the unit interval. The ratio of these two is the time \( t \) estimate of \( \gamma. \)

After revealing \( \Delta \alpha_{t+1} \) additional winners and \( \Delta \beta_{t+1} \) additional losers during period \( t \), Bayes’ rule implies that the posterior estimate of \( \gamma \), which is the prior at the beginning of period \( t+1 \), equals (see, e.g., Kvam and Vidakovic (2007, chapter 4))

\[
\hat{\gamma}_{t} = \frac{\alpha_t + \Delta \alpha_{t+1}}{\alpha_t + \Delta \alpha_{t+1} + \beta_t + \Delta \beta_{t+1}}.
\]  

From equations (6) and (7), one can confirm the intuitive notion that the application of Bayes’ rule leads to an upward revision of the expected fraction of reform beneficiaries (\( \gamma \)) when \( \Delta \alpha_{t+1} \) additional winners are revealed (and vice versa after the revelation of losers). Because this implies that beliefs about \( \gamma \) can change over time, the critical value for \( a \) \((\alpha^*)\) also becomes time varying. In particular, after plugging (6) into (5), we obtain

\[
\alpha_t^* = 2 \hat{\gamma}_t + \beta_t - 1.
\]  

Now, the key question is whether we can obtain \( \alpha_t \geq 1/2 \) before \( \alpha_t \geq \alpha_t^* \). If this is the case, the government is able to reveal that the median voter is a reform winner (which
happens when $\alpha_t$ crosses $1/2$) before this pivotal voter begins opposing the reform package (which happens if and only if $\alpha_t$ exceeds $\alpha^*_t$ while $\alpha_t < 1/2$). Subsequently, the government can complete the reform with no risk of losing majority support.

Because empirical studies such as those of Carlin and Mayer (1992), Frydman, Rapaczynski, and Earle (1993),Marcincin and Van Wijnbergen (1997), and Gupta, Ham, and Svejnar (2008) all present evidence that reforms start by revealing ex post winners, it is interesting to see what our model predicts would happen if the reform follows such a selective path. To investigate this, we make the following assumption on the sequencing within the reform:

**Assumption 1.** Sequencing is such that the reform starts by revealing ex post winners.

The reason for the presence of this selection bias can be twofold. First, it can result from a situation of asymmetric information in which the government knows ex ante who will benefit and who will lose from reform (but, as in Perotti (1995), the government is unable to transmit this information credibly to the public).

Especially in our privatization example, this assumption seems realistic because the government (as the incumbent owner of the firms that are to be privatized) has inside information on firm profitability and future policies that may benefit or harm each firm. If this government then follows the conventional wisdom and begins by reforming the ex post winners (which is often recommended to reformers in practice; cf. Roland 2000, 49), Assumption 1 materializes.

Second, in light of our application to privatization, Assumption 1 can also result from the fact that better firms tend to find buyers more rapidly (Roland, 2000, 248). This point has also been recognized by policy makers. According to Egyptian government officials in the *New York Times* of June 27, 2010, Egypt suspended its privatization program in 2009 because “most of the likely candidates had already been either privatized or dissolved, leaving hard-to-sell industries that were technologically outdated and overstaffed with ill-trained workers”.

In the Appendix, we explore the alternative case in which the government is not able to identify winners and losers in advance. That case is probably more relevant to trade reform because this type of reform does not come with a natural selection process, and it is not clear that the government knows the identity of the winners and losers up front in that setting. Then, reform outcomes are sampled randomly from the true underlying distribution. Crucially, the Appendix shows that the “sampling without replacement” effect continues to be

13. Roland (2008, 4) nicely summarizes this literature by noting, “The few studies on the determinants of privatizations suggest that the more profitable firms were privatized first, which is consistent with political economic theories of privatization where the sequencing of privatization is used to gather support for further privatization”.

14. In this sense, the government in our model is a bit like Monty Hall in the “Monty Hall problem”: he knows ex ante behind which doors the gains and losses are located. Note that the counterintuitive solution to the Monty Hall problem follows from the fact that sampling takes place without replacement.
present under random sampling. This leads to two regions in the \((a, b)\) space where the dynamics are anomalous (i.e., favorable reform outcomes decreasing support for reform and vice versa). More generally, the importance of the “sampling without replacement” effect is increasing in the tightness of the prior belief on \(\gamma\): the tighter the prior on \(\gamma\) is, the less responsive voters’ beliefs on \(\gamma\) are to news and the more dominant the “sampling without replacement” effect becomes (because the latter works independently of the tightness of \(\gamma\)’s prior). In the limit, because the prior on \(\gamma\) converges on a point, the model collapses to the one discussed in the previous subsection (without aggregate uncertainty).

Turning to the setting in which Assumption 1 holds, it is instructive to first think through what would happen if voters do not account for the selection bias and hold a diffuse prior belief on \(\gamma\) at the start of the reform (call this “time 0”). In particular, let us assume that both \(a_0\) and \(b_0\) (the fractions of winners and losers whose identities are clear ex ante) are close to zero (which minimizes the tightness of the prior). Then, Bayes’ rule implies that voters’ beliefs about \(\gamma\) are revised upward when winners are revealed \((\partial \hat{\gamma}_t / \partial a_t = b_t / (a_t + b_t)^2 > 0)\). In particular, voters’ beliefs about \(\gamma\) will quickly converge to 1 because voters only observe favorable reform outcomes and erroneously think that this is the result of random sampling from the underlying true distribution of winners and losers. This implies that \(a_t^* \rightarrow 1\) (cf. equation (8)), which allows the reforming government to reveal that the median voter is a reform winner before \(a_t \geq a_t^*\) (provided that \(\gamma > 1/2\), of course). Subsequently, the reform can be completed with no risk of losing majority support.

Therefore, when voters have a diffuse prior belief on \(\gamma\) at the start of the reform and when they do not take the selection bias into account, the government is able to complete efficient reforms gradually by revealing winners while running no risk of losing majority support.

This case, however, imposes an unrealistically high degree of naivety on voters; they think that the reform is sequenced in a truly random way and do not take into account that the government (or nature; recall our discussion following Assumption 1) starts by revealing reform winners.

Consider, therefore, the more realistic case in which the public does consider the selection bias. Then, the revelation of additional outcomes provides no valuable information; the public realizes that these draws do not come from the true underlying distribution, as a result of which, Bayes’ rule no longer leads to a revision of the prior belief.\(^{15}\)

\(^{15}\) The only thing that is revealed if \(a_t\) is increased to \(a_{t+1}\), is that \(\gamma \geq a_{t+1}\) (which was already known given that \(a_{t+1} < 1/2\), whereas voters know that the reform is efficient, i.e., \(\gamma > 1/2\)). However, under Assumption 1, this is by no means informative about how many winners are located beyond 1/2. More formally, Bayes’ rule states that \(P(A|B) = P(A)P(B|A)/P(B)\), where “A” is a particular hypothesis of interest (e.g., \(\gamma > 0.6\)) and “B” represents the new incoming data \((\Delta a_{t+1}\) or, equivalently, \(\Delta b_{t+1}\)). When sampling is such that the winners are revealed first, \(\Delta b_{t+1}\) will always be equal to 0 as long as we move to an \(a_{t+1} < \gamma\), both conditional on hypothesis \(A\) as well as unconditionally (remember that the latter case still conditions on voters knowing that \(\gamma > 1/2\)). Consequently, \(P(\Delta b_{t+1} = 0|A) = P(\Delta b_{t+1} = 0) = 1\) and Bayes’ rule implies that the posterior belief equals the prior belief (i.e., \(P(A|B) = P(A)\)). In this case, beliefs are no longer revised.
Consequently, agents cannot update their estimate of \( g \), and \( \hat{g}^t \) remains constant at \( g_0 \) \( \forall t \) (where \( g_0 \) is the exogenously given belief on \( g \) at the start of the reform). This leads to the following results.

**Proposition 1.** If the public believes that the reform starts by revealing the ex post winners and if it believes that the reform is “sufficiently efficient” (in the sense that \( \hat{g}_0 \geq 3/4 - 1/2\beta_0 \)), the reform can still be completed gradually by revealing only winners from time 0 onwards.

**Proof.** From equation (8), it follows that \( \hat{g}_0 \geq 3/4 - 1/2\beta_0 \Leftrightarrow \hat{\alpha}_0^* \geq 1/2 \). Revealing only winners (keeping \( \beta_t \) constant at \( \beta_0 \)) implies that \( \hat{\alpha}^*_t \) remains constant at \( \hat{\alpha}_0^* \geq 1/2 \) over time. This implies that the threshold \( \hat{\alpha}^*_t \geq 1/2 \ \forall t \), as a result of which the reforming government can reveal that the median voter is a reform winner before this voter begins opposing the reform (i.e., the government can push \( \hat{\alpha}_t \geq 1/2 \) before \( \alpha_t > \hat{\alpha}_t^* \)).

**Proposition 2.** If the public believes that the reform starts by revealing the ex post winners but if \( \hat{g}_0 < 3/4 - 1/2\beta_0 \), even reforms that are believed to be efficient (i.e., reforms for which \( \hat{g}_0 > 1/2 \)) can never be completed gradually by revealing only winners from time 0 onwards.

**Proof.** From (8), it now follows that \( \hat{g}_0 < 3/4 - 1/2\beta_0 \Leftrightarrow \hat{\alpha}_0^* < 1/2 \). This implies that the reform is not believed to be “sufficiently efficient” (as defined in Proposition 1). Revealing only winners (i.e., keeping \( \beta_t \) constant at \( \beta_0 \)) then decreases the expected value of the median voter via (3). Because the constancy of \( \beta_t \) again implies constancy of \( \hat{\alpha}^*_t \) (at \( \hat{\alpha}_0^* < 1/2 \)), \( \alpha_t > \hat{\alpha}_0^* \) before \( \alpha_t > 1/2 \), and majority support is lost before the reform is completed.

Hence, those reforms on which initial prior beliefs are not sufficiently optimistic (such that \( \hat{g}_0 < 3/4 - 1/2\beta_0 \)) can no longer be completed gradually by revealing only winners. This result arises even if the reform enjoys majority support at its beginning and even if the reform is believed to be efficient (in the sense that it is believed to generate more winners than losers, i.e., \( \hat{g}_0 > 1/2 \)).

The intuition for what is going on is exactly as in the previous subsection: every additional winner revealed reduces the perceived probability of ending up as a winner for those who remain uncertain. As a result, the median voter will, at some point, begin opposing the reform that he used to support.

Similar dynamics arise when we drop Assumption 1 and instead assume that reform outcomes are sampled *randomly* from the underlying distribution (see the Appendix for a discussion of this case). Then, the revelation of winners also implies that voters become more enthusiastic about the reform as they revise their estimate of the aggregate fraction of winners (\( \hat{\gamma} \)) in the upward direction. However, when the prior belief on the aggregate state is sufficiently tight, the
updating process in the aggregate dimension will proceed at a rather slow pace, and the “sampling without replacement” effect will dominate.\footnote{To see how the speed of updating is inversely related to the tightness of the prior, consider two priors with an identical mean estimate, \( \gamma = \alpha_t / (\alpha_t + \beta_t) = \alpha' / (\alpha' + \beta') \), where \( \alpha' > \alpha_t \) and \( \beta' > \beta_t \). Consequently, by the standard formula for the variance of the Beta-distribution, \( \text{Var}_\gamma(\alpha_t, \beta_t) > \text{Var}_\gamma(\alpha', \beta') \), so the prior driven by \( (\alpha', \beta') \) is tighter than that driven by \( (\alpha_t, \beta_t) \). Now consider a given \( \Delta \alpha_{t+1} \) (of equal size in both cases). Because \( (\alpha' + \beta') > (\alpha_t + \beta_t) \), equation (7) implies that this will lead to a smaller upward revision for the tighter prior driven by \( (\alpha', \beta') \). This holds for other distributions as well. Intuitively, an agent with a tighter prior is more certain that the true \( \gamma = \gamma_t \), as a result of which he revises his beliefs by less after receiving new information. Consequently, learning will occur at a slower pace for such an agent.}

Returning to the setup in which Assumption 1 does hold, revealing losers immediately ends majority support.

**Proposition 3.** If the public believes that the reform starts by revealing the ex post winners, any reform will lose majority support as soon as a loser is revealed before \( \alpha_t > 1/2 \).

The proof is intuitive and simply follows from the fact that the public expects the government to start by revealing reform winners. If a loser appears, the public thinks that all winners have already been revealed and that those individuals who are still uncertain about their identity will all be losers. If this happens while \( \alpha_t < 1/2 \), majority support is immediately lost.

**Summarizing**

Fernandez and Rodrik (1991) pointed out that welfare-enhancing reforms that would enjoy majority support ex post may not enjoy majority support ex ante because the reform winners cannot always be identified up front.

In a way, the message of this paper is more discouraging: even welfare-enhancing reforms that enjoy both ex ante and ex post majority support may still not come to completion because of the learning dynamics that are triggered through the initiation of the reform process. Revealing winners launches the “sampling without replacement” effect (as a result of which majority support will be lost at some point if initial beliefs about the aggregate dimension of the reform are not sufficiently optimistic), whereas revealing losers immediately ends support. The reforming government thus finds itself seemingly trapped and destined to lose majority support irrespective of what action it takes.

**How Can the Loss-of-Support Problems Be Avoided?**

Is there anything reformers can do to overcome these loss-of-support problems? The Chinese reform experience in particular suggests that a route exists toward successful gradual reform. After all, China also followed a more gradual path, and with quite some success. In sharp contrast to the experiences of many Latin American and Central and Eastern European countries, the initial Chinese reform successes seem to have only increased support for further reforms.
Obviously, the voting mechanism is absent in China (also recall footnote 9), but reforms there could still generate dissatisfaction and opposition, which does not seem to be the norm in China (cf. Litwack and Qian 1998; Qian, Roland and Xu, 1999, 2006). This raises the question of why the experiences with gradualism have been so different across countries.

In this respect, it is crucial to note that the Chinese gradual reform strategy differs from the Latin American and Central and Eastern European approaches. Whereas most countries in the latter regions tried to reform gradually along the sectoral dimension (which implies that one firm or sector is reformed after the other; cf. our discussion of the “case-by-case” approach to privatization in the introduction), China reformed gradually along the spatial dimension. In particular, China first introduced market forces in 1978 in the inland province of Sichuan (which was the first province to abolish collective agriculture and begin state-owned enterprise reform) and in the coastal province of Guangdong in 1980.

By reforming gradually along the spatial dimension, Chinese policy makers enabled the Chinese public to learn about the effects of new policies by looking at outcomes in reformed regions. Of course, the citizenry will only find the information generated via the spatial dimension useful if those regions are believed to be informative to the rest of the country. Here, China had an advantage over many other countries. As noted by Qian, Roland, and Xu (2006, 394), the Chinese economy is organized along territorial lines. This implies that its regions (such as Sichuan) are rather self-contained and relatively representative of the Chinese economy as a whole. As Démurger et al. (2002) argue, this is the result of a conscious decision made by Mao Zedong. In addition to the two key principles of Soviet development (common ownership and central planning), Mao added a third principle: regional economic self-sufficiency. This principle required each region to be self-sufficient, not only in food production but also in industrial goods. The Soviet Union did not adhere to this principle at all; their ideology called for an organization of the country along industrial lines with high degrees of industrial concentration (Qian and Xu 1993). Consequently, each Soviet region was much more specialized, dependent on other regions, and less representative of the Union as a whole.

Because of this, China had (in contrast to, for example, Russia) the possibility to start reforms by taking informative samples of small mass (in the form of certain regions) and using them to show the public where the gains and losses of the proposed reform were likely to occur. In particular, the coastal Special Economic Zones were instructive to inhabitants of other coastal regions (with approximately one-third of the total Chinese population living near the coast), whereas Sichuan fulfilled a similar role for inland areas. This strategy reduces individual uncertainty about the distribution of gains and losses (the root of all problems), but - crucially - insulates the main part of the country from the “sampling without replacement” effect.17

17. Obviously, this effect will be present within the reform region itself, but if that is only a region of relatively small mass, then the reform runs no risk of losing country-wide majority support.
For the main part of the country, this sampling strategy does not affect the distribution from which future sampling will take place because it is a form of sampling from a different, smaller urn (where the distribution of balls in this smaller urn is taken randomly from the large urn, the latter representing the main part of the country, which remains untouched in this sampling strategy).

An alternative way to think about this is by conceiving a model that has two dimensions: a sectoral one (as in our model presented above) and a spatial one. For a country that is perfectly diversified, these two dimensions are orthogonal to each other, as a result of which reforming gradually along the spatial dimension does not trigger the “sampling without replacement” effect (which exists along the sectoral dimension). When there is some correlation between the two dimensions (as is the case when certain sectors are concentrated in particular areas), orthogonality no longer holds, and sampling without replacement begins to play a role again. In the extreme situation where there is a perfect correlation between “sectors” and “space”, the two dimensions merge and we are back in our original one-dimensional setup where any gradual reform strategy launches the counterintuitive support dynamics.

For this reason, a spatial reform strategy would not have been a viable option for Russia: over there, reform outcomes in one region were not only less relevant to those in other regions (due to the higher degree of spatial heterogeneity) but the higher degree of industrial concentration would have also given a “sectoral flavor” to any spatial reform strategy. After all, if certain sectors are concentrated in certain areas, reforming one area is equivalent to reforming one sector. Then, “sampling without replacement” would re-enter the story.

For the spatial strategy to work, it is crucial that agents who know that they will be among the winners (i.e., those with $i < \alpha_t$) cannot self-select into the reformed regions because doing so would imply that the zone becomes less instructive to the relevant other parts of the country. To continue the urn analogy, the smaller urn needs to be isolated from the larger one. Interestingly, this is precisely what the Chinese “hukou” system (which restricts the mobility of citizens within

18. Thanks to our referees for putting us on track for this interpretation. Formally developing such a two-dimensional model goes beyond the scope of this paper but could be an interesting avenue for future research.

19. The “sampling without replacement” effect exists due to heterogeneity (the existence of winners and losers), but in a country that is perfectly diversified, all regions are alike in terms of economic structure (each district is a miniature version of the country as a whole), so there is no heterogeneity along the spatial dimension.

20. As Dani Rodrik pointed out to us, the spatial strategy could be applied to reforms that entail only one particular sector (“sector X”). The “sampling without replacement” effect would then operate within sector X (because cross-sectoral heterogeneity no longer plays a role). In the case of one-sector reforms, governments could get around this effect by first transforming a region of small mass that contains a representative sample of sector X firms into a Special Economic Zone. Subsequently, the sector X firms inside the Zone could be used to reduce individual uncertainty by giving other firms in the sector outside the Zone an idea of what the reform would do to them. The Special Economic Zones in Malaysia and Mauritius seem to have fulfilled such a role successfully for the electronics and apparel sectors, respectively (Auty 2011).
China) achieves. Therefore, although one could debate the fairness of this system (just like one could debate the fairness of mobility restrictions between different countries), it does seem to play an economic role in the Chinese reform process.

Finally, this view of Special Economic Zones shines a new light on their raison d’être. In a static setup, Hamilton and Svensson (1982) show that Special Economic Zones are actually welfare decreasing in a second-best world where the suboptimal regime continues to apply outside the zone. This raises the question of why governments bother installing them in the first place. In this respect, the present paper argues that Special Economic Zones could produce large dynamic gains because they can facilitate the implementation of reforms that bring the entire country closer to the first-best.

Our results thus indicate that countries can ease their reform process if they have the possibility to begin the reform by first implementing it in a region of small mass that is instructive to the rest of the country (because such an action decreases individual uncertainty). In this sense, there is an important difference between sectoral and spatial gradualism. This difference may be key as to why the gradual reform strategy has worked for China but has worked much less for many other countries.

**Conclusion**

In this paper, we have modeled the learning process surrounding economic reforms when there is both aggregate and individual uncertainty. The process of revealing reform outcomes entails sampling without replacement. We have shown that this implies that the revelation of winners early in the reform process makes those who remain uncertain about whether they will gain or lose from reform more pessimistic about their chances of ending up as reform winners. This channel can be so strong that it can even induce the median voter to start opposing a reform that he used to support - in which case the reforming government loses majority support. As a result, even reforms that enjoy both ex ante and ex post majority support may still not come to completion. The conditions under which such a destructive interaction between rational learning and political support will occur are relatively mild. As soon as one combines the presence of individual uncertainty with rational belief updating, the “sampling without replacement” effect kicks in.

This situation challenges the conventional wisdom that sequencing should be such that favorable reform outcomes are revealed first. Instead, our model illustrates that a reform strategy based on revealing winners first may backfire. The reason is that such an approach leads to a deterioration in the quality of the remaining pool, triggering reform fatigue in spite of the successes of those firms that have already been reformed. Strikingly, this is consistent with the puzzling experiences that many reformers have had in practice. There are numerous examples of reforms that began while enjoying majority support but subsequently lost this support even though they were progressing in a successful way.
We have also outlined a strategy that is able to overcome the problems related to the learning process. In particular, if a country happens to contain individual regions that are instructive to the rest of the country, the “sampling without replacement” effect can be avoided by reforming gradually along the spatial (rather than along the sectoral) dimension. This could explain the success of the “Special Economic Zone” approach to reforms taken by China.

On a more general level, this paper has developed a theory of agents who are learning from realizations that are sampled without replacement, which may have other applications as well. In many environments, the distribution from which future sampling will take place is not static and time invariant; it changes over time, often as a result of past sampling actions (the process of authorities revealing the identity of good and bad banks in a financial crisis is a clear example). From the so-called “Monty Hall problem”, we know that this can give rise to counterintuitive dynamics that are not easy to understand without formalism (whereas a formal model clarifies things significantly). Applying the concepts developed in this paper to problems that entail sampling without replacement might therefore be a fruitful avenue for further research.21

Appendix

Although there are certain reform types (such as privatization) for which the government is likely to have an ex ante idea about where the gains and losses of the reform will be located, there are also cases in which the reforming government does not have such information. This Appendix therefore explores the properties of our model when we drop the assumption that the reform is sequenced in a non-random, selective way.

Suppose that the government cannot identify the reform winners and losers up front. In addition, we assume that there is no natural selection process that could lead to a non-random sequencing of events. Consequently, reform outcomes are sampled randomly from the true distribution. Moreover, we assume that the public believes that these outcomes are sampled randomly, as a result of which they perceive new observations to be informative and apply Bayes’ rule to update their estimate of $\gamma$ (the aggregate fraction of winners) in response to new information.

Because we feel that this case may deserve a closer study in its own right, we leave a full analysis for future work, but we present some main results in this

21. In this respect, it was pointed out to us that the mechanisms underlying this paper also relate to those in Konrad (2004, although he does not establish the link to sampling without replacement, nor does he allow for aggregate uncertainty). However, that paper addresses strategic rather than dynamic issues and shows that a party that is campaigning to implement a particular reform may obtain an incentive to point out ex ante who will be a reform loser (a phenomenon that Konrad refers to as “inverse campaigning”).
Appendix. In particular, we show that the “sampling without replacement” effect discussed in the main text continues to in this alternative setting. The core of the model is unaffected, and the critical value for $\alpha$ ($\alpha^*$) is still given by

$$\alpha^*_t = 2 \frac{\alpha_t}{\alpha_t + \beta_t} + \beta_t - 1. \quad (A.9)$$

Using expression (A.1), one can analyze how the distance between $\alpha_t$ (the fraction of sure-winners) and $\alpha^*_t$ (the cut-off level for $\alpha_t$ above which the median voter starts opposing the reform) varies with the revelation of additional winners and losers. In particular, it holds that

$$\frac{\partial (\alpha^*_t - \alpha_t)}{\partial \alpha_t} = \frac{2\beta_t}{(\alpha_t + \beta_t)^2} - 1. \quad (A.10)$$

Here, the first term shows that the revelation of winners pushes up $\alpha^*_t$ (because it leads to an upward revision of the expected fraction of reform winners $\hat{\gamma}_t$ through application of Bayes’ rule), whereas the second term (“\(-1\)”) indicates that the revelation of winners simultaneously makes those who remain uncertain more pessimistic about their individual chances of ending up winners. In particular, this term reflects the fact that revealing a reform outcome is an example of sampling without replacement. From (A.2), one can derive that as long as

$$\alpha_t > \sqrt{2\beta_t} - \beta_t, \quad (A.11)$$

the “sampling without replacement” effect dominates. Under condition (A.3), the public’s estimate of $\gamma$ increases less than one-for-one with $\alpha_t$ (mathematically, $\partial (\alpha^*_t - \alpha_t)/\partial \alpha_t < 0 \Leftrightarrow \partial \hat{\gamma}_t/\partial \alpha_t < 1$), and the median voter becomes more pessimistic as favorable reform outcomes are increasingly revealed. Hence, under this condition, the revelation of additional winners produces an increase in $\hat{\gamma}_t$ that is insufficient to compensate for the fact that sampling occurs without replacement.

Similarly,

$$\frac{\partial (\alpha^*_t - \alpha_t)}{\partial \beta_t} = \frac{-2\alpha_t}{(\alpha_t + \beta_t)^2} + 1 \quad (A.12)$$

captures the same two effects for the revelation of losers. In this case, the median voter becomes more optimistic when additional losers are revealed (i.e., $\partial (\alpha^*_t - \alpha_t)/\partial \beta_t > 0$) as long as

$$\beta_t > \sqrt{2\alpha_t} - \alpha_t. \quad (A.13)$$
Conditions (A.3) and (A.5) yield two regions of \((\alpha, \beta)\) combinations, displayed as the shaded areas in Figure 2, where one can characterize the learning dynamics as “anomalous”. That is, in region A3, good reform outcomes decrease support for the reform, whereas the revelation of bad reform outcomes increases support for reform in region A5.

Now, one can ask whether the government is able to complete the reform without a loss of majority support along the way. Because the sequencing of the reform is random in this case (because the government (or nature) is no longer able to select the winners up front), it is no longer possible to analyze this question analytically. Instead, one would have to simulate the reform process, and the answer to the question would depend upon the amount of time a typical simulation spends in the shaded areas of the state space. Because we feel that this issue deserves a full discussion in its own right, we leave this for future work.

The main point to take away from this Appendix is that the “sampling without replacement” effect continues to be present when reform outcomes are revealed in a truly random fashion. This leads to two regions in the \((\alpha, \beta)\) space where the support dynamics can be characterized as “anomalous”.

REFERENCES


A Helping Hand or the Long Arm of the Law? Experimental Evidence on What Governments Can Do to Formalize Firms

Gustavo Henrique de Andrade, Miriam Bruhn, and David McKenzie*

We conducted a field experiment in Belo Horizonte, Brazil to test which government actions work to encourage informal firms to register. We find zero or negative impacts of information and free cost treatments and a significant but small increase in formalization from inspections. The local average treatment effect estimates of the inspection impact are larger, providing a 21 to 27 percentage point increase in the likelihood of formalizing. The results show that most informal firms will not formalize unless forced to do so, suggesting that formality offers little private benefit to these firms. JEL codes: O17, O12, C93, D21, L26

Spurred by the work of Hernando de Soto (1989) and the World Bank/IFC’s Doing Business project, governments around the world have spent much of the past decade extending a helping hand to informal businesses by trying to make it cheaper and less burdensome to formalize. Since 2004, 75 percent of countries have adopted at least one reform to make it easier to register a business (IFC 2009). However, despite these efforts, the majority of firms in most developing countries remain informal. Studies that have examined the impact of these regulatory reforms find that much of the action comes from increases in the entry of new firms rather than from the formalization of existing firms (e.g., Klapper

* Miriam Bruhn is a Senior Economist in the World Bank’s Development Research Group. Her email address is mbruhn@worldbank.org. Gustavo Henrique de Andrade is a Public Policy Specialist in the State Government of Minas Gerais. His email is gustavo.andrade@planejamento.mg.gov.br. David McKenzie (corresponding author) is a Lead Economist in the World Bank’s Development Research Group. His email address is dmckenzie@worldbank.org. We thank Priscila Malaguti, Arianna Legovini, Leticia Silva Palma, Milla Fernandes Ribeiro Tangari, João Luiz Soares, and Renato Braga Fernandes for their help in developing and implementing this project and the editor, two anonymous referees, and participants at seminars at UCL/LSE, Warwick, DFID, Princeton, and the World Bank for helpful comments. We are grateful for funding from the Knowledge for Change Trust Fund and from DFID as well as to the State Government of Minas Gerais, which funded the baseline data collection and collaborated on this project. All opinions expressed in the paper are those of the authors and do not necessarily represent those of the institutions to which they belong.
Policymakers worry that a large stock of informal firms will result in a loss in tax revenue, unfair competition for formal firms, and a culture of informality (Perry et al. 2007; Levy 2008). Although policymakers and researchers have devoted attention to reducing the costs of formalizing, much less attention has been given to the issue of increasing the costs of remaining informal. The most obvious way to raise these costs is to use the long arm of the law to increase the enforcement of existing regulations. Yet, to our knowledge, there is very little empirical evidence on whether enforcement attempts can induce firms to register or whether, instead, they cause informal firms to close down and prevent other firms starting up. There is a large body of related literature in developed countries showing that increases in the probability of detection and enforcement lead to an increase in tax compliance and that other factors, such as a household’s sense of moral or social obligation, are also important (e.g., Alm et al. 1992; Andreoni et al. 1998). More recent non-experimental studies in developing countries have found evidence that the degree of enforcement matters for labor informality (Ronconi 2007; Almeida and Carneiro 2012), but we are aware of no reliable evidence on the impacts of enforcement on firm informality.

We conducted a field experiment with the state government of Minas Gerais in the city of Belo Horizonte in Brazil to test which government actions work to encourage informal firms to register. Brazil began a process of simplifying firm registration in 1996 with the introduction of the SIMPLES tax system, which consolidated multiple taxes and contributions into a single payment and reduced the tax burden on small firms. Within Minas Gerais, the Minas Fácil service was started in 2005 with the purpose of additionally reducing the number of procedures and time needed to start a business. Minas Fácil is a one-stop-shop system where firms obtain municipal, state, and federal tax registrations simultaneously instead of having to request these from separate offices. However, despite these efforts, survey data from 2009 revealed that 72 percent of firms were still informal. As a result, the state government wanted to test several competing mechanisms for reducing formality.

A listing survey was used to identify potentially informal firms, which were then randomized into four treatment groups and a control group. Survey data revealing a lack of knowledge about how to formalize motivated the first treatment, which provided information about how to register by means of a glossy brochure and a dedicated helpline. A second treatment coupled this information with an exemption in the registration fees and free use of mandatory accounting services for a year to test whether reducing registration costs would induce

---

1. An exception is Mullainathan and Schnabl (2010), who find that formalization of existing informal businesses accounts for approximately 75 percent of the increase in the number of newly licensed firms in Lima, Peru, after a simplification of municipal licensing. Even then, however, the numbers involved suggest that the vast majority of informal firms chose not to formalize.
formalization. The third treatment randomly assigned municipal inspectors to firms to see whether increased enforcement would encourage firms to formalize. The final treatment consisted of having a neighboring firm visited by an inspector to test whether there was a spillover impact of inspection on the formalization behavior of other firms.

We find that efforts to help firms formalize by giving them information and by reducing the initial cost to zero along with offering a free accountant resulted in an increase in knowledge about the role of accountants in the formalization process but did not lead firms to formalize. In fact, this approach resulted in a small reduction in firms registering through a separate formal category – that of individual entrepreneur – which the information campaign did not target and for which the eligibility criteria were relaxed during the course of our study. Moreover, firms that were assigned to either of these two treatments expressed less trust in the government in our follow-up survey.

In contrast, assigning firms to receive a visit from a municipal inspector did result in an increase in municipal registration, although the impact was much less than anticipated by the government, with only an additional 3 percent of those assigned to treatment formalizing. This low rate is due to the inspectors finding some firms closed, not finding others, and from some firms in our sample already being formal to start with. An instrumental variables estimate suggests that the impact of actually receiving an additional inspector visit is much higher – resulting in a 21 to 27 percentage point increase in registration. We find no evidence of spillovers on neighboring firms, perhaps due to the relatively low increase in inspections and to many firms saying they do not communicate very much with neighboring businesses.

The papers that are most closely related to ours are two recent randomized experiments that test the role of “carrots” in inducing informal firms to register. De Mel et al. (2012) found no significant impact of information alone in getting firms to register with the tax authority in urban Sri Lanka, but they found that many firms are willing to register when offered money to do so, although formalizing does not seem to benefit the performance of most of these firms. Alcázar et al. (2010) and Jaramillo (2009) offered firms in Lima, Peru, information and reimbursement of direct costs to encourage municipal registration (which was separate from federal tax registration). They found that approximately one-quarter of those treated registered. This larger impact is consistent with municipal registration imposing fewer costs on firms than tax registration and, potentially, with municipal enforcement being higher.² Our paper builds on these studies by offering information and free-cost “carrots” in a context in which simplification has recently occurred but in which registration costs and complexity still remain much higher than in the Sri Lankan and

² An information-only intervention designed to encourage registration in Bangladesh also found a zero effect of information (De Giorgi and Rahman 2013). Bruhn and McKenzie (2014) provide a more detailed review of the literature.
Peruvian cases, and by also testing simultaneously the role of “sticks” in the form of inspections.³

Other related literature looks at the impact that formalizing has on firms. In addition to the experimental work of de Mel et al. (2012), there are several non-experimental studies that examine this question. Fajnzylber et al. (2011) and Monteiro and Assunção (2012) both analyzed the impact of the SIMPLES program on firms. These authors find that firms created after the reform invested more, were larger, and were more likely to operate in a permanent location than firms created just before the reform. However, it is unclear how much of this is an impact of formalizing versus a difference in the selection of which firms formalize. McKenzie and Sakho (2010) used an instrumental variables strategy in Bolivia based on distance to tax offices and find that some firms in Bolivia that face high costs of formalizing would gain on net from registering for taxes, whereas other firms would lose from doing so and appear to be rationally informal. Taken together, these studies imply that many informal firms would not benefit from becoming formal and are thus consistent with our results that information and reducing the costs of formalizing are not enough to induce formalization.

The remainder of the paper is structured as follows. First, we describe the process of becoming formal as a small firm in Belo Horizonte, the information firms have about the process of formalizing, and what firms see as the costs and benefits of becoming formal. We then describe our interventions and subsequently outline the data used to evaluate their impact and the experimental design. Next, we provide the results of the interventions on the rate of formalization among informal firms and conclude with a discussion of implications for policy and further studies in this area.

**Context and the Process of Formalizing**

Belo Horizonte is the capital city of the state of Minas Gerais in Brazil and has a city population of almost 2.5 million, with 5.5 million in the official metropolitan area (the third largest in Brazil after São Paulo and Rio de Janeiro). A 2009 survey by the Brazilian statistical agency IBGE along with government records was used by SEBRAE, the government agency for supporting micro and small businesses, to estimate that Belo Horizonte had a total of 561,310 businesses, of which 402,744 were informal (72 percent).

*Registering a Microenterprise in Belo Horizonte*

The Complementary Federal Law 123 defines micro-enterprises as firms with annual revenues up to R$360,000 (US$177,000),⁴ provided they are not a

---

³ The cost of registering in Sri Lanka was approximately US$10 and took between two and eight days, with the median firm in the de Mel et al. study below the threshold for income taxes. The Peru cost was approximately $45, and firms were not liable for taxes after municipal registration. This compares to a cost of approximately US$180 and ongoing taxes of at least 5 percent of income for the typical firm in this study.

⁴ One US dollar was approximately 2 Reais during the period of our intervention.
subsidiary of another firm. Microenterprises that meet several other conditions (the key ones being that they do not have a foreign owner or partner and are not in certain sectors such as financial services, consulting, alcohol or tobacco, or transportation) are eligible to register their businesses formally under a national simplified taxation system called SIMPLES. The SIMPLES regime combines several ongoing tax and contribution payments into a single payment (including employee taxes and the state sales tax), but it does not simplify the registration process itself. In addition, at the time our study began, enterprises with one or fewer employees and R$36,000 or less in annual revenues could instead register as individual microentrepreneurs (MEIs). This is an even simpler tax status for proprietorships in which only a fixed amount per month is paid for all taxes.\(^5\) Eligibility for this category was changed after our intervention had begun; the eligibility threshold was raised to R$60,000 after a law change in September 2011.

The state government created a unit called Minas Fácil in June 2008 to simplify the registration process. Registration under this new system involves registering at the federal, state and municipal levels through a single process (presented in Appendix S1, available at http://wber.oxfordjournals.org/). Many steps in the process are online, and the entire process is estimated to take seven days for an average firm. The key documents obtained are federal tax registration, evidenced by obtaining a Cadastro Nacional de Pessoas Jurídicas (CNPJ) number; state-level registration with the Chamber of Commerce (JUCEMG); and a municipal license (Alvará de Localização e Funcionamento, or ALF). The initial cost of registration is R$236 for a sole proprietorship and R$320 for a limited liability company.

The annual costs of being formal include a sanitary tax (R$53.40–106.76, depending on activity); the TFLF, a municipal inspection tax (R$77.26) that firms need to pay within 30 days of formalizing and again at the beginning of each year; and a revenue tax. The revenue tax is a flat rate of R$51.65 per month for individuals who qualify as a MEI; otherwise, SIMPLES ranges from 4 percent to 8.21 percent depending on the revenue level and industry.\(^6\) In addition, it is mandatory for formal firms with two or more employees to use an accountant, who must prepare cash flow and accounting statements each month, ensure the firm makes the monthly tax payments, and submit a form each year to the federal tax authority. In Belo Horizonte, accountants charge an average of R$300 per month for this service. Accountants are not required for MEIs with one or fewer workers and revenue under the MEI revenue threshold.

We use our baseline survey data to calculate the estimated costs in the first year from formalizing as a percent of baseline annual profits. Figure 1 shows the distribution of the cost of formalizing. The cost of formalizing is 10.7

---


\(^6\) The first R$180,000 in annual revenue is taxed at 4 percent for firms in commerce, 4.5 percent for firms in industry, and 6 percent for firms in services. The next R$180,000 is taxed at 5.47 percent for firms in commerce, 5.97 percent for firms in industry, and 8.21 percent for firms in services. This tax includes a number of taxes such as income tax, contributions to social security, and employer pension contributions.
percent of baseline profits for a firm at the 25th percentile of firms in our sample, 15.6 percent for a firm at the 50th percentile, and 27 percent for a firm at the 75th percentile. Furthermore, 1.2 percent of firms have a cost of formalizing that exceeds 100 percent of baseline annual profits. Thus, the cost of becoming formal is large for some firms and greatly exceeds what owners would pay in personal income tax if this were taxed as wage income.7 These calculations assume that formal firms report their full revenues and all of their workers; in our surveys, firms say that they think firms typically report only 50 percent of revenues for tax purposes. Firms may also underreport workers to escape the requirement of having an accountant. A firm in our sample reporting no more than one worker and reporting only half its revenues for tax purposes would face a median annual cost of being formal of 8 percent of its annual profits.

**Figure 1.** Distribution of First-Year Costs from Formalization as Percent of Baseline Profits

Source: Authors’ analysis based on survey data described in the text.

Are Firm Owners Well Informed about the Process of Formalizing?

Our baseline survey reveals that most of the interviewed informal firms lack key information about the process of formalizing, that there is substantial

7. By way of comparison, personal income tax rates are 0 for income below R$20,529, 7.5 percent for income between R$20,529 and R$30,767, and 15 percent for income between R$30,768 and R$41,023, with the highest rate of 27.5 percent applying to income over R$51,250. The mean annual profits in our sample are R$24,255, so they would have an average tax rate of only 1.1 percent if they were taxed as wage income.
heterogeneity in their beliefs and that most think formalizing is more time-consuming and costly than it is in reality. Only 46 percent of firm owners claim that they know what is needed to register, and only 19 percent know that the Junta Comercial (Chamber of Commerce) is where firms need to go to register. The mean (median) time firm owners think it takes to register after all documents are provided is 51 days (30 days), whereas in practice, the average time is seven to nine days. Almost 30 percent have no idea of the cost of registering, and among those who estimate the cost, the mean estimate is R$1,304, and the 90th percentile of R$2,500 is more than 10 times the 10th percentile of the estimated cost of R$200. As described above, the actual upfront cost of registering is only R$236, or R$366 if one includes the sanitary and municipal taxes due within 30 days of registering. The mean (median) estimated tax rate is 22 (20) percent, compared with the actual tax rate of 4 to 8 percent.

The baseline survey asked firm owners open-ended questions about what they saw as the main benefits and costs of formalizing. The main benefits mentioned were the ability to open a bank account in the business name (51 percent), a better reputation for the business (47 percent), reduced risk of being fined (44 percent), ability to get business loans (43 percent), and the ability to sell to other firms that require registration (39 percent). Only 13 percent said that they saw no advantages. The main disadvantages mentioned were the initial costs of registration (62 percent), having to pay taxes (58 percent), having to pay for an accountant (34 percent), and the process of registering taking too much time (32 percent).8

The baseline survey also asked firms if they had received an inspection visit from various types of inspectors in the past year. Thirty-two percent of firms responding to the baseline survey had received a visit from the municipal inspector (typically to check whether they had paid for a sign outside), 5.5 percent from a state tax inspector, and 3.1 percent from a federal tax inspector. The other main form of inspection was sanitary inspections, which 20.1 percent had received. Only two percent reported having paid a fine for being informal; the mean (median) fine was R$2,340 (R$600), with the majority of these fines paid to the municipality. Note that these data come from surviving firms and may understate the rate of inspection and fine receipt among all informal firms if many firms that are inspected or fined close as a result.

Interventions

The context is thus one of pervasive informality, despite the introduction of the simplified taxation program SIMPLES and the efforts by the Minas Facil unit to streamline the registration process. Given this context, Descomplicar, a unit

---

8. There is also an option value to remaining informal because firms can always decide to formalize later when asked by an inspector or when a law changes, whereas it is much harder to de-formalize. Firms did not list the loss of this option value as a disadvantage, but this may have been harder for them to express.
within the state government of Minas Gerais that has the mandate to simplify relations between citizens, firms and the state, worked with the World Bank to test various mechanisms that could be used to induce more firms to formalize under the existing system in place. The focus was on trying to target firms that fell under the eligibility criteria for SIMPLES, which, at the time of design, was for revenues in the range of R$36,000 to R$240,000 or having two or more workers if revenues were below this level.9

The following three interventions were designed, along with a fourth, indirect treatment and a process to test them experimentally (described in the next section).

**Communication Treatment**

Given the lack of information many firm owners have about the process, time, and costs of formalizing, the first intervention considered was an information treatment. An attractive and colorful brochure titled “formalization of enterprises” (see Appendix S2 for example pages) was designed by professional marketing staff. This 18-page brochure included (i) information on the advantages and importance of formalizing, explaining benefits such as the availability of lines of credit, the ability to participate in tenders and public bids, increased credibility, and compliance with social obligations; (ii) the disadvantages of being informal, including the risk of seizure of goods and application of fines, difficulty dealing with medium and large suppliers and customers, limited business growth prospects, the inability to practice judicial recovery, and the inability to obtain financing due to a lack of accounting records and formal status; (iii) explaining how firms can tell if they qualify as a microenterprise; (iv) discussion of opportunities in business procurement and the simplicity of selling to the Government of Minas Gerais; (v) opportunities for lines of credit for small formal businesses through the state development bank; (vi) the importance of working with an accountant; (vii) how to calculate taxes; and (viii) the 10 steps needed to register for SIMPLES (see appendix S1) and a telephone number firms could call for help.

Firms selected for this intervention were given this brochure in person by a trained interviewer from the survey company Sensus Pesquisa e Consultoria. Descomplicar staff trained these interviewers on the content of the brochures and on an accompanying short speech explaining the content. Firms that stated they were formal and that could produce a federal taxpayer number (CNPJ) were not given the brochures. Those that stated they were formal but could not document this assertion were also given the brochure.

**Free-Cost Treatment**

The second intervention combined the information brochure given in the communication treatment with an effort to eliminate as many of the costs of

---

9. Note the upper threshold was raised to R$360,000 after a law change in late September 2011, while at the same time, the revenue threshold for MEI registration was raised from R$36,000 to R$60,000.
registering as possible. As part of this intervention, Descomplicar made an arrangement with its counterparts in the other agencies involved in registration for all registration fees to be waived for the firms selected for this pilot intervention. This arrangement included waiving the JUCEMG registration fee and municipal license fees as well as paying the first year’s sanitary tax and municipal inspection fee, which are due within 30 days of registering. The fees waived amounted to between R$366 and R$504 (US$183–250) depending on the type of firm. In addition, an arrangement was made with the local accountants’ association whereby 50 accountants would be available to provide one year of free accounting services to these firms. This service has an effective value of R$3,600 given the prevailing cost of accounting services and the mandate for certain types of firms to use an accountant. Thus, firms participating in this treatment and being formalized through this offer would pay no initial registration fees, and the only cost of formalizing in the first year would be their SIMPLES taxes. This offer was again delivered in person to firm owners by trained enumerators, with a phone number of a government office firm owners could call for further information or help. Firm owners were given 90 days to take advantage of the offer.

**Inspector Treatment**

In addition to informing firm owners about how to register and making it cheaper for them to do so, the other main instrument governments can use to influence the behavior of informal firms is enforcement. As evidenced by our baseline data, the most common source of enforcement comes from municipal inspectors, which is also the typical pattern in other countries. In countries where municipal registration is separate from tax registration, the result is that many firms tend to be registered with the municipality but not with other levels of government. In Minas Gerais, because registration is a streamlined process requiring registration at all three levels, it is possible that municipal inspection may also result in full formalization. Note that before Minas Fácil was introduced in 2005, municipal, state, and federal registration processes were not linked in Minas Gerais, so it was possible for firms that registered pre-reform to have one type of registration but not the others.

The Prefeitura de Belo Horizonte (PBH) is the authority in charge of municipal inspections within the city of Belo Horizonte. The PBH has approximately 100 inspectors divided across nine semi-autonomous subregions, each with their own decision-making processes about which firms to inspect. These inspectors perform two types of inspections. The first is to check whether firms have a permit to display a sign outside the firm (*placa*). Second, the inspectors can conduct visits that request proof that a firm has a current municipal license (ALF), which expires every five years, and can check whether a firm has a CNPJ (tax registration). Firms that are lacking the ALF receive and notification and are given 30 to 45 days to formalize, at which time the inspector returns. If the firm is still lacking municipal registration, the inspector fines the firm owner (the fine
amount varies depending on the area of the premises)\textsuperscript{10} and closes the firm.\textsuperscript{11} If firms can prove they are in the process of registering, they can receive more time. The inspectors do not have the power to fine firms for not being registered with the state or federal authorities, but they can threaten to report un-registered firms to these authorities. In practice, this usually does not occur. When a firm applies for a municipal license following the inspection, it should technically also receive the state and federal licenses because all three registration processes are now integrated.

The third intervention consisted of giving these inspectors a list of selected firms to receive this second type of inspection, which involves requesting proof of the municipal license (ALF) and checking whether the firm has a CNPJ.

\textit{Indirect Inspector Treatment}

The cost effectiveness of inspection as a means of getting firms to formalize depends in part on whether there are spillover impacts from inspected to non-inspected firms. Our final treatment is therefore an indirect one, whereby a firm does not receive an inspection but firms very closely located to it do receive an inspection. The next section explains our experimental design to measure these spillovers.

\textbf{DATA AND EXPERIMENTAL DESIGN}

To experimentally test these interventions, we needed a sample of informal firms. However, since no recent sample frame of informal firms was available, we had to construct one through a listing exercise. The presence of the inspector treatment added a complication to this listing process for ethical and survey-response reasons. In particular, if a firm owner voluntarily provided information about the firm’s formality status in an interview, it may not be considered ethical to then use this information to potentially assign an inspector to visit them.\textsuperscript{12} Even if it were considered ethical (because the government has a right to ask firm owners about their formality status and a right to conduct inspections), we were still concerned that individuals who were interviewed in a baseline survey and then received an inspection may be unwilling to respond to a follow-up. Therefore, a listing stage was performed that did not involve talking to the firm owner.

\textsuperscript{10} Note that these fines occur only if the owner fails to respond to the request to formalize after an inspector visit. There are no back taxes or fines for having operated informally before the first inspection visit.

\textsuperscript{11} Closing the firm involves the inspector physically shutting the door of the firm, saying the firm is closed, and then coming back three times to check that the firm is still closed.

\textsuperscript{12} Our inspection experiment here differs from almost all other experiments we are aware of, in which the treatment given to participants is something they privately want. Here, inspections are likely socially good but privately undesired. McKenzie (2013) discusses this issue in the context of this experiment in more detail.
Listing Survey

The survey firm Gauss Estatística & Mercado was hired by the Minas Gerais government through a public procurement process to undertake a listing survey in 600 census blocks in Belo Horizonte. Appendix S4 describes how these census blocks were selected. Listing consisted of enumerators visiting every firm operating out of a fixed building in the census block. It excluded individuals operating informally on the street because our interest was in larger informal firms, and it excluded transportation firms because the rules for formalizing are different for them. Enumerators recorded basic information about the firm that could be observed without talking to the firm owner, including the full street address, the business sector, the “fantasy name” of the firm (the name on a sign outside the firm if they had one), whether the business had a sign, the approximate area in square meters of the premises, and the approximate number of employees in the business. A photo was also taken of the firm to aid in subsequent identification.

Through this process, more than 10,000 firms were listed during January and February of 2011 (appendix S3 provides a timeline). The firms were then matched by Gauss against two databases of formal firms – a database from PBH of 140,628 firms with municipal registration and a database from JUCEMG (the Chamber of Commerce) of 117,350 firms with state registration. This approach was used to eliminate the “definitely formal” firms (i.e., firms who appeared on both of these lists), resulting in a sample of 7,852 listed potentially informal firms in 574 census blocks. In terms of the listed sector, 48 percent were in commerce, 45 percent were in services, only 1 percent were in manufacturing, and 6 percent were undefined. The large number of firms to be matched in a short timeframe, the requirement for firms to be on both formal lists, and possibly the inexperience of the survey firm in performing this matching exercise meant that, as we will see, a number of formally registered firms remained in this listed sample.

Because part of our design involves examining spillovers within blocks, we wanted to minimize the risk of spillovers across treatment blocks. To do this, we used the address of each listed firm and obtained from this the GPS latitude and longitude. We then calculated the number of firms in other census blocks that lay within 100 meters of the listed firm in a straight line. We examined in detail the 239 census blocks in which at least 10 percent of the firms were close to at least one firm in another census block. We were most concerned with adjacent census blocks, where firms on one side of the street that was a block boundary were in one block and those on the other side of the street were in a different block. We then used an algorithm to reclassify these into new blocks, resulting in a total of 662 geographic blocks. Of these blocks, 57 contained only one potentially

13. For each street that formed part of more than one block (144 streets in total), we calculated the median street number in each block. We then took the difference in median street numbers across blocks for each street and, for all blocks where this difference was smaller than 250 street numbers, combined the firms that were on the same street but in different blocks into a new block.
informal firm each, so we dropped these blocks for a final sample of 605 geographic blocks containing 7,795 firms.

**Randomization into Treatments at the Geographic Block Level**

We randomized these 605 geographic blocks into three groups: control blocks, communication blocks, and inspector blocks. Randomization was stratified by the 9 sub-districts in Belo Horizonte and by whether the block had firm density above or below the median (measured by the number of listed informal firms divided by the area of the block).\(^{14}\) This method resulted in 201 blocks being chosen as control blocks, 202 inspector blocks, and 202 communication blocks (figure 2).

**Baseline Survey of Control and Communication Blocks**

All of the firms listed in the communication and control blocks were targeted for a baseline survey that took place between May and August 2011 and was conducted by the same firm (Gauss) that performed the listing.\(^{15}\) Of these 5,419 firms, 1,455 were found to be formal (through the presentation of documents). In 832 cases, the firm had closed, and neighbors said this was a permanent closure. In 871 cases, the owner was unable to be contacted on three visits at different times and days. In 699 cases, the owners said that they were too busy and/or refused outright to be interviewed. There were errors in records for 209 firms (such as being listed twice), and 1,353 firms were interviewed. Thus, the interview rate was 25 percent of all listed firms and 48 percent of non-formal, non-closed firms without listing data errors.

Firms that appeared in the baseline survey were almost evenly split between commerce and services, with less than 5 percent in manufacturing. The most common types of firms were hairdressers/salons (20 percent), bars (14 percent), automobile mechanics (8 percent), clothing (4 percent), and grocery stores (4 percent), with a wide range of other types of firms, such as restaurants, bookstores, photocopying, flower shops, laundromats, and dance and language schools. The average owner was 44 years old and had run the business for eight years; 37 percent of the owners were female, and 42 percent had completed a high school education or higher. The average firm had 1.3 employees (not including the owner) and reported annual revenues of R$52,000 (US$26,000) and monthly profits of R$2,000 (US$1,000).

**Randomization to Treatment Status at the Individual Level**

Given that only one-quarter of the listed firms answered the baseline survey, it was decided to focus the communication treatments on this subgroup because

---

14. Because firm density (area of the block) was not known for the blocks that we reclassified to avoid having neighboring firms in different blocks, we had three strata within each subdistrict: above median density, below median density, and reclassified block.

there was little point in trying to provide information on the process of formalization to firms that were already formal, closed, or for which the owner could not be found. Therefore, we randomly chose half of the firms that had responded to the baseline survey in each communication block to receive the communication treatment and the other half to receive the free-cost treatment. These firms would then be directly comparable to the firms in the control block that had answered the baseline survey. This method produced a sample of 1,348 firms, which were divided into 689 control firms, 331 communication firms and 328 free-cost firms for use in evaluating the effectiveness of the communication and free-cost treatments (indicated by the solid black box in figure 2).16

The first three columns of table 1 show that randomization succeeded in generating comparable firms across the different treatment groups. The first column shows the control group mean, whereas the second and third columns show the

16. Due to data coding issues, seven of the firms that answered the baseline were not assigned to the control, communication or free-cost groups, whereas two firms that were not in the baseline were assigned to the communication treatment. We work with the 1,348 observations that were assigned for treatment or control.
Table 1. Confirming Randomization

<table>
<thead>
<tr>
<th>Listing Variables</th>
<th>Communication vs Control Blocks</th>
<th>Inspector vs Control Blocks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control Mean</td>
<td>Inspector Free Cost Mean</td>
</tr>
<tr>
<td>In commerce</td>
<td>0.46</td>
<td>0.0254</td>
</tr>
<tr>
<td>In services</td>
<td>0.50</td>
<td>-0.0108</td>
</tr>
<tr>
<td>Has a sign outside</td>
<td>0.34</td>
<td>-0.0104</td>
</tr>
<tr>
<td>Num. employees</td>
<td>0.99</td>
<td>0.0511</td>
</tr>
<tr>
<td>Area (square meters)</td>
<td>43.8</td>
<td>7.293</td>
</tr>
</tbody>
</table>

Baseline Variables

| Owner is female | 0.37       | -0.00542                | 0.0307                        |                         |                        |                        |
| Owner’s age     | 44.3       | 0.172                   | -0.175                        |                         |                        |                        |
| Owner has primary or lower education | 0.17 | 0.00315 | 0.00641 | 0.17 | 0.00315 | 0.00641 |
| Owner has completed high school | 0.42 | -0.0319 | -0.00110 | 0.42 | -0.0319 | -0.00110 |
| Owner is married | 0.58       | -0.00277                | -0.0291                       |                         |                        |                        |
| Age of business (Years) | 7.82       | 0.309                   | 0.150                         |                         |                        |                        |
| Hours owner works per week | 54.1 | -1.565 | -1.244 | 54.1 | -1.565 | -1.244 |
| Total number of employees | 1.14 | 0.154 | 0.201 | 1.14 | 0.154 | 0.201 |
| Keeps business records | 0.32 | 0.0574* | 0.0728** | 0.32 | 0.0574* | 0.0728** |
| Annual revenue (Reais) | 48,595     | 3,637                   | 12,299                        |                         |                        |                        |
| Monthly profits (Reais) | 1797       | 159.3                   | 824.3*                        |                         |                        |                        |
| Owned capital stock (Reais) | 46,319     | 8,243                   | 13,486                        |                         |                        |                        |
| Claims to have SIMPLES but no proof | 0.15 | 0.0132 | 0.00256 | 0.15 | 0.0132 | 0.00256 |
| Claims to be registered with JUCEMG | 0.13 | -0.00379 | -0.00268 | 0.13 | -0.00379 | -0.00268 |
| Claims to have a CNPJ | 0.23 | -0.0164 | -0.0108 | 0.23 | -0.0164 | -0.0108 |
| Claims to have an ALF | 0.24 | 0.0419 | 0.0515 | 0.24 | 0.0419 | 0.0515 |
| Visited by municipal inspector in past year | 0.33 | -0.0539* | -0.0272 | 0.33 | -0.0539* | -0.0272 |
| Visited by state inspector in past year | 0.06 | -0.00700 | -0.0287* | 0.06 | -0.00700 | -0.0287* |
| Visited by federal tax inspector in past year | 0.03 | 0.00692 | 0.00380 | 0.03 | 0.00692 | 0.00380 |
| Sample Size | 689        | 328                     | 331                           | 1398        | 577                | 593                          |

Notes: *, **, and *** indicate significantly different from control mean at the 10, 5 and 1 percent levels, respectively, after controlling for randomization strata and clustering at the block level. Inspector vs control blocks comparison uses sampling weights to account for uneven sampling probabilities.

Source: Authors’ analysis based on data described in paper.
coefficients on the communication and free-cost assignment to treatment dummies in the following regression for firm $i$ in geographic block $s$:

$$BaselineVariable_{i,s} = \alpha + \beta Communication_{i,s} + \gamma FreeCost_{i,s} + \sum \delta_s d_s + \epsilon_{i,s},$$  \hspace{1cm} (1)

where the $d_s$ are dummies for the 27 sub-district*firm density strata used in the block-level randomization, and the standard errors are clustered at the block level. The first five rows show variables obtained from the listing survey, and the remainder show variables from the baseline survey. Of the 48 coefficients shown, five are significantly different from zero at the 10 percent level, and only one is significantly different from zero at the 5 percent level, which is in line with what would be expected by pure chance.

In contrast, because the inspector block firms were not interviewed at baseline, randomization for this group required randomizing among all firms originally listed (indicated by the dashed line box in figure 2). Our original plan called for assigning half the firms listed in these blocks to receive the inspection treatment and half to be indirect inspector firms. However, this was not feasible with the existing number of inspectors, so we had to randomly choose only one-quarter to be inspector firms. This method resulted in 577 firms being assigned to the inspector treatment. The remaining firms in this block were all indirectly inspected. Budget constraints required us to randomly choose 593 of the indirectly inspected firms as the sample we would target for the follow-up survey. The correct control group for these firms consists of listed firms in the control blocks. Recall that we already have a sample of 689 firms in the control blocks that had answered the baseline survey. We therefore randomly chose an additional sample of 709 control block firms that had not answered the baseline survey and assigned them for follow-up. We then reweighted the control group sample to take account of the fact that we had all firms that answered the baseline and only a sample of those that did not in this group.

The last three columns of table 1 compare the inspector and indirectly inspected firms to this weighted sample of control firms in terms of the characteristics available from the listing survey using the following intention-to-treat regression:

$$ListingVariable_{i,s} = \alpha + \beta Inspector_{i,s} + \gamma IndirectInspector_{i,s} + \sum \delta_s d_s + \epsilon_{i,s},$$  \hspace{1cm} (2)

Although this comparison offers far fewer variables for checking balance than is possible for the communication and free-cost treatments, the results are still reassuring, with only one of the 10 coefficients being significant at the 10 percent level.
Follow-up Surveys

A very short phone survey of firms selected for the communication and free-cost treatments was conducted between April 10 and 18, 2012, by Sensus. The purpose of this survey was to follow up with the firms that had received the communication or free-cost content to assess whether they had started the process of formalization, their intent to formalize, and whether the information in the brochure had been useful. It was also intended to serve as a last prompt to use the information and free-cost offer. This survey was not given to firms to which Sensus had not made the treatment offers; thus, it was given to 464 firms, of which 367 responded (79 percent). In this survey, 86 percent of the firms said they remembered receiving the information, and 48 percent had read the brochure after having had it explained in person. Of those who read the brochure, 48 percent said they had learned that there were more potential benefits of formalizing than they had realized, but 42 percent also said they had learned that formalization involved more costs than they had known. Fifty-seven percent said that they had learned the process of formalizing was simpler than they had thought, whereas 18 percent said that they learned the process was more complicated than they had thought.

The full follow-up survey consisted of an in-person survey administered by Sensus between July and September 2012. The target sample size was 3,227 firms consisting of 328 free-cost firms, 331 communication firms, 577 firms assigned to inspectors, 593 indirectly inspected firms, 689 firms in the control blocks that had completed the baseline, and 709 firms in the control blocks that had not completed the baseline. Three attempts on different days at different times were made to contact these firms. A total of 1802 (56 percent) of the targeted firms were interviewed, with a further 14 percent closed, 20 percent absent and 10 percent refusing to be interviewed (see the last row of figure 2).

This attrition rate is relatively high. Appendix S5 examines whether attrition differed systematically between the treatment and control groups and does not find any evidence for such a difference. Nevertheless, in this appendix, we also calculate bounds to show the robustness of our main survey outcome variables to attrition. Importantly, this attrition is only an issue for survey-based outcomes. For our most important outcomes regarding formalization, we are able to use administrative data that are free of attrition.

Administrative Data

We received a list from JUCEMG of all firms that registered with the chamber of commerce in Belo Horizonte between October 25, 2011 (the day the communication treatment started) and September 19, 2012. This list contained the official business name, fantasy name, street address, phone number, and CNPJ of the business as well as registrations for MEI status and SIMPLES. In addition, we obtained a list from PBH of all firms that registered for an ALF license during this same period. We used a matching algorithm and manual checking to match this
to the listing survey data described in Appendix S6. For each formalization measure, we define “definite” matches as cases where there are sufficient data in both data sources to make it almost certain we are matching the same firm and “definite or probable matches” to include cases where it seems very likely to be the same firm but less data are available to confirm the match. We also construct an overall measure of whether a firm has formalized that measures whether the firm is a definite or probable match for at least one of these three forms of formalization (MEI, SIMPLES, or ALF).

Estimating Treatment Effects

To estimate treatment effects, we estimate versions of equations (1) and (2) for different outcomes. Because baseline data are not available for the inspector block firms, we do not control for baseline levels of the outcome variable, although the results for the communication versus control blocks are robust to doing so. A key point that is clear from figure 2 is that we are estimating the communication and free-cost treatment effects for a subpopulation (those who answered the baseline survey) of the group for which we estimate the inspector treatment. We can write the following:

\[
\text{ITT (full sample)} = \Pr(\text{Answer Baseline}) \cdot \text{ITT (Baseline Sample)} \\
+ (1 - \Pr(\text{Answer Baseline})) \cdot \text{ITT (Not in Baseline Sample)},
\]

where ITT is the intention to treat effect. Because the group that did not answer the baseline survey consists of firms that were either closed, already formal, or could not be found, it seems reasonable to assume that any intervention in this group would have smaller impacts than it would on the group that agreed to answer the baseline survey. Therefore, we view our ITT estimates for the communication and free-cost treatments as an upper bound of what they would be in the full sample listed, whereas a lower bound can be obtained by multiplying these estimates by 0.25, the probability of being in the baseline. This approach allows direct comparison of the inspector treatment effects to the communication and free-cost treatment effects.

A further issue to address is that we are examining a number of different outcomes. The tests of significance provided for these outcomes are appropriate if we are interested in a particular outcome, such as whether the communication treatment increases the likelihood of a firm having obtained a municipal license (ALF). However, when looking at the range of outcomes, we need to make adjustments for multiple hypothesis testing. Two approaches are commonly used in the literature. The first is to aggregate outcomes into indices and test whether the overall impact of the treatment on a family of outcomes is different from zero (e.g., Kling and Liebman 2004). We use this approach when considering formalization as an outcome because it is natural to aggregate different types of permits
into a single measure of whether a firm has obtained any permit. This approach is less useful when we are interested in the individual outcomes themselves, so the alternative is to adjust the p-values used to test each individual null hypothesis. We use the Benjamini and Hochberg (1995) correction, which controls the false discovery rate within families of outcomes. Fink et al. (2012) provide more detail on this method and the need for its use in the analysis of development experiments.

**RESULTS**

We begin by discussing the implementation of the different interventions and use our survey data to examine whether this implementation resulted in changes in knowledge or in inspection frequencies. We then examine whether these interventions changed the rate of business closure and, ultimately, the impact on different aspects of formalization.

**Implementation of Interventions**

The communication treatment began on October 26, 2011, with up to four attempts made to deliver the brochure to the owner. This method resulted in 208 of the 331 firm owners assigned to this treatment receiving the brochure (63 percent). A further 100 firm owners declared themselves to be formal and were not given the brochure. Only two owners outright refused the brochure, and the remainder were either unable to be found or absent.

The free-cost treatment took place in February 2012 once arrangements had been finalized with both the government agencies that would waive the costs and with the accountants’ association. We realized that many of the firms that claimed to be formal in the communication treatment may have possessed one document (a federal, state, or municipal registration) but may not have SIMPLES (i.e., all types of required registrations). Therefore, the instructions were to give the offer to all firms without SIMPLES and to explain that it was also valid to take the firms from partially formal to fully formal. As a result, this offer was delivered to 255 of the 328 firms assigned to this treatment (78 percent). Take-up of the offer was incredibly low: one month after the offer, our partner government agency and hotline had received only five calls and two visits; three months after the offer, only 10 to 15 people had called, and one had started the formalization process. Ultimately, only one firm in this treatment group took the offer to formalize and use one of the free accountants.

The inspector treatment began in December 2011 and lasted through April 2012. Of the 577 firms identified for inspection, the inspectors said they were able to locate 530, of which 387 firms were open. Among these 387 firms, 170 were found to have a municipal license, although some of these had expired and others were using more space than licensed or had other infractions. The inspectors notified 269 firms that they were operating without the proper licenses. The
inspectors reported that in their follow-up visits, 143 firms were closed and 88 firms were in the process of formalizing.\footnote{Note that most of these “closed” firms were found to have subsequently re-opened by the time of our follow-up survey, making it unclear what “closed” means in practice.} Their final report showed that 17 of the notified firms had produced a valid municipal license, four had produced a license as a microentrepreneur, and the rest were still classified as in process.

**Impacts on Knowledge and Inspection Likelihood**

The first part of table 2 examines whether the treatments changed individuals’ knowledge about the process of formalization. We obtain these treatment impacts through estimating equations (1) and (2) with different knowledge measures as the outcome of interest. We see first that approximately 60 percent of the control group firms claim to know the steps required to fully register a business and that none of the treatments has a statistically significant impact on this outcome. This high rate of self-assessed knowledge contrasts dramatically with objective measures of knowledge; firm owners were asked the cost of registering and the tax rate faced by firms that register. As in the baseline survey, knowledge of this information is very low, with only 3 percent of the control group giving an answer in the right range for the cost and only 4 percent in the right range for knowing the tax rate. The communication and free-cost treatments have no significant impact on this outcome, whereas the marginally significant impacts of the inspector treatment on knowledge of the tax rate and of the indirect inspector treatment on knowledge of the cost of registration are not significant once a correction is made for multiple hypothesis testing.

We find stronger impacts on whether individuals claim to use an accountant (a requirement of being formal for most firms) and whether they know the cost of an accountant. The free-cost treatment raises the likelihood of claiming to use an accountant by 11.6 percentage points ($p = 0.014$) and the likelihood of knowing the cost of an accountant by 11.8 percentage points ($p = 0.0002$). The communication treatment has a similar magnitude impact on claiming to use an accountant ($p = 0.022$). However, applying the Benjamini-Hochberg (1995) false discovery correction procedure to the 10 information and knowledge estimates results in only a significant impact of the free-cost treatment on knowing how much an accountant costs. Firms assigned to the inspector treatment are also significantly more likely to know the cost of the accountant, even after controlling for false discoveries. Taken together, these results suggest little impact of the treatments on precise details of the formalization process, such as costs and tax rates. However, firm owners did gain information about the role and cost of accountants in this process.

The second part of table 2 examines a second family of outcomes relating to inspector visits. Approximately 47 percent of firms in the control blocks report having been visited by a municipal inspector in the past year. This result increases by 13.5 percentage points for the firms assigned to the inspector
Table 2. Impacts on Knowledge about Formalization and Inspections

<table>
<thead>
<tr>
<th></th>
<th>Communication vs Control Blocks</th>
<th>Inspector vs Control Blocks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample Size</td>
<td>Control Mean</td>
</tr>
<tr>
<td>Claims to know procedures for formalizing</td>
<td>817</td>
<td>0.59</td>
</tr>
<tr>
<td>Knows cost of registration is between R$200 and R$400</td>
<td>817</td>
<td>0.03</td>
</tr>
<tr>
<td>Knows tax rate if registered is between 4% and 8%</td>
<td>817</td>
<td>0.04</td>
</tr>
<tr>
<td>Claims to use an accountant</td>
<td>817</td>
<td>0.27</td>
</tr>
<tr>
<td>Knows cost of an accountant is R$200-R$400/month</td>
<td>817</td>
<td>0.10</td>
</tr>
<tr>
<td>Inspections</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Municipal inspector visit in past year</td>
<td>817</td>
<td>0.42</td>
</tr>
<tr>
<td>Received information on how to formalize from any inspector</td>
<td>817</td>
<td>0.14</td>
</tr>
<tr>
<td>Reports a neighboring firm inspected in past year</td>
<td>817</td>
<td>0.19</td>
</tr>
<tr>
<td>Was notified or fined by an inspector in past year</td>
<td>817</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Notes: Standard errors in parentheses, clustered at the block level. *, **, and *** indicate significantly different from control mean at the 10, 5, and 1 percent levels, respectively, after controlling for randomization strata. Sampling weights are used for the Inspector vs Control blocks comparisons. Coefficients in bold remain significant applying the Benjamini-Hochberg (1995) procedure within a family of outcomes to control false discoveries.

Source: Authors’ analysis based on data described in text.
treatment \((p = 0.0002)\), which is consistent with the inspector treatment increasing the likelihood of receiving an inspection. Note, however, that the combination of firms that were closed or unable to be located by the inspectors coupled with the fact that some firms would have received an inspection anyway means that this difference between the treatment and control groups is much less than 100 percentage points. However, note also that we do not know how many of these reported visits were only to check on the signage of the firm versus how many also checked the municipal license. The inspector treatment firms are marginally more likely to say that they obtained information on how to formalize from an inspector (not significant after adjustment for multiple testing) and are no more likely to be notified or fined than control firms.

The indirectly inspected firms are no more likely to report having seen or heard that a neighboring firm had received an inspection in the past 12 months. This may result from the fact that inspections occur to some extent anyway, and that firm owners do not always communicate amongst each other – 35 percent of firm owners say they do not talk at all to other firm owners about business matters. Additional support for the view that many firms do not notice inspections in neighboring firms comes from the fact that twice as many firms report having been inspected themselves than those that report having seen a neighboring firm inspected. These findings imply that we should expect the spillover impact from inspecting the inspector group firms on the indirectly inspected firms to be minimal in terms of formalization.

**Impacts on Firm Survival**

La Porta and Shleifer (2008) noted that a competing view to the De Soto/Doing Business view of the informal economy as home to potentially productive entrepreneurs held back by regulatory barriers is the dual economy view associated with Tokman (1992) and Rauch (1991). In this view, the informal sector is a source of subsistence livelihoods for individuals with relatively low levels of human capital, and any increase in firm value that these owners would be able to generate by formalizing would not be large enough to offset the additional costs of taxes and other regulatory requirements. The result is that enforcing formalization may cause these firms to shut down because they cannot afford to operate formally. In both cases, the entrepreneur is making a cost/benefit calculation of whether to formalize, but the competing views differ in terms of what they see as the main costs and benefits entering this calculation.

Table 3 examines this possibility by looking at the impact of the different treatments on firm closure. Firm closure is measured by whether the firm is observed to be closed at the time of the follow-up survey coupled with information obtained from neighboring businesses. Between 14 and 16 percent of the control group was verified as closed at follow-up, with none of the treatments having any sizeable or significant impact on this closure rate. In particular, it is not the case that firms that received enforcement through the inspector treatment are more likely to have closed.
Table 3. Business Closure Impacts by Treatment Group

Panel A: Control vs Communication Blocks

<table>
<thead>
<tr>
<th></th>
<th>Control Mean</th>
<th>Free Cost Difference</th>
<th>Communication Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closed at Follow-up</td>
<td>0.142</td>
<td>0.0139</td>
<td>0.0297</td>
</tr>
<tr>
<td></td>
<td>(0.0268)</td>
<td>(0.0277)</td>
<td></td>
</tr>
<tr>
<td>Sample Size</td>
<td>685</td>
<td>329</td>
<td>328</td>
</tr>
</tbody>
</table>

Panel B: Control vs Inspector Blocks

<table>
<thead>
<tr>
<th></th>
<th>Control Mean</th>
<th>Inspector Assigned Difference</th>
<th>Indirectly Inspected Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closed at Follow-up</td>
<td>0.166</td>
<td>−0.0184</td>
<td>−0.00850</td>
</tr>
<tr>
<td></td>
<td>(0.0188)</td>
<td>(0.0209)</td>
<td></td>
</tr>
<tr>
<td>Sample Size</td>
<td>1383</td>
<td>577</td>
<td>593</td>
</tr>
</tbody>
</table>

Notes: Standard errors in parentheses, clustered at the block level. *, **, and *** indicate significantly different from control mean at the 10, 5 and 1 percent levels, respectively, after controlling for randomization strata. Sampling weights are used in Panel B to account for uneven sampling probabilities.

Source: Authors’ analysis based on data described in text.

Impact on Formalization

Table 4 turns to the main outcome of interest, whether the treatments succeeded in getting firms to formalize. We use the administrative data to measure formalization because these data offer impacts without attrition and substantially larger samples to measure impacts, providing the most power. The first few columns compare the control firms to the free-cost and communication firms. We see that assigning firms to the free-cost treatment has a strongly significant negative impact on the likelihood of MEI registration ($p = 0.008$ for definite matches, $p = 0.010$ for probable matches). One possible explanation for this is that the free-cost and communication interventions explained how to register for SIMPLES but not how to register as a MEI. The free-cost treatment also emphasized the need to have an accountant if the firm formalized, which is not required for MEIs. The unexpected policy change immediately after we launched this intervention increased the revenue thresholds under which firms could register as MEIs. It is plausible that firms that received the communication and free-cost intervention were less aware of this policy change given the information about the
## Table 4. Impacts on Formality

| Administrative data measures of formalizing after interventions began | Communication vs Control Blocks | Inspector vs Control Blocks |
|---|---|---|---|---|---|---|
| | Sample Size | Control Mean | Free Cost Difference | Communication Difference | Sample Size | Control Mean | Inspector Assigned Difference | Indirectly Inspected Difference |
| Definite match for SIMPLES | 1346 | 0.007 | 0.00618 | -0.00300 | 5186 | 0.006 | 0.00390 | -0.00144 |
| | | (0.00662) | (0.00446) | | | (0.00443) | (0.00229) | |
| Definite or probable match for SIMPLES | 1346 | 0.015 | 0.00177 | -0.00706 | 5186 | 0.014 | 0.00421 | 0.00102 |
| | | (0.00773) | (0.00761) | | | (0.00826) | (0.00469) | |
| Definite match for MEI | 1346 | 0.060 | -0.0349*** | -0.0155 | 5186 | 0.026 | 0.00313 | -0.00557 |
| | | (0.0131) | (0.0139) | | | (0.00826) | (0.00469) | |
| Definite or probable match for MEI | 1346 | 0.067 | -0.0370*** | -0.0143 | 5186 | 0.033 | 0.0138 | -0.00339 |
| | | (0.0142) | (0.0161) | | | (0.0106) | (0.00561) | |
| Definite match for ALF | 1346 | 0.030 | 0.00311 | 0.000227 | 5186 | 0.032 | 0.0218** | -0.00540 |
| | | (0.0113) | (0.0117) | | | (0.0110) | (0.00535) | |
| Definite or probable match for ALF | 1346 | 0.041 | -0.000335 | -0.00597 | 5186 | 0.041 | 0.0327*** | 0.00206 |
| | | (0.0125) | (0.0125) | | | (0.0124) | (0.00653) | |
| Definitely obtained any type of formal status | 1346 | 0.083 | -0.0281* | -0.0120 | 5186 | 0.056 | 0.0245* | -0.0100 |
| | | (0.0159) | (0.0174) | | | (0.130) | (0.0068) | |
| Definitely or most likely obtained any type of formal status | 1346 | 0.104 | -0.0350* | -0.0182 | 5186 | 0.075 | 0.0392*** | -0.0034 |
| | | (0.0183) | (0.0209) | | | (0.0149) | (0.0083) | |

**Notes:** Standard errors in parentheses, clustered at the block level. *, **, and *** indicate significantly different from control mean at the 10, 5, and 1 percent levels, respectively, after controlling for randomization strata. Sampling weights are used for the Inspector vs Control blocks comparisons.

**Source:** Authors’ analysis based on data described in text.
need to register for SIMPLES and to obtain an accountant that they were given in person, causing them to decide not to register. The point estimates are also negative for the communication treatment. Testing for equality of treatment effects, we cannot reject that the communication-only treatment has the same negative impact as the free-cost treatment – but neither can we reject a zero treatment effect for communication only.

We see no significant impacts of the free-cost or communication treatments on registering for SIMPLES or obtaining an ALF license. Pooling together all three measures, we see that 8–10 percent of the control group that was interviewed at baseline obtained some form of formal status after our interventions began and that this was approximately 3 percentage points lower for the free-cost group. Given the size of the free-cost group, this equates to approximately 10 firms not formalizing that would have in the absence of this intervention.

Turning to the last three columns, which consider the inspector versus control block comparisons, we see that assignment to the inspector treatment group leads to a strongly significant 2 to 3 percentage point increase in the likelihood of obtaining an ALF. Recall that the ALF license is the only one the municipal inspectors are legally able to enforce. Under the one-stop shop for registration, we expected that firms would register and obtain SIMPLES and an ALF all at once. However, if firms had already obtained a CNPJ or if they had previously had an ALF that had expired (they are valid for five years), firms could just register and obtain an ALF only. It therefore appears that this extra formalization was by firms that were already partially formal. There are small and insignificant impacts of inspections on the other forms of formalization; thus, the overall impact of formalizing comes from the ALF registrations. We find a significant overall impact of between 2 and 4 percentage points, which is equivalent to between 11 and 22 extra firms formalizing of the 577 assigned to the inspector group.

In contrast, we find a rather precise zero effect of the indirect inspector treatment on the likelihood of formalizing. The 95 percent confidence for the treatment effect on “definitely or most likely obtained any type of formal status” is [−0.020, +0.013]. This result is consistent with the evidence in table 2 that the firms in this treatment group did not notice any increase in inspections of neighboring firms.

Estimates of the Impact of Being Inspected

Although we find a significant impact of being assigned to receive an inspector visit on obtaining an ALF license, the effect of 3 percentage points is very small. There are several reasons for this small effect: many of the firms were closed or could not be found by the inspectors, some firms were already formal, and some

19. Recall that because the formalization measures naturally aggregate, we consider impacts on this aggregate to deal with concerns with multiple hypotheses testing.
firms would have been inspected anyway. To estimate the causal impact of being inspected on formality, we therefore run the following instrumental variables regression:

$$Formalize_{i,s} = \alpha + \beta ReceiveMunicipalInspection_{i,s} + \sum \delta_s d_s + \epsilon_{i,s},$$

(5)

where we instrument the follow-up survey report of whether the firm had received a municipal inspection in the past year with assignment to the inspector treatment group. We estimate this equation using the follow-up survey data for the control group and inspector group only. We consider both ALF registration, which is the registration form most closely tied to municipal inspection, and our overall measure of formalizing.

Table 5 displays the results. We see that the point estimates range from 0.214 to 0.265, so receiving an inspection results in a 21 to 27 percentage point increase in the likelihood of formalizing. The statistical significance is greatest for ALF registration, where the p-value is 0.108 for definite registration and 0.051 for definite or probable registration. This is the impact on the group of firms that answered the follow-up survey. It therefore removes firms that had closed or that could not be found easily but still includes some firms that were already formally registered.

The estimated cost of an inspection is R$64.34, which is based on an estimated inspection taking 56 minutes per visit plus 17 minutes of travel time (estimates provided by PBH). The inspectors visited 387 firms (the rest were closed or not found), so the total cost of inspections is estimated at R$24,900. Taking our estimated impact of 11 to 22 more firms formalizing, the cost per firm formalized is R$1132–2264. Under the more questionable assumption that our inspections did not differ from the inspections these firms would have received anyway, the IV estimates suggest that approximately four additional inspections are required to get one firm to register for an ALF license, so the estimated cost of formalizing one firm is approximately R$256.

Annual tax revenue is R$620 for a MEI; based on 4 percent revenue tax on the average revenue of R$57,000 for newly formalized firms with an ALF, the annual tax revenue is R$2280. Firms report that firms like theirs typically only report only half their revenues, which would take the SIMPLES annual tax down to R$1140. Therefore, it appears that the cost of formalizing a firm in our experiment via inspections would be gained back within the first year of tax payments (or more than gained back if we consider only the marginal visits under the IV estimation), with subsequent years of tax payments then a net gain for the government.

20. Note that this assumes that being assigned to the inspection treatment does not affect formalization for firms that would have been inspected anyway. This assumption may not hold if the inspection they would otherwise receive only checked their signage, whereas our inspection also checked for the municipal license. As a result, we consider these IV results suggestive only.
However, because our estimates above suggest that the main effects are for firms that were already partly formal, it is unclear whether all of these tax gains would be realized in practice. The municipality would gain the fixed renewal fees, but it is less clear whether these firms would now pay SIMPLES taxes. Even if they did, the municipality (which pays for the municipal inspectors) only receives a share of this additional revenue, with the remainder going to the state and federal governments. Apart from the annual inspection tax, under SIMPLES, municipalities only directly obtain tax revenue from service firms. One component of the SIMPLES tax, called the ISS, is 2 percent of revenues on the first R$180,000 of revenues and 2.79 percent after that. Therefore, the municipality could gain approximately R$570 per year from formalizing a typical service firm in our sample. This result suggests that the municipality could benefit from well-targeted attempts to formalize service firms, but our results also suggest that this can be difficult in practice.

Comparing Actual Impacts with Expectations of Treatment Impacts

A standard question regarding impact evaluations is whether they deliver new knowledge or merely formally confirm the beliefs that policymakers already have (Groh et al. 2012; Hirshleifer et al. forthcoming). To measure whether the results differ from what was anticipated, in January 2012 (before any results were known), we elicited the expectations of the Descomplicar team regarding what they thought the impacts of the different treatments would be. Their team expected that 4 percent of the control group would register for SIMPLES between

Table 5. Instrumental Variable Estimates of the Impact of an ALF Inspection on Formalization

<table>
<thead>
<tr>
<th></th>
<th>Definitely Got ALF</th>
<th>Definitely or Most likely got ALF</th>
<th>Definitely Formalyzed</th>
<th>Definitely or Most likely Formalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reports receiving an ALF inspection</td>
<td>0.204 (0.127)</td>
<td>0.265* (0.136)</td>
<td>0.214 (0.148)</td>
<td>0.222 (0.157)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,100</td>
<td>1,100</td>
<td>1,100</td>
<td>1,100</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors in parentheses, clustered at block level. *, **, and *** indicate significance at the 10, 5, and 1 percent levels, respectively. Regressions also control for randomization strata and are only for the control group and firms assigned to receive inspectors. Assignment to receive an inspector is used as an instrument for receiving an inspection. Table 2 provides this first stage.

Source: Authors’ analysis based on data described in text.

21 Municipalities also receive part of the taxes that go to federal and state governments indirectly through transfers from these levels of government back to the municipality. However, we ignore this indirect component, which is based on a complicated revenue-sharing procedure that does not depend only on municipal tax takes.
the baseline and follow-up surveys. We see from table 4 that this is an overesti-
mate of the SIMPLES registration rate, but given the change in MEI require-
ments, it is in line with the combined SIMPLES and MEI registration level.

The communication-only group was expected to double this rate so that
8 percent would register, the free-cost treatment would lead to 15 percent regis-
tering, and the inspector treatment would lead to 25 percent registering. The
team did not expect there to be any indirect inspector effect and so expected that
only 4 percent of the untreated firms in the inspector blocks would register. The
zero or negative impacts of the communication and free-cost treatments are
therefore surprising. The overall impact of the inspector treatment is much lower
than expected but is in line with the IV estimates, suggesting that the
Descomplicar team has a reasonable sense of what to expect when an inspection
actually occurs but may have overestimated the amount of new inspections that
would take place. Their expectation of a lack of impact for the indirect inspector
treatment was also accurate.

**Impacts on Trust and Attitudes towards Government**

De Mel et al. (2012) offer Sri Lankan firms monetary payments to get them to
formalize and find that one outcome of formalization is that firms have more
trust in local government. In their case, formalization is much cheaper and
quicker than firms had believed. They note that one possible reason for this in-
crease in trust was that firms experienced better services from the government
than they had expected, whereas an alternative could be that they were less
afraid of being shut down after registering.

Our follow-up survey asked firms about their trust and views of government.
Individuals were asked on scale of 1 to 10 how much they trusted different
actors, where 10 denoted the most trust and 1 denoted the least trust. They were
also asked whether they believed that government acts in the interests of the
people or in its own interest. Table 6 reports the results of estimating the impacts
of our treatment assignments on these outcomes. We see a strong contrast to the
results in Sri Lanka: the attempts at formalization in Belo Horizonte appear to
have generally worsened trust in government. The results for the communication
and free-cost treatments all remain significant after controlling the false discovery
rate at $\alpha = 0.10$, whereas the impact of the free-cost treatment on believing the
government acts in its own interests also remains significant when controlling the
false discovery rate at $\alpha = 0.05$.

The impacts are not that large: a reduction of 0.3 to 0.5 points on a 10-point
trust scale, which represents a 0.1 to 0.15 change, and an increase of 4 to 10 per-
centage points in the likelihood that firm owners think the government acts in its
own interest rather than in the interests of the people. Nevertheless, in an envi-
ronment of widespread informality, efforts to reach out to particular firms and
bring them into the formal system using either carrots or sticks may run the risk
of increasing distrust in government if firm owners do not see any benefits from
being brought into this formal system. The distrust effect is more significant for
Table 6. Impacts on Trust and Views of Government

<table>
<thead>
<tr>
<th></th>
<th>Communication vs Control Blocks</th>
<th>Inspector vs Control Blocks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample Size</td>
<td>Control Mean</td>
</tr>
<tr>
<td>Trust state governor</td>
<td>805</td>
<td>4.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trust state officials</td>
<td>799</td>
<td>4.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trust state and municipal inspectors</td>
<td>793</td>
<td>4.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Believe people in govt. act in own interests rather than interests of the people</td>
<td>733</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Standard errors in parentheses, clustered at the block level. *, **, and *** indicate significantly different from control mean at the 10, 5, and 1 percent levels, respectively, after controlling for randomization strata. Sampling weights are used for the Inspector vs Control blocks comparisons. Coefficients in bold remain significant applying the Benjamini-Hochberg (1995) procedure within a family of outcomes to control false discoveries.

Source: Authors’ analysis based on data described in text.
information and free-cost efforts than for the inspection treatment, possibly because a government initiative that invited individual firms to register is different from usual activities and appears to have aroused suspicion.

**Conclusion**

Despite reforms that make it faster and simpler for informal firms in Brazil to register, the majority of firms remain informal. Although simply paying firms to formalize has been found to have large impacts on formalization rates in Sri Lanka, this approach is unlikely to be on the policy menu for most governments. Instead, governments can use a range of carrots and sticks to attempt to bring firms into the formal sector. Our experiment tests some of the most common ones: informing firms, making it cheaper for them to register, and increasing the enforcement of rules. Our findings suggest that sticks rather than carrots seem more effective at getting firms to formalize, but we also find limits to this approach.

The process of registering in Belo Horizonte still requires more steps and complications than in a number of other countries that have pursued entry reforms. In addition to facing taxes, firms that do register face a relatively large cost in terms of the need to hire an accountant. Faced with these costs of being formal, it appears that few informal firms want to formalize unless they are forced to do so by enforcement. We are unable to measure whether firms benefit from being forced to formalize because the number of firms induced to formalize is too small, and our follow-up survey suffered from high item non-response on sales and profits questions. However, evidence from other countries (McKenzie and Sakho 2010; de Mel et al. 2012) suggests that although some informal firms benefit from formalizing, the majority appear not to. Being informal is thus likely to be privately optimal for many firms.

This finding suggests three directions for government policy. The first is to reconsider where it is desirable to even attempt to bring these firms into the formal sector. It may not make much sense for the smallest firms, but given the limited tax base and the fact that firm owners with revenues in the range that qualifies for SIMPLES are likely to be at least in the middle of the income distribution, there may be a public benefit to formalizing these firms even if there is no private benefit. Formalization may also bring other wider benefits, such as reducing a “culture of informality” and allowing more efficient reallocation by protecting formal firms from “unfair competition” from less efficient informal firms. The second avenue for policy is to further simplify the ease of formalizing and, more importantly, to revisit the need for an accountant, which dramatically increases the cost of being formal. Efforts to link formality with access to government programs and bank financing might help to induce some firms to register, but many firms will not benefit from such approaches. Improved enforcement is thus the third part of policy efforts. Our research shows that enforcement can induce formalization, but there are limits. Rather than having separate inspectors for
different forms of registration, having municipal inspectors who are able to enforce municipal, state, and federal registration should have stronger impacts. Furthermore, given that many of the firms the inspectors said they had closed were open again at the time of our follow-up survey, there appears to be scope for improving the degree of enforcement that inspection actually entails. Combining enforcement with carrots may offer the greatest impact because firms may be far more receptive to information and lower costs of registering when they have an enforcement incentive to register.

 References


Economic Shocks and Subjective Well-Being: Evidence from a Quasi-Experiment

Jacob Gerner Hariri, Christian Bjørnskov, and Mogens K. Justesen

This article examines how economic shocks affect individual well-being in developing countries. Using the case of a sudden and unanticipated currency devaluation in Botswana as a quasi-experiment, we examine how this monetary shock affects individuals’ evaluations of well-being. We do so by using microlevel survey data, which—incidentally—were collected in the days surrounding the devaluation. The chance occurrence of the devaluation during the time of the survey enables us to use pretreatment respondents, surveyed before the devaluation, as approximate counterfactuals for post-treatment respondents, surveyed after the devaluation. Our estimates show that the devaluation had a large and significantly negative effect on individuals’ evaluations of subjective well-being. These results suggest that macroeconomic shocks, such as unanticipated currency devaluations, may have significant short-term costs in the form of reductions in people’s sense of well-being. JEL codes: E50, H0, I31, O23

Few tasks are more important in the social sciences than discovering the sources of human well-being. While this remains a contested issue (Frey and Stutzer 2000; Clark et al. 2008; Frey 2008; Bjørnskov et al. 2010; Deaton 2012), the question of whether “money buys happiness” attracts particular attention, no doubt because of the seemingly paradoxical finding—first reported by Easterlin (1974, 1995)—that income growth is not associated with corresponding increases in happiness and well-being (Clark et al. 2008; Easterlin et al. 2010). However, recent work has emphasized that subjective well-being does seem to fluctuate with banking and financial crises (Deaton 2012; Bjørnskov 2014; Montagnoli and Moro 2014) and macroeconomic factors like inflation, unemployment, and gross domestic product (GDP; Oswald 1997; Di Tella et al. 2001,
2003; Stevenson and Wolfers 2008; Kahneman and Deaton 2010; Sacks et al. 2012a), providing some support for the claim that income is correlated with happiness and well-being.

In this article, we contribute to this literature by examining how macroeconomic shocks affect individual well-being. Using the case of an unanticipated and rapidly implemented currency devaluation in Botswana—a middle-income country in sub-Saharan Africa—we examine how individual evaluations of well-being respond to such a monetary policy shock. We do so by analyzing microlevel data from the Afrobarometer, which happened to be in the field conducting interviews for a survey at the time when the citizens of Botswana were exposed to the news of the national currency devaluation. Specifically, two days into the survey—late in the day on May 29, 2005—the central bank of Botswana and the Ministry of Finance and Development Planning issued a public statement saying that the national currency would be devalued by 12 percent, with effect from the following morning.¹ Our analysis exploits the fact that the chance occurrence of the devaluation creates a clear demarcation between respondents surveyed before the devaluation and respondents surveyed in the days following the devaluation.

The incidental occurrence of the central bank’s intervention during the time of the survey provides us with a quasi-experimental research design allowing us to examine the effect of a monetary shock on subjective evaluations of well-being. However, the fact that the devaluation was an unanticipated shock—a claim we will validate later—is not sufficient to treat it as exogenous. Identification of the causal effect on subjective well-being requires that the devaluation—the treatment—is orthogonal to the error term, that is, uncorrelated with other factors that may affect the outcome. As we discuss in detail below, this assumption may not be satisfied unconditionally due to geographically imbalanced sampling of respondents in the pre- and post-treatment groups, caused by a shift in the sampling of respondents from urban to rural areas in the days surrounding the devaluation. However, since we can identify and measure the source of nonrandom treatment assignment with relative precision, the exogeneity of the devaluation is plausible conditional on adjusting for the urban-rural shift.

On this assumption, our results show that the devaluation caused an instant and observable discontinuity in the data. The change in reported levels of well-being occurred literally overnight, reflecting that individuals’ responses were immediate and most likely based on expectations about the future consequences of the devaluation. Thus, respondents in the treatment group—surveyed after the devaluation—report levels of well-being that are both substantially and significantly lower than respondents in the control group—surveyed immediately before the devaluation. This result is robust to adjusting the data for nonrandom treatment assignment in various ways and to centering the sample on the discontinuity in the data created by central bank intervention. However, we also report

¹. Press Release, 17:00 hours, Sunday 29 May 2005, issued by the Ministry of Finance and Development Planning.
evidence that respondents with more education and larger consumption of media news react more strongly to the policy shock, suggesting that the effect of monetary shocks may be conditional on individuals’ information and cognitive sophistication.

The article contributes to the broader literature on the determinants of individual happiness and well-being (Oswald 1997; Dolan et al. 2008; Frey 2008; Bjørnskov et al. 2010). It is also closely related to contributions linking macroeconomic variables like GDP and inflation to subjective well-being (Di Tella et al. 2001, 2003; Stevenson and Wolfers 2008; Deaton 2012). In particular, our results support the conclusion of Di Tella et al. (2003: 823) that “macroeconomics matters,” at least with respect to monetary shocks. However, the quasi-experimental nature of our design distinguishes it from standard correlational studies, which mostly regress well-being or life satisfaction on some potentially endogenous micro- or macrolevel explanatory variable. The “shock nature” of the currency devaluation allows us to avoid most of the problems caused by the usual endogeneity of macroeconomic and policy variables like GDP and inflation (Besley and Case 2000; Di Tella 2003). In this respect, our article adds to the small literature using large-scale exogenous shocks to study changes in subjective well-being (e.g., Frankenberg et al. 2003; Frijters et al. 2004; Montagnoli and Moro 2014).

The rest of the paper is structured as follows. Section 2 outlines theoretical mechanisms linking currency devaluations to subjective well-being. The research design and the experimental situation are described in section 3. Section 4 introduces the data, and section 5 provides empirical estimates of the effects of the devaluation. Section 6 concludes.

I. DEVALUATION AND SUBJECTIVE WELL-BEING

The response of individuals to the news of a devaluation might depend on at least two different mechanisms—price responses and a signaling mechanism.

Price Responses and Expectations

First, following a devaluation, the prices of imported consumer goods will increase. If contracts are written in foreign currency—in the case of Botswana most likely South African rand or US dollars—the price increase will be virtually immediate. If contracts are denoted in Botswana Pula, the price correction may occur gradually as import contracts are renegotiated over a period of weeks or months to reflect the new exchange rate. Depending on the price elasticity of the good, the degree of competition in the product market, and the availability of domestic substitutes—all of which would reduce the price response—some (or all) of the price increase will be reflected in proportionately increasing consumer prices. The devaluation thus makes imported goods more expensive and therefore reduces real wages for the population at large. Since Botswana is a net
importer of food and other consumables like fuel and energy from, for example, South Africa (Rakotoarisao et al. 2011), the economic costs of the currency devaluation mainly accrued to consumers, at least in the short term.

Second, the general price level is also likely to increase following a devaluation for two reasons associated with the price of domestically produced goods and services. One is that the devaluation affects final goods through its effect on import prices of raw materials and intermediate goods. By increasing input prices in production, the devaluation affects the prices of final goods that are produced domestically but relies on imported raw materials or intermediates. The second reason is that an import price increase is likely to cause an increase in the demand for domestically produced substitutes (or near-substitutes). As such pairs of goods tend to have substantial cross-price elasticities, the price of substitutes is also likely to increase proportionally to the devaluation.

These price effects are likely to affect individuals and households to approximately the same extent. However, household reactions to the shock can differ substantially, as documented by Frankenberg et al. (2003). In particular, one might expect that households directly engaged in the production of near-substitutes to imports may benefit in the medium-run, as demand patterns react to the changing relative prices. Conversely, all households would be harmed by a general drop in aggregate demand and an increase in uncertainty, making it difficult from a theoretical angle to make any systematically heterogeneous predictions (see Montagnoli and Moro 2014).

Although price increases may occur immediately following the news of a currency devaluation, they do not adjust fully or instantly to their new equilibrium. Subsequent changes in subjective well-being are therefore likely to at least partially reflect expectations about the future (Graham 2008; Guriev and Zhuravskaya 2009; Sacks et al. 2012b). If prices of imported goods increase, price changes will take place almost instantly. Changes in economic expectations can therefore occur very rapidly given that individuals rely on consumption of imported final goods. If the general price level increases, the inflationary effects of the devaluation are likely to spread over time to most goods and services and lead to changes in expected and actual economic well-being for larger segments of society. However, the speed of adjustment of expectations is likely to depend on individuals’ economic and cognitive sophistication. If individuals have little information about the economy, their economic expectations are likely to adapt gradually as the consequences of the devaluation become observable in prices, real wages, and unemployment risk. In contrast, if individuals have sufficiently sophisticated mental models of the economy, a devaluation enables the formation of rational expectations (Muth 1961; Phelps 1967) that change rapidly after the news of the devaluation but presumably before the actual changes in absolute or relative prices. In this case, individuals with more sophisticated mental models of the economy will be better at foreseeing the consequences of devaluation and thus change their expectations earlier and more precisely. The extent to which people form and internalize expectations of how the economy is likely to develop
in the longer run also depends on their cognitive sophistication, as well as information obtained from past experiences with similar policy shocks.

Signaling and Uncertainty

Another type of mechanism may also affect individuals’ well-being. As stressed by Graham (2011), well-being is not only affected by individuals’ current status and expectations of the immediate future but also their perceived uncertainty. A devaluation announcement might therefore have two additional effects.

First, relatively well-informed individuals are probably able to assess the direction of the price effects that we described above but may only have a vague idea about their magnitude. Policy changes with complex consequences, such as devaluations, may thus release a perceived demand for insurance of some kind, which in all forms must reduce current consumption possibilities. Whether this demand can be covered in actual insurance markets is questionable in middle-income countries. A likely consequence of a perceived and unanticipated increase in uncertainty is therefore likely to be an increase in either current savings or changed savings behavior in the near future. In either case, expected consumption and economic welfare is likely to decrease (Graham 2011).

Second, the announcement of a devaluation can easily be taken as a signal that the economy is moving in a direction that is inconsistent with individuals’ prior information. With limited information on the state of the domestic economy and less knowledge and information about that of major trading partners and the general world economy, governments’ policy decisions may work as signals of the direction of economic change. Changes such as devaluations can therefore be perceived as signals of future economic slowdown—particularly by more well-informed citizens—that induce changes to expectations and financial plans.

These nontechnical theoretical considerations lead us to expect the following: First, people’s evaluations of subjective well-being will on average decrease following a devaluation, all else equal. Second, however, since price effects may not materialize immediately and signals from government policy changes may be complex, we also expect that individuals with more sophisticated mental models and more complete information are able to form more accurate predictions of the consequences of a devaluation, and that their self-reported well-being will respond more strongly to the news of a devaluation. Against this background, we proceed by describing the quasi-experimental design.

II. Quasi-Experimental Research Design

Late in the afternoon on May 29, 2005, the Bank of Botswana—the country’s central bank—and Botswana’s Ministry of Finance and Development Planning issued a press release stating that the national currency—the Pula—would be devaluated by 12 percent against a basket of international currencies, with effect from the following morning, May 30, 2005. The central bank’s decision to
devaluate the Pula came as a shock to the general public, the business community, and currency markets in Botswana, as we will show in more detail below. Our research design exploits this sudden and unanticipated intervention to examine the effect of economic shocks on individuals’ subjective well-being. We are able to do so because, incidentally, the devaluation occurred during the period where the Afrobarometer—an independent research project conducting surveys of political and social issues in Africa—was interviewing a representative sample of citizens in Botswana.2 The chance occurrence of the devaluation two days into the survey demarcates the sample of respondents into a treatment group surveyed after the intervention and a control group surveyed immediately before the intervention.3

The terms “natural” and “quasi” experiments are often used in an imprecise and interchangeable sense. However, we advertently refer to the Botswana devaluation as a quasi-experiment and distinguish it from natural experiments. While a common feature of natural and quasi experiments is that an intervention generated by some force outside the control of the researcher assigns subjects into treatment and control groups (Meyer 1995; Robinson et al. 2009), the defining characteristic of natural experiments is that treatment assignment occurs in a random or ‘as-if’ random way (Dunning 2008, 2012). However, as emphasized by Cook and Campbell (1979) and Achen (1986), what distinguishes quasi-experimental designs from natural and controlled experiments is that assignment to treatment is nonrandom, which means that the treatment and control groups are imbalanced—or nonequivalent—at the outset. This means that even a macroeconomic shock, for example, a surprise devaluation, may not be strictly exogenous because nonrandom treatment assignment may make treatment status correlated with other factors that affect the outcome. In a regression framework, nonrandom assignment to treatment may therefore imply that treatment status is not statistically independent of the error term—at least not unconditionally—and that confounding is a potential challenge to a causal interpretation of the estimated treatment effect.

While the survey data we use are a random and representative sample of 1200 adult citizens in Botswana, the key source of nonrandom assignment to treatment and control is that the sampling of respondents before and after the devaluation is geographically imbalanced. Overall, 216 respondents—corresponding to 18 percent of the sample—were surveyed before the devaluation (the control group), while 984 respondents were surveyed after (the treatment group). However, pre-treatment respondents are predominantly from the capital of

---

2. The data are published as part of the third round of the Afrobarometer. Technical details on the sampling of respondents and the methodology of the survey are available on the Afrobarometer website http://afrobarometer.org/. See also Bratton et al. (2005) for descriptions of the Afrobarometer, and Mattes (2007) for a discussion of survey research in developing countries.

3. The survey started on May 28 and ended on June 12, 2005. Since the devaluation was announced late in the afternoon (17:00 hours) on May 29, no interviews started after the announcement of the devaluation (the final interview began at 16:57 hours).
Botswana—Gaborone—and from urban areas more broadly. Specifically, 63 percent of the pre-treatment respondents were from Gaborone; 85 percent were from urban areas. In the two days following the devaluation, only 10 percent of the respondents were from urban areas. Therefore, the treatment coincides with a shift in the sampling of respondents from urban to rural areas, which is also likely to correlate with respondents’ evaluations of their living conditions and well-being. Part of the treatment effect might therefore be due to preexisting differences in subjective well-being between people in rural and urban areas, or may arise if, for example, more confident, optimistic, or resourceful individuals self-select into cities and urban areas (cf. Cook and Campbell 1979; Achen 1986). Despite this initial imbalance between the control group and the treatment group, there are at least two reasons to believe that we can plausibly mitigate the consequences of nonrandom assignment. First, since we can identify the source of nonrandom treatment assignment—geographically imbalanced sampling—with relative precision, we can also go a long way towards making the treatment and control groups comparable by adjusting for the relevant covariates. As we explain in more detail below, we do so in a number of ways; most importantly by controlling for whether respondents live in urban or rural areas; by excluding respondents in the Gaborone area; and by zooming in on the discontinuity in the data generated by the devaluation. Second, since the imbalance between the pre- and post-treatment groups is a result of the fact that the Afrobarometer simply happened to conduct interviews mainly in Gaborone and urban areas prior to the devaluation, we can rule out other sources of nonrandom treatment assignment caused by the actors generating the data. First, it is highly implausible that the Afrobarometer’s timing of the survey was related to the central bank’s decision to devalue in any way, or vice versa. Second—and more importantly—there is little reason to believe that respondents could somehow sort or directly self-select into treatment or control, since they did not have the information, incentive, or capacity to do so (cf. Dunning 2012: 236). Indeed, qualitative evidence suggests that people in Botswana did not have any prior information about the central bank’s decision to devalue. For instance, media reports by the Mmegi (The Reporter)—an independent Botswana newspaper—and the BBC in the days following May 29, 2005, consistently refer to the devaluation as a “surprise” or “shock.” One report notes that the reduction of the value of the Pula “has taken consumers by surprise.”4 In another report, a woman being interviewed in the wake of the devaluation said that “this information should be disseminated while we can act. This was a pre-emptive action.” These statements clearly suggest that the central bank’s intervention was a surprise move to citizens. Indeed, even business actors in currency markets—who should, a priori, be among the most likely candidates to be well-informed about a monetary policy

intervention—expressed great surprise at the news of the devaluation. For instance, a BBC report stated that “Botswana has surprised the currency market by devaluing the Pula by 12%.” On May 31, 2005—the day after the devaluation became effective—the Mmegi newspaper quoted a chief executive officer of Stockbrokers Botswana—a registered member of the Botswana Stock Exchange—for saying that “the move has taken the market by surprise, particularly the magnitude of the devaluation and the timing.”5 A few days later, Stockbrokers Botswana (2005) issued a briefing paper commenting on the devaluation. While the company acknowledged the potential benefits of the devaluation to import-competing domestic producers and export companies, for example, the mining industry, it also stated that “we take issue with the brute force of the devaluation. It may have been more appropriate to introduce the new mechanism, explain it, and then take steps to devalue to the desired level in a more measured fashion. This would allow corporates and investors to plan for the adjustments and reduce the shock premium that the move will command. The danger is that where the market is shocked it will overreact . . .” (Stockbrokers Botswana 2005: 1).

This qualitative evidence supports two important points: First, neither the devaluation nor its timing was anticipated by the general public, and not even by businesses operating in currency markets. In that sense, it was an “exogenous” economic shock to citizens and the outcome we study, subjective evaluations of well-being. Second, although citizens are able to self-select into categories (like living in an urban area) that are correlated with treatment assignment, neither respondents nor the Afrobarometer had information, incentive, or capacity to decide whether respondents were interviewed before or after the devaluation, making direct self-selection into treatment highly improbable. Rather, the currency devaluation by the Bank of Botswana was an event that demarcated the respondents of the Afrobarometer survey into two groups, not because of the knowledge or decisions of respondents, but simply by chance. In Appendix S1, we provide further tests of the equivalence of the treatment and control groups on socio-economic background variables (the appendix is available at http://wber.oxfordjournals.org/). Appendix S2 and S3 also show results from regression and matching for observations that are on common support on the propensity score. These results do not change the main conclusions below.

III. DEVALUATION AND WELL-BEING: SIMPLE PRE- AND POST-TREATMENT COMPARISONS

To get a sense of the differences between pre- and post-treatment groups, this section shows the simple relationship between exposure to the devaluation and subjective well-being, as well as the development in food prices in the months surrounding the devaluation. The latter is important because it illustrates the

most plausible mechanism connecting the currency devaluation to individuals’ evaluations of well-being.

As dependent variable, we use respondents’ answers to the following question: “In general, how would you describe: Your own present living conditions?” Answers are given on a scale consisting of the categories “very bad,” “fairly bad,” “neither,” “fairly good,” and “very good,” where high values denote good living conditions. While the literature often uses questions concerning “life satisfaction” (Deaton 2008, 2012; Bjørnskov et al. 2010; Kahneman and Deaton 2010; Asadullah and Chaudhury 2012), the question we use asks respondents to evaluate their present living conditions on a scale from “very bad” to “very good,” which is clearly a constitutive feature of subjective well-being. We therefore use this question to measure subjective well-being.

Figure 1 shows a simple time-series plot of respondents’ average evaluations of their present living conditions (subjective well-being) for each day of the survey. Figure 2 shows a plot of the development in an index of food prices from July 2004 to September 2006, with the value of September 2006 indexed at 100 (Central Statistics Office 2008). The vertical lines indicate the timing of the devaluation. As is clearly visible in Figure 1, upon the devaluation of the Pula, there is an immediate and substantial drop in respondents’ average evaluations of living conditions in the magnitude of 0.16 on a scale from 0 to 1. Compared to individuals surveyed prior to the devaluation, the subjective well-being of people surveyed after the devaluation was much lower. The immediacy of this drop in well-being is important too, as prices are unlikely to have adjusted very much already on the first day after the devaluation. While there were media reports of upward re-pricing by retailers and a consequent “shock of skyrocketing prices” shortly after the devaluation, the price level of consumables did not fully adjust to its new equilibrium within the short period where the survey data were collected. As shown in Figure 2, food prices developed as expected in the months following the devaluation. While the food price index was relatively stable in the year preceding the devaluation, it increased dramatically in the year after the devaluation. This suggests that people’s reaction to the devaluation—the drop in

6. We note that although the wording is not identical to most surveys asking about the satisfaction with life as a whole, the two questions tend to produce quite similar results. Using the 2011 wave of the World Values Survey in Ghana, we note the similarity between the regular life satisfaction question and a question specifically on satisfaction with one’s financial situation. Less than 5 percent of respondents who declare themselves satisfied with their financial situation (rating it 8–10 on a 1–10 scale) declare themselves unsatisfied with their life as a whole.

7. The Afrobarometer also contains a related question, asking respondents to evaluate their living conditions relative to other people. Replications using this variable—evaluations of relative living conditions—does not change our findings substantially or statistically. Detailed results are available upon request. We do not think the two variables are sufficiently distinct to treat them as alternative measures, partly because questions asking people to rate their situation relative to others may pick up absolute and not relative differences (Karadja et al. 2014), and partly because the two living conditions questions are asked immediately after each other, which may make responses quite similar.

**Figure 1. Subjective Well-Being Around Time of Devaluation**

Notes: Vertical line shows timing of devaluation. Dashed curves are 95% confidence intervals.
Source: Authors’ analysis based on data sources discussed in text.

**Figure 2. Food Prices in Period Surrounding Devaluation**

Notes: Vertical line shows timing of devaluation. m1 = January; m7 = July.
Source: Authors’ analysis based on data sources discussed in text.
their evaluations of subjective well-being shown in Figure 1—is in large part driven by (qualitatively correct) expectations about the effects of the devaluation. Indeed, while the price effect of the devaluation materialized over months, there are good reasons to believe that people in Botswana knew what to expect, because 16 months earlier—in early February 2004—the Bank of Botswana also implemented a 7.5 percent devaluation of the Pula. While this did not make the May 2005 devaluation any less of a shock to people in Botswana, the prior experience with the consequences of a sizeable currency devaluation means that people may have rationally updated their expectations concerning the effects of the devaluation rapidly, even though the consequences of the May 2005 had not fully materialized at the time of the survey.

Figure 2 illustrates a second important point, namely that the Pula devaluation increased the price of imported food products and consumables in general, making consumers the major losers of the devaluation. A likely causal mechanism linking the currency devaluation to subjective well-being is therefore (expectations about) the development in prices, particularly the price level of food and consumables. During the time of the Afrobarometer survey in Botswana in late May and early June 2005, this was a very salient feature of the devaluation to the Batswana. In a report in the Mmegi newspaper, several people being interviewed who were employed in various low-wage jobs expressed concern at the consequences of the devaluation. A taxi driver reportedly stated that the expected price increases “. . . will have a devastating impact on our business and the economy at large.” In the same report, another employee is quoted for saying that “putting food on the table will empty wallets” and that “I am concerned and feel impoverished.” These examples suggest that people in Botswana had clear expectations about what consequences the devaluation would have for the price level of consumables and, therefore, for their own well-being. They also suggest that the expectations of increasing prices could be an important factor driving individuals’ feelings of being impoverished and are therefore the most likely causal mechanism linking the currency devaluation to the drop in subjective well-being we observe in Figure 1.

Although the relationship between the Pula devaluation and subsequent drops in subjective well-being is clear in Figure 1, we can use pretreatment observations as approximate counterfactuals for post-treatment observations only on the assumption that the devaluation is a plausibly exogenous shock to the citizens of Botswana. Given the imbalanced sampling of the pre- and post-treatment groups, the plausibility of the exogeneity assumption of course requires that we successfully condition on relevant confounders, most importantly by adjusting for rural-urban differences between the two groups as discussed above. However, as we show in the next section, neither the urban-rural shift nor a range of other potential confounders can fully account for the observed drop in subjective well-being following the devaluation. Detailed descriptions of all variables used in the econometric analyses along with summary statistics are available in Appendix S4.
IV. Empirical Results

To estimate the effect of the currency devaluation on subjective well-being, our econometric analyses use models for continuous and categorical data. First, we treat the dependent variable as continuous by converting the categorical responses into a variable that assigns a number to each response. Following this strategy, we construct a variable, which holds the values 0, 0.25, 0.5, 0.75, and 1 corresponding to the five response categories and use this as our dependent variable in a series of linear regressions. 9 As an alternative, we maintain the categorical nature of the data and estimate an ordered logit model, using the appropriate link function. In what follows, we report the coefficients of interest using both estimators to show that the results are qualitatively identical. Our starting point is the following linear regression model.

\[ y_i = a + dT_i + bX_i + e_i, \]  

where the dependent variable, \( y_i \), is respondent \( i \)'s evaluation of her present living conditions; \( T_i \) is the devaluation treatment indicator; and \( X_i \) is a vector of controls. The identifying assumption in (1) is that that \( T \) and \( e \) are orthogonal, \( \text{Cov}(T, e) = 0 \), conditional on \( X \), where the most import element in \( X \) is respondents’ rural-urban status. Throughout, standard errors are regionally clustered to allow for arbitrary correlation among respondents living in the same region. Table 1 shows the results.

Main Results

Panel A in Table 1 show results obtained using OLS regressions. Panel B shows the treatment coefficient from identical specifications obtained using ordered logit regressions. Throughout all models in Panel B, the ordered logits confirm the basic conclusion from the linear models of a negative association between the devaluation and respondents’ evaluations of their living conditions. Since the results are substantially similar, we comment only on the results in Panel A.

Column (1) in Panel A shows the unconditional association between the treatment and respondents’ evaluation of their living conditions. The point estimate of the treatment effect is negative and with a magnitude about 16 percentage points corresponds to the finding in Figure 1. The association is highly significant and corresponds to 60% of a standard deviation. In columns (2) and (3), respectively, we include an urban dummy and a capital (Gaborone) dummy. This serves to immediately alleviate concerns that our results are in fact driven by a shift in the sampling of respondents from urban (predominantly Gaborone) to rural areas. In column (2), the urban dummy barely changes the estimated association. In column (3), the Gaborone dummy does attenuate the association

9. This effectively amounts to a rescaling of the numerical values assigned to each response in the Afrobarometer survey such that our variable runs in the interval from 0 to 1.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
<th>(11)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Least Squares</strong></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><strong>Dependent variable: Subjective evaluation of living conditions</strong></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>Treatment</td>
<td>−0.16***</td>
<td>−0.14***</td>
<td>−0.09***</td>
<td>−0.12***</td>
<td>−0.13***</td>
<td>−0.16***</td>
<td>−0.15***</td>
<td>−0.12***</td>
<td>−0.14***</td>
<td>−0.07***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.035)</td>
<td>(0.028)</td>
<td>(0.025)</td>
<td>(0.027)</td>
<td>(0.042)</td>
<td>(0.059)</td>
<td>(0.028)</td>
<td>(0.030)</td>
<td>(0.025)</td>
<td></td>
</tr>
<tr>
<td>Botswana economic condition</td>
<td>0.11***</td>
<td>0.09***</td>
<td>0.08***</td>
<td>0.05***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.013)</td>
<td>(0.007)</td>
<td>(0.008)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Own past economic situation</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>Male dummy</td>
<td>0.03*</td>
<td>0.00</td>
<td>0.00</td>
<td>(0.012)</td>
<td>(0.015)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban dummy</td>
<td>0.05***</td>
<td>0.00</td>
<td>0.00</td>
<td>(0.022)</td>
<td>(0.026)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaborone dummy</td>
<td></td>
<td>0.12***</td>
<td>0.25***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.026)</td>
<td>(0.024)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poverty</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>Age</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>Age squared</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>Unemployment</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>No children</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>District fixed effects</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>(Continued)</td>
</tr>
<tr>
<td>Tribal fixed effects</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Occupational FE</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Sample centred on</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>discontinuity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
Table 1. Continued

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
<th>(11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>1,198</td>
<td>1,198</td>
<td>1,198</td>
<td>1,198</td>
<td>1,198</td>
<td>375</td>
<td>216</td>
<td>1,152</td>
<td>1,188</td>
<td>1,109</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.053</td>
<td>0.061</td>
<td>0.061</td>
<td>0.105</td>
<td>0.105</td>
<td>0.113</td>
<td>0.070</td>
<td>0.068</td>
<td>0.267</td>
<td>0.150</td>
<td>0.362</td>
</tr>
</tbody>
</table>

Panel B: Ordered logit

<table>
<thead>
<tr>
<th>Treatment</th>
<th>-1.07***</th>
<th>-0.89***</th>
<th>-0.62***</th>
<th>-0.91***</th>
<th>-1.12***</th>
<th>-0.91***</th>
<th>-0.97***</th>
<th>-0.97***</th>
<th>-0.93***</th>
<th>-0.97***</th>
<th>-0.65***</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.206)</td>
<td>(0.207)</td>
<td>(0.212)</td>
<td>(0.202)</td>
<td>(0.158)</td>
<td>(0.230)</td>
<td>(0.314)</td>
<td>(0.174)</td>
<td>(0.178)</td>
<td>(0.195)</td>
<td></td>
</tr>
</tbody>
</table>

**Sources:** Authors’ analysis based on data sources discussed in text. Treatment denotes the Pula devaluation. Days before devaluation are coded as 0; days after devaluations are coded as 1. All models contain a constant term (not reported to save space). Robust standard errors clustered at the region level in parentheses. *** p < .01, ** p < .05, * p < .1.
somewhat, but it remains sizeable and statistically significant. In the next section, we tackle the fundamental problem of nonrandom assignment in more depth.

In column (4) we proceed to include a full set of dummies for the geographical regions of Botswana registered by the Afrobarometer to remove as much idiosyncratic geographical variation as possible in how respondents evaluate their living conditions. The association drops marginally to 0.12 and remains negative and highly significant. In columns (5) and (6), we include fixed effects for respondents’ tribal affiliation (column [5]), and for each of the 25 occupational categories available in the Afrobarometer survey (column [6]). In both cases, the association between the devaluation treatment and subjective well-being remains substantively and statistically significant.10

In columns (7) and (8) we zoom in on the discontinuity in the data, i.e. the days immediately surrounding the devaluation. We do so to minimize the likelihood that some unobserved event occurring after—and close to—the treatment is confounding the results. In column (7), we focus on the four days surrounding the devaluation (two days before, two days after); in column (8) we focus on the first day before and the first day after the devaluation. This drastically reduces the sample size, but it does not change the main result: The size of the treatment coefficient is virtually unaffected as is its level of statistical significance. That is, zooming in on narrow bands around the discontinuity generated by the devaluation does not change the negative association between the devaluation treatment and subjective well-being.

In column (9), we control for respondents’ assessments of the country’s economic conditions, since this could plausibly affect how they perceive their own living conditions by supplying a signal of the existence of an overarching macro-economic problem. The treatment coefficient barely changes, however, and remains highly significant. In column (10), a control has been added for how respondents perceive their own past personal economic situation. This shows that even after removing the effect of respondents’ past economic situation, there is a very sizable and significantly negative change in the perception of living conditions following the Pula devaluation.

Finally, in column (11) both of these controls have been included together with the urban-rural indicator variable, gender, age and age squared, as well as a measure of poverty.11 While this lowers the coefficient of interest to −0.07, it is

10. In addition, we have experimented with categorizing particular occupations as export-exposed. However, we cannot know whether individuals within those occupations are indeed engaged in export activities or not. Furthermore, for any clear theoretical implication to hold, we would need to know whether the Marshall-Lerner condition holds in the short run for the particular occupation. As results are as mixed as the theoretical prerequisites, we refrain from showing them.

11. The poverty index is based on the work of Bratton et al. (2005) and measures poverty as respondents’ experience with lack of access to five basic types of household necessities: food, water, medicine, fuel to cook food, and cash income (Justesen and Bjørnskov 2014). The index comprises the sum of these five survey items. A principal component analysis show that all five items load onto the same component (alpha=0.74).
still highly significant and substantive, corresponding to approximately a third of a standard deviation. Since these observable variables are unable to account for the negative effect of the devaluation, we do not suspect that equally important unobservables are driving the estimated effect.

_Tackling Nonrandom Treatment Assignment_

As mentioned above, there are systematic differences between pretreatment and post-treatment responses since the former group was predominantly from urban areas (particularly the capital, Gaborone). This provides reason for caution because the shift from urban to rural respondents could plausibly coincide with a drop in evaluations of living conditions if, for example, more confident or optimistic individuals self-select into urban areas. While we dealt with this issue above, this section provides further tests that tackle the issue of nonrandom treatment assignment in more detail. We do so in Table 2 chiefly by removing respondents from the Gaborone area and respondents from urban (or rural) areas in general from the sample.12

In column (1), we report the basic unconditional association after omitting all respondents from Gaborone, which reduces the sample from 1,198 to 1,063 respondents. In absolute terms, the coefficient is reduced from 0.16 to 0.09, but it remains highly significant and shows that the relationship between the currency treatment and subjective well-being cannot be accounted for by the presence of respondents from the Gaborone area in the pretreatment group. In column (2), we continue to exclude respondents from Gaborone but also zoom in on the two days surrounding the devaluation (the first day before; the first day after). This does not change the results substantially either.

Column (3) shows the basic unconditional association, this time omitting all urban respondents. The familiar conclusion obtains also in a sample of rural respondents, which shows that our results are not driven by differences in evaluations of living conditions between urban and rural respondents. The model in column (4) again omits urban respondents and zooms in on the two days surrounding the devaluation, with little impact on the treatment effect. Column (5), finally, omits all rural respondents, focusing only on respondents from urban areas. This also leaves conclusions unchanged. For all model specifications, we find very similar results using ordered logit instead of OLS (as reported in Panel B).

To further document that the effect of the devaluation on subjective well-being cannot be reduced to the shift in the sampling of respondents from Gaborone to rural areas, we have performed a series of placebo tests, repeating some of our analyses using data from Round 4 (2008) of the Afrobarometer. In

12. A separate issue is that the treatment divides the sample between weekend and weekdays. If subjective evaluations were, for some reason, more positive during weekends, our results would be biased (Helliwell and Wang 2011). However, in further estimates (available upon request) we show that this is not the case in the present sample or the subsequent fourth round of the Afrobarometer survey in Botswana.
these tests, we define a placebo treatment indicator as living outside Gaborone (or urban areas more generally). If our results were in fact driven by differences in evaluations of living conditions between respondents in the capital (or urban areas) and elsewhere, the coefficient on this placebo treatment indicator should be similar in size to the coefficient on the treatment indicator reported above. However, as we document in Appendix S5, across various model specifications the difference between Gaborone and the rest of Botswana is never more than 0.07 in Round 4 of the survey. And in some cases it is both statistically and substantively indistinguishable from zero. With the Round 3 data we use here, in contrast, the coefficient of interest is consistently significant and negative, in the magnitude of $-0.16$. This provides additional confirmation that our results are not driven by nonrandom treatment assignment of survey respondents. We did similar placebo tests using as treatment the first two days of the survey from Round 4 (Appendix S5). This reveals that in Round 4 there was no discontinuity in respondents’ evaluations of living conditions after two days of surveying.

**Conditioning Effects of Information and Cognitive Sophistication**

So far we have documented a strong effect of the shock devaluation on subjective well-being. However, as mentioned earlier, there may be reason to expect that people with higher levels of information and cognitive sophistication display

### Table 2. Robustness Tests

<table>
<thead>
<tr>
<th>Panel A: Least Squares</th>
<th>Dependent variable: Subjective evaluation of living conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>$-0.09^{***}$</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
</tr>
<tr>
<td>Excluding Gaborone</td>
<td>Yes</td>
</tr>
<tr>
<td>Sample centered on discontinuity</td>
<td>No</td>
</tr>
<tr>
<td>Excluding urban respondents</td>
<td>No</td>
</tr>
<tr>
<td>Excluding rural respondents</td>
<td>No</td>
</tr>
<tr>
<td>Observations</td>
<td>1,063</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.008</td>
</tr>
</tbody>
</table>

**Panel B: Ordered logit**

| Treatment              | $-0.63^{***}$ | $-0.64^{**}$ | $-1.06^{***}$ | $-0.89^{***}$ | $-0.61^{*}$ |
|                        | (0.173)       | (0.091)      | (0.129)       | (0.114)       | (0.317)       |

*Sources: Authors’ analysis based on data sources discussed in text. Treatment denotes the Pula devaluation. Days before devaluation are coded as 0; days after devaluations are coded as 1. All models contain a constant term (not reported to save space). Robust standard errors clustered at the region level in parentheses. *** $p < .01$, ** $p < .05$, * $p < .1$.**
stronger and more immediate responses to the news of the devaluation. Specifically, individuals with more informed and sophisticated mental models of the economy may make more accurate predictions of the consequences of the devaluation and update their expectations about the future more rapidly. In Table 3 we examine whether the association between subjective well-being and the macroeconomic shock depends on respondents’ level of information and cognitive sophistication. To operationalize information we construct a dummy variable where we treat informed respondents as those who report getting daily news from the radio, television, or newspapers (coded 1). News consumption must be on a daily basis to moderate the observed drop in subjective well-being already on the day following the devaluation. If respondents do not follow the news on a

Table 3. Information, Education, and the Effect of the Treatment

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Least squares</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>-0.08***</td>
<td>-0.05</td>
<td>-0.05*</td>
<td>-0.08***</td>
<td>-0.07***</td>
<td>-0.06**</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.029)</td>
<td>(0.027)</td>
<td>(0.023)</td>
<td>(0.020)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Daily news consumption</td>
<td>0.16***</td>
<td>0.15***</td>
<td>0.13***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.014)</td>
<td>(0.015)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment-news interaction</td>
<td>-0.08***</td>
<td>-0.08***</td>
<td>-0.07**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.024)</td>
<td>(0.023)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>0.04***</td>
<td>0.03***</td>
<td>0.03***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment-education interaction</td>
<td>-0.01**</td>
<td>-0.01**</td>
<td>-0.01*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.004)</td>
<td>(0.005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban dummy</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.019)</td>
<td>(0.020)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Own past economic situation</td>
<td>0.07***</td>
<td></td>
<td></td>
<td>0.07***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td></td>
<td></td>
<td>(0.006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupational fixed effects</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>1,196</td>
<td>1,196</td>
<td>1,186</td>
<td>1,194</td>
<td>1,194</td>
<td>1,184</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.083</td>
<td>0.135</td>
<td>0.211</td>
<td>0.103</td>
<td>0.133</td>
<td>0.205</td>
</tr>
<tr>
<td>Panel B: Ordered logit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>-0.57***</td>
<td>-0.34*</td>
<td>-0.37*</td>
<td>-0.56***</td>
<td>-0.44***</td>
<td>-0.36*</td>
</tr>
<tr>
<td></td>
<td>(0.164)</td>
<td>(0.203)</td>
<td>(0.203)</td>
<td>(0.157)</td>
<td>(0.133)</td>
<td>(0.189)</td>
</tr>
<tr>
<td>Daily news consumption</td>
<td>1.11***</td>
<td>1.03***</td>
<td>0.97***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.082)</td>
<td>(0.100)</td>
<td>(0.101)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment-news interaction</td>
<td>-0.54***</td>
<td>-0.56***</td>
<td>-0.48**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.174)</td>
<td>(0.203)</td>
<td>(0.193)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>0.28***</td>
<td>0.22***</td>
<td>0.20***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.017)</td>
<td>(0.038)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment-education interaction</td>
<td>-0.08**</td>
<td>-0.08***</td>
<td>-0.09**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.026)</td>
<td>(0.044)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sources: Authors’ analysis based on data sources discussed in text. Treatment denotes the Pula devaluation. Days before devaluation are coded as 0; days after devaluations are coded as 1. All models contain a constant term (not reported to save space). Robust standard errors clustered at the region level in parentheses. ***p < .01, **p < .05, *p < .1.
daily basis, we treat them as uninformed (coded 0). As a proxy for cognitive sophistication, we use respondents’ level of education (see Appendix S4 for details).

To examine whether information and cognitive sophistication condition the relationship between the currency devaluation and subjective well-being, we augment the regression model (1) with, first, an interaction of the treatment indicator and our measure of information and, second, an interaction of the treatment and education, our proxy for cognitive sophistication. Panel A in Table 3 shows results from linear regressions, while Panel B shows coefficients from identical ordered logit models. As in Tables 1 and 2, across specifications the conclusion that follows from these models confirms the OLS models in Panel A.

Consistent with our expectations, the coefficients in column (1) show that the association between the devaluation and subjective evaluations of living conditions is stronger if respondents are well informed. Thus, while the coefficient on the treatment indicator remains significantly negative at 0.08, treated respondents with daily news consumption evaluate their living conditions to worsen by an additional and significant 0.08. Similar conclusions follow from the specifications in columns (2)–(3), where controls for urban residence, respondents’ perceptions of their past personal economic situation, and occupation fixed effects are added. This suggests that individuals with higher levels of information more quickly update their perceptions of well-being.14

In columns (4)–(6), we interact the treatment indicator with respondents’ education. Here we find that higher levels of education strengthen the association between the treatment and respondents’ negative evaluations of their living conditions. We show this in Figure 3 by plotting the marginal effect of the currency treatment at different values of education (cf. Brambor et al. 2006) along with 90 percent confidence intervals (indicated by the dotted lines). While the devaluation shock causes a drop in subjective well-being even for people with no formal education (values of zero on the education variable), Figure 3 clearly shows that the negative effect increases and becomes more significant as respondents’ educational level increases.

The conditioning effects of information and education are both intuitive. In order to understand the effect of a devaluation on (future) living conditions, people must be reasonably informed about the devaluation and have mental models that allow them to predict the future consequences of the devaluation. Even so, the fact that respondents who follow news on a daily basis give more negative responses following the devaluation need not reflect cognitive sophistication but can also reflect respondents’ ability to mimic and absorb the evaluation of experts reported in the news. However, higher levels of cognitive

14. The devaluation might plausibly affect rich and poor individuals differentially. Within occupational groups, however, there are no signs of a heterogenous treatment effect between rich and poor (results available on request).
sophistications in the form of education also seem to strengthen the effect of the devaluation on respondents’ subjective well-being. This probably reflects both increased consumption of daily news among this group of respondents and that education increases individuals’ knowledge about the future consequences of the devaluation and their consequent ability to form rational expectations. Overall, these results suggest that the devaluation shock did on average result in drops in subjective well-being for all citizens of Botswana, but that the negative effect is conditional in nature and larger for people with higher levels of information and education.

V. Conclusions

This article documents a strong and significantly negative effect of monetary shocks on subjective well-being. Using the case of a central bank devaluation in Botswana as a quasi-experiment, our results show that people’s subjective well-being dropped immediately after the news of the devaluation was released in the public. As we have documented, this result is extremely robust and persists even when plausible sources of nonrandom treatment assignment are dealt with. The results therefore provide robust evidence that monetary shocks in the form of unanticipated currency devaluations have a strong and negative causal effect on how people rate their living conditions and personal well-being.
Moreover, people who are well informed through higher levels of news consumption and people with higher levels of education respond more strongly to the news of the devaluation. This suggests that the effect of monetary shocks on subjective well-being is conditional on individuals’ levels of information and cognitive sophistication and not merely an effect of real economic change in the very short run. Given the short time period for which we have data—the days in which the survey was conducted in Botswana—we cannot say anything about how quickly well-being might recover following an economic shock like the one we study. However, our results strongly suggest that macroeconomic shocks, such as unanticipated currency devaluations, may have significant short-term costs in the form of reductions in people’s sense of well-being.

References


Does Access to Foreign Markets Shape Internal Migration? Evidence from Brazil

Laura Hering and Rodrigo Paillacar

This paper investigates how internal migration is affected by Brazil’s increased integration into the world economy. We analyze the impact of regional differences in access to foreign demand on sector-specific bilateral migration rates between the Brazilian states for the years 1995 to 2003. Using international trade data, we compute a foreign market access measure at the sectoral level, which is exogenous to domestic migration. A higher foreign market access is associated with a higher local labor demand and attracts workers via two potential channels: higher wages and new job opportunities. Our results show that both channels play a significant role in internal migration. Further, we find a heterogeneous impact across industries, according to their comparative advantage on the world market. However, the observed impact is driven by the strong reaction of low-educated workers to changes in market access. This finding is consistent with the fact that Brazil is exporting mainly goods that are intensive in unskilled labor. JEL codes: F16, F66, R12, R23

INTRODUCTION

A considerable amount of literature provides evidence that a country generally benefits from opening up to international trade. However, within the country, these benefits are often unevenly distributed. This can cause a rise in regional wage disparities, both across and within industries, which may lead to changes in the spatial distribution of the domestic economic activity.

In this paper, we investigate how internal migration is affected by Brazil’s increased integration into the world economy. More specifically, we analyze the impact of changes in foreign demand for Brazilian goods on sector-specific bilateral migration rates between the 27 Brazilian states for the years 1995 to 2003.

Laura Hering (corresponding author) is an assistant professor at the Erasmus School of Economics (Department of Economics) and a Tinbergen Research Fellow; her email address is laura.hering@gmail.com. Rodrigo Paillacar is an assistant professor at the University of Cergy-Pontoise (Laboratoire THEMA); his email address is rodrigo.paillacar@gmail.com. This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01). We thank the editor, two anonymous referees, Maarten Bosker, Matthieu Couttenier, Fabian Gouret, Philippe Martin, Sandra Poncet, Loriane Py, Cristina Terra, Vincent Rebeyrol and Gonzague Vannoorenberghe for their helpful suggestions. A supplemental appendix to this article is available at http://wber.oxfordjournals.org/.

THE WORLD BANK ECONOMIC REVIEW, VOL. 30, NO. 1, pp. 78–103 doi:10.1093/wber/lhv028 Advance Access Publication April 29, 2015 © The Author 2015. Published by Oxford University Press on behalf of the International Bank for Reconstruction and Development / THE WORLD BANK. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.
In order to identify the effect of international trade on the local labor market in a specific sector, we compute a region-sector specific measure of foreign demand, which is derived from a standard gravity equation that can be obtained from various trade models. The location of the region with respect to its potential trading partners plays a key role in determining a region’s market access. Firms located in regions closer to large consumer markets have a higher market access due to lower trade costs, thereby giving them a competitive advantage in these markets. An increase in a region’s market access therefore reflects a higher demand for its products and consequently a higher labor demand.

We show in this paper that an increase in a region’s access to foreign markets attracts migrants via two channels: i) an indirect effect via an increase in the local wage premium and ii) a direct effect resulting from the creation of new job opportunities.

The positive effect of foreign market access on wages is already well documented for various countries, including Brazil (Fally et al. 2010). In this paper we focus on the second channel, which captures the impact of market access on migration beyond its effect via a change in local wages.

Higher market access is expected to also have a direct effect on migration essentially due to a higher number of vacancies, which increases the probability of employment. Alternatively, the type of jobs created as a result of an increased foreign demand can be considered to be of better quality. In Brazil, as in many emerging countries, firms in the export industry are preferred employers. Next to a higher employment probability, an increase in the market access variable can thus also capture long-term considerations in the migration decision. These aspects are typically excluded when migration is modeled as depending only on spot wages, which themselves cannot capture the workers’ wage profile or non-pecuniary aspects linked to the job (Aguayo-Tellez et al. 2010).

Our sector-specific foreign market access measure identifies the net effect of foreign demand on the local labor market. Note that a positive shock to foreign market access does not necessarily mean that only jobs in exporting firms will be created. Due to spillovers or an increase in connected activities (e.g., outsourced tasks), the increase in demand for exported goods may also lead to a change in labor demand in non-exporting local firms in the same sector.

The main advantage of our market access measure is that it is by construction exogenous to domestic factors, such as local labor market regulations or a region’s comparative advantage in the supply of goods in a specific sector. Thus, we do not risk confounding the role of foreign demand with local characteristics,

---

1. The impact of market access on wages is by now well studied empirically. See, among others, Hanson (2005) for the United States, Head and Mayer (2006) for Europe, and Hering and Poncet (2010) for China. The theoretical link is modeled explicitly in the so-called “New Economic Geography wage equation” (Fujita et al. 1999), but Head and Mayer (2011) point out that such wage equations can be established in numerous trade models.

2. Exporters are likely to offer more long-term employments, propose a steeper wage gradient and better working conditions (see e.g., Wagner (2012) for an overview).
in particular the local export capacity, which may be affected by domestic migration.³

Performing the analysis of bilateral migration at the sectoral level is motivated by some recent studies on Brazil’s labor market, which present evidence for a very low sectoral mobility of Brazilian workers (Menezes-Filho and Muendler, 2011; Muendler, 2008). Therefore, in this paper we focus on labor migration that takes place within sectors.⁴

The sectoral approach has two important advantages, which we exploit in our identification strategy. First, in contrast to our sectoral measure, an aggregated market access variable would be potentially correlated with the evolution of other unobserved migration determinants that vary over time and across states (i.e., amenities, price levels, institutional quality). Constructing migration rates and market access by sector allows us to include year-location fixed effects, which control for these unobserved location characteristics. Second, this allows us to study the heterogeneous effect of market access across industries.

Our results show that regional differences in access to international markets indeed affect internal migration patterns. Foreign demand impacts migration also directly and not only by means of an increased wage level. These findings suggest that new job opportunities created by higher foreign demand are important location determinants.

Further, our results indicate that the effect of market access is generally stronger, the higher the industry’s comparative advantage is on the world market.

Moreover, we find that the impact of market access on sectoral migration rates is driven by the low-educated workers. This could be explained by Brazil’s relative abundance of low-skilled labor. A higher market access represents a stronger increase in demand for goods intensive in low-skilled labor, in which Brazil has a comparative advantage on the world market (Muriel and Terra, 2009). Thus, these workers are more likely to be affected by a change in the foreign demand.

Although several studies explore the link between trade and migration, they have mostly focused on international migration patterns (cf., for example, Ortega and Peri, 2013, and Letouzé et al. 2009). Yet, internal migration flows have a far greater magnitude than international flows and hence may modify a country’s development path much more sensibly. This is of particular relevance in fast urbanizing developing countries like Brazil.

Closest to our work is the paper by Aguayo-Tellez et al. (2010), which also applies to Brazil. These authors show that workers in formal employments are attracted to states with a higher concentration of foreign owned establishments. We differ from that paper in that we we focus only on employment opportunities

³. This is possible because our approach allows us to separate the foreign demand from a region’s production and export capacity. By excluding all supply side factors from our market access measure, we eliminate the possibility of reverse causality between internal migration and international market access.

that are created by a change in foreign demand. However, as explained above, these new vacancies can also be in non-exporting and domestically owned firms. Further, our analysis also includes informal workers, who account for at least 38 percent of the Brazilian workforce (Henley et al. 2009).

A few papers have studied the role of imports in the location choice of individuals and can be considered as complementary to our work. Kovak (2011; 2013) studies the effect of import competition on internal migration patterns in Brazil. He finds that regions specialized in industries experiencing larger tariff cuts see their wages decrease, which in turn triggers outmigration. In the same spirit, Autor et al. (2013) show how import competition from China affects local labor markets in the United States. They find that stronger import competition is associated with a higher reduction in manufacturing employment. However, their setting requires internal migration in reaction to trade shocks being negligible.

**Empirical Methodology**

The empirical specification of our migration equation is based on an additive random utility model. Every individual $k$ from location $i$ maximizes the indirect utility $V_{kij}$ across all possible destinations $j$. In a general utility differential approach, the individual location choice $M_{kij}$ can then be written as:

\[ M_{kij} = 1 \quad \text{if and only if} \quad V_{kij} = \max(V_{ki1}, \ldots, V_{kij}), \quad \text{otherwise} \quad M_{kij} = 0. \]

The indirect utility $V_{kij}$ can be decomposed as follows:

\[ V_{kij} = X_{ij}\beta + \xi_{ij} + e_{kij} \tag{1} \]

where $X_{ij}$ are the characteristics of location $j$. The subscript $i$ is included, as characteristics of $j$ can vary across original locations $i$ (e.g., bilateral distance). $\beta$ is a vector of marginal utilities and $\xi_{ij}$ represents unobserved location characteristics. The idiosyncratic error term $e_{kij}$ is included to allow individuals from the same origin to choose different locations. We make the standard assumption that this error term follows an i.i.d. Type I extreme value distribution.

5. In our empirical analysis, the presence of exporters and foreign owned firms is controlled for via location-year fixed effects.

6. Note also that their proxy of trade exposure is only region-time specific. Since we exploit the sectoral dimension and control for location-time fixed effects, we automatically account for this measure.

7. This model choice is standard in the recent migration literature and is used, for example, in Grogger and Hanson (2011) and Kovak (2011). For a detailed description on the derivation of the empirical specification see Bertoli and Fernández-Huertas Moraga (2013).
Given that individuals select the location that maximizes their utility, the probability that an individual from $i$ will choose destination $j$ is defined by

$$\Pr(V_{kij} > V_{kim}) \forall j \neq m$$

Replacing the indirect utilities by their definitions of equation 1 and rearranging terms, the probability that individual $k$ will move from $i$ to $j$ is given by:

$$\Pr(e_{kij} - e_{kim} > X_{im}\beta - X_{ij}\beta + \xi_{im} - \xi_{ij}) \forall j \neq m$$

McFadden (1974) shows that under the assumption of an i.i.d. extreme value distribution of the individual error term, migration probabilities can be expressed as

$$\Pr(M_{kij} = 1) = \frac{\exp(X_{ij}\beta + \xi_{ij})}{\sum_{j=1}^{J}\exp(X_{ij}\beta + \xi_{ij})} = s_{ij}$$

Following Berry (1994), this individual migration probability can be interpreted as the share of individuals from $i$ migrating to $j$, $s_{ij}$. Similarly, the share of stayers of region $i$, $s_{ii}$, can be written as

$$\Pr(M_{kii} = 1) = \frac{\exp(X_{ii}\beta + \xi_{ii})}{\sum_{j=1}^{J}\exp(X_{ii}\beta + \xi_{ij})} = s_{ii}$$

Dividing equation 4 by equation 5 and taking the log yields

$$\ln\left(\frac{s_{ij}}{s_{ii}}\right) = \ln\left(\frac{\exp(X_{ij}\beta + \xi_{ij})}{\exp(X_{ii}\beta + \xi_{ii})}\right) = \beta(X_{ij} - X_{ii}) + \xi_{ij} - \xi_{ii}$$

We now have an aggregate discrete choice model that accounts for unobserved location characteristics $\xi$ and whose parameters can be estimated using conventional linear estimation techniques.

To obtain our empirical specification, we add the time dimension $t$ and the sectoral dimension $s$ and replace the vector $X$ with our location-sector specific variables of interest. This gives us our first benchmark specification:

$$\ln m_{ijst} = \ln\left(\frac{s_{ijst}}{s_{ii}}\right) = \alpha + \beta_1 \Delta MA_{ijst} + \beta_2 \Delta w_{ijst} + FE_{ij} + FE_{st} + FE_{it} + FE_{jt} + e_{ijst}$$

8. Here we make the implicit assumption that workers do not switch sectors, and thus their migration decision depends only on state characteristics (e.g., price level) or the characteristics of their own sector (e.g., sectoral market access).
$m_{ijst}$ is the observed migration rate between state $i$ and $j$ for sector $s$ in the household survey of year $t$. It is simply defined as the number of migrants going from $i$ to $j$ divided by the number of stayers. Individuals are considered as migrants when they declare having lived five years ago ($t - 5$) in a different state than their current state of residence. Since we do not know the exact moment of migration, all independent variables are constructed as means over the years $t-4$ to $t-1$. This is indicated by the index $\tilde{t}$.9

Our main variable of interest is the market access gap between states $i$ and $j$ for sector $s$, $\Delta MA_{ijst} = ln\frac{MA_{ijst}}{MA_{is}}$. An increase in this variable makes state $j$ relatively more attractive, either because of i) a higher wage level or ii) new job opportunities (more or better jobs). We can isolate the second channel by including the wage gap, $\Delta w_{ijst}$, in our benchmark specification. Adding the wage variable has an additional important advantage: it also captures other sector and time varying characteristics of the local labor market that we cannot observe but which are potentially correlated with foreign market access (e.g., sector-specific productivity differentials).

A lower number of available jobs typically also corresponds to a higher unemployment rate. But a higher unemployment rate can also reflect limitations on the labor supply side or a mismatch on the local labor market between vacancies and job seekers. While in some specifications we explicitly include regional differences in unemployment rates, our benchmark estimation includes $FE_{it}$ and $FE_{jt}$, which correspond to origin-year and destination-year dummies. These account for time-varying differences across states, including the unemployment rate, amenities or price levels, which are also considered to be important determinants of migration.

Bilateral fixed effects $FE_{ij}$ take into account time-invariant specificities concerning migration between two particular states (e.g., moving costs, migration networks). $FE_{it}$ represents sector-year fixed effects.

In the presence of these numerous sets of fixed effects, we identify $\beta_1$ by exploiting the variation of market access within the same pair of states over time and across industries. The exact ranking of market access across states or sectors is therefore not of importance.

By definition, $\varepsilon_{ijst}$ is a i.i.d. bilateral error term. However, using equation 6 it can be shown that all $\varepsilon_{ijst}$ from the same origin $i$ depend on the same $\xi_{it}$. This leads to a non-zero covariance of $\varepsilon_{ijst}$ for observations with the same origin $i$ in year $t$. In all our regressions, we therefore cluster our standard errors by the state of origin-year level. Appendix S3 discusses the assumption of the independence of irrelevant alternatives (IIA) that is underlying our model.

---

9. Our benchmark results hold also when specifying our independent variables as four-year lags instead of the mean over the previous four years.
MARKET ACCESS: DERIVATION AND CONSTRUCTION

Theoretical Derivation of Market Access

In this subsection, we provide the formal definition of market access and how it can be derived from a standard gravity model of trade.\(^{10}\)

According to structural gravity models, exports \(EX_{ij} \) in sector \(s \) from region \(i \) to partner \(j \) can be written as

\[
EX_{ij} = \phi_{ij}S_{is}M_{js} = \phi_{ij} \frac{Y_{is}}{P_{is}} \frac{E_{js}}{P_{js}} \frac{S_{is}}{M_{is}}
\]  

(8)

with \(0 \leq \phi_{ij} \leq 1\). This equation decomposes exports into three components: The term \(\phi_{ij} \) reflects the accessibility of market \(j \) for the exporters from location \(i \) in sector \(s \). A \(\phi_{ij} \) of 1 indicates free trade and \(\phi_{ij} = 0 \) refers to prohibitively high trade costs and thus zero exports.

The terms \(S_{is} \) and \(M_{js} \) are often referred to in the literature as the supply and market capacity. They capture all the considerations that make exporter \(i \) a competitive exporter and partner \(j \) an attractive destination in sector \(s \). More precisely, the supply capacity depends on the total output \(Y_{is} = \sum_j EX_{ij} \) of sector \(s \) in location \(i \), as well as the local firms’ price competitiveness, \(\Pi_{is} \). The market capacity of \(j \) in sector \(s \) depends on location \(j \)’s total expenditure on goods from sector \(s \), \(E_{js} = \sum_i EX_{ij} \), and the prevailing price index in sector \(s \) on market \(j \), \(P_{js} \).

The terms \(\Pi_{is} \) and \(P_{js} \) are the so-called outward and inward “multilateral resistance terms” (Anderson and van Wincoop 2003). These terms take into account that bilateral trade relationships are affected by competition from third countries.

Given equation 8, region \(i \)’s relative access to every individual market \(j \) for sector \(s \) is defined by \(\frac{E_{js}}{P_{js}}\phi_{ij} \). Region \(i \)’s total market access in sector \(s \) can be obtained by summing over all destinations \(j \):

\[
MA_{is} = \sum_j \frac{E_{js}}{P_{js}}\phi_{ij} = \sum_j \phi_{ij}M_{js}
\]  

(9)

\(MA_{is} \) measures the overall ease for firms in location \(i \) to access all domestic and foreign markets \(j \) in sector \(s \). It represents an expenditure-weighted average of

---

10. This subsection borrows from the presentation of the general framework in Head and Mayer (2013). Although initially derived from a trade model of monopolistic competition, these authors show how market access can be obtained also in other market structures, notably in a setting with perfect competition and technology differences (Eaton and Kortum 2002), or in trade models accounting for firm heterogeneity (Chaney, 2008).
relative access, as it weights the market capacity of each potential destination $j$ by their accessibility from region $i$.

By summing only over foreign countries, we obtain an international market access measure, which solely captures the demand for goods from location $i$ coming from abroad.

**Market Access Calculation**

We estimate the market access measure presented in equation 9 via a gravity trade regression, following Redding and Venables (2004). This methodology is rarely applied in regional studies because of data limitations: bilateral trade flows are often unavailable at the subnational level, particularly for developing countries. Brazil is a fortunate exception since it provides information on international trade flows at the sectoral level for each of its twenty-seven states.

Our trade data set covers the years 1991 to 2002 and eleven sectors.\textsuperscript{11} It contains international trade flows between the twenty-seven Brazilian states and 170 partner countries and flows among the 170 foreign countries.

The empirical specification of the trade equation follows from equation 8. After taking the logs, we obtain

$$\ln EX_{ijs} = \ln f_{ijs} + \ln S_{is} + \ln M_{js} \quad (10)$$

For the calculation of a sector-state specific market access variable that varies over time, we estimate equation 10 separately for every sector-year pair.

In the regressions, sector-specific market capacity ($M_{js}$) and supply capacity ($S_{is}$) of every trading partner are captured by sector-importer ($FM_{js}$) and sector-exporter ($FX_{is}$) fixed effects. $\phi_{ijs}$ can be specified using different measures of trade costs. Specifically, we consider bilateral distance ($d_{ij}$), whether partners share a common border ($B_{ij}$), the presence of a free trade agreement between the two trading partners ($RTA_{ij}$) and whether the two are members of the WTO or its predecessor GATT ($WTO_{ij}$). Since we estimate the trade equation separately for every sector-year pair, we can drop the subscript $s$. Our empirical specification of the trade equation can then be written as

$$\ln EX_{ij} = \delta \ln d_{ij} + \lambda_1 B_{ij} + \lambda_2 RTA_{ij} + \lambda_3 WTO_{ij} + FX_i + FM_j + \nu_{ij} \quad (11)$$

where $\nu_{ij}$ is a random bilateral error term.

In total, we run 132 regressions (12 years $\times$ 11 sectors). Given that all coefficients and fixed effects are allowed to vary over time and across sectors, this enables us to build a time-varying market access specific for each state-sector combination.

\textsuperscript{11} For details on sectoral classification and data sources for the variables used in this section see appendix S4.1 and S4.2.
Market access for state \( i \) in sector \( s \) in year \( t \) is built by weighting each predicted market capacity, \( M_{jst} \), by the estimates of the corresponding bilateral trade costs, \( \phi_{jist} \). These weighted market capacities are then summed up to one single variable per state-sector pair:

\[
MA_{ist} = \sum_{j} \phi_{jist} M_{jst} = \sum_{j} \exp(\delta_{st} \ln d_{ij} + \lambda_{1st} B_{ij} + \lambda_{2st} RTA_{ij} + \lambda_{3st} WTO_{ij} + \lambda_{FM} M_{jst})
\]

We sum over \( R \) countries, where \( R \) includes only foreign countries and not the Brazilian states. This way, market access exclusively captures the foreign demand addressed to each Brazilian state. Market access thus differs from predicted exports as it excludes the local supply capacity.

Our measure can be considered exogenous to bilateral migration rates since all effects of internal migration on the states’ exports (imports) are captured by the estimates of the export (import) capacities of the Brazilian states. These are however not included in our measure. By excluding the exporter fixed effects, we ensure that our measure is exogenous to all domestic factors that affect the state’s export supply capacity, such as its comparative advantage in sector \( s \), the local infrastructure or changes in the labor force.

By focusing on foreign market access we eliminate the possible reverse causality that can arise when immigrants raise local consumption and hence the local market capacity: a local shock inducing the arrival of additional migrants may increase consumption in the host region and thus domestic market access but does not affect the access to foreign markets.

Finally, also the variables to proxy trade costs can all safely be regarded as exogenous to internal migration within Brazil (at least for the time horizon under study).

Table A-1 summarizes by industry the coefficients obtained from the trade regressions (equation 11). Coefficients on the trade cost variables have the expected sign, and magnitudes are in line with the literature (cf. Head and Mayer, 2013). However, there are some important differences across sectors, in particular in the distance coefficient. The last column summarizes the time varying importer fixed effect, representing the sector-specific market capacity of each destination country. Appendix S1.1 provides some descriptive statistics. Appendix S1.2 calculates various alternative market access measures.

12. To be consistent across sectors and years, each \( MA_{is} \) is constructed using the estimated market capacities and trade costs of always the exact same one hundred countries. These are the countries that import goods from all sectors in all years and thus provides us for all sector-year combinations with the necessary estimates for trade costs and importer fixed effects.

13. We present also robustness checks including the difference in the states’ exporter fixed effects as control variable (\( \Delta \text{supply}_{ijst} \)), to verify that our market access coefficient is not correlated to supply factors.
HOUSEHOLD SURVEY DATA

Our main data set is the yearly household survey Pesquisa Nacional por Amostra de Domicílios (PNAD) collected by the Brazilian Institute of Geography and Statistics (IBGE). The PNAD does not follow individuals but interviews a different random and representative sample of residents each year (between 310,000 and 390,000 per year). We use the PNAD for the years 1992 to 2003 (with data missing for 1994 and 2000).14

Migration Rates

We identify an individual as a migrant when the answer given to the question “In which Brazilian state did you live five years ago?” differs from the actual state of residence. Our sample is limited to individuals who declare having a job in a tradeable sector, earning a positive wage, having lived in Brazil five years ago and being between twenty and sixty-five at the time of the interview.

We distinguish eleven tradeable sectors that can be matched with the trade data and construct bilateral migration rates separately for each sector.15 We do not have any information about the individual’s work five years ago. Nevertheless, as argued above, we can make the reasonable assumption that individuals already worked in the same sector as in the year of the survey. Bilateral migration rates are then defined as the number of migrants from state \(i\) to \(j\) over the number of workers that stayed in state \(i\) and declare working in sector \(s\) at the time of the interview. In table 3, we rely on sectoral migration rates constructed separately by educational attainment. The workers are treated as highly educated if they attended high school for at least one year; otherwise they are regarded as low educated.16

Despite the presence of a relatively high number of zero migration flows among the states, the PNAD is considered to be representative of overall migration rates and thus adequate for studying migration patterns within Brazil (Fiess and Verner, 2003; Cunha, 2002). In robustness checks, we will also address the problem of unobserved flows by running Poisson-Maximum-Likelihood estimations including zero-flows.17

In our final data set, close to 3 percent of the individuals have moved states at least once within five years prior to the interview. Even though most of the migrants are low qualified in absolute terms, the highly educated individuals are the

14. In 1994 the PNAD was not conducted because of a strike. 1991 and 2000 were years of the population census.
15. See appendix S4.1 for details on the industrial classification.
16. In our empirical analysis, we exclude migration rates that are constructed with less than six observations. Results are robust when maintaining all observed flows and when omitting the top five and bottom five percent of migration rates. Also, using a sample limited to household heads yields overall very similar results (available upon request).
17. We have 7722 potential origin-destination-sector cells \((27 \times 26 \times 11)\) but observe at least one positive migration rate for only 1748 cells. In the Poisson estimations we replace all missing values with zeros for these 1748 sector-origin-destination combinations.
more mobile throughout the years (2.75 percent versus 3.53 percent). Table A-2 compares interstate mobility across sectors. Whereas less than 3 percent of the workers in basic metals, machinery, textile, and agriculture migrated within the last five years, this percentage is above 4 percent in the wood industry.

### Sector-State Specific Wages

Our key control variable is the sector-specific wage gap, $\Delta w_{ijsT}$. This variable accounts for most sector-state specific characteristics of the Brazilian labor market (such as sectoral and regional variations in employment regulations and labor productivity). Moreover, it controls for the indirect impact of market access on migration.

However, due to endogeneity concerns, we do not rely on the observed average wage levels. The main potential source of endogeneity in our case stems from self-selected migration. The personal characteristics (e.g., education, age) that drive the location choice are also major wage determinants. Thus, the observed wage level in a region depends on the composition of the local labor force, including the immigrants. We treat this issue by correcting for self-selected migration, following the methodology developed by Dahl (2002).\(^{18}\)

$\Delta w_{ijsT}$ is constructed from estimates of a modified Mincerian wage equation that is run separately for every state-year combination $jiT$.\(^{19}\) The obtained parameters on the individual characteristics are then used to predict wages that each individual $k$ would potentially earn in each of the twenty-seven states in year $t$. The effect of sector-specific market access on the wage level is accounted for by sector fixed effects.

The final wage gap for year $t$ is defined as

$$\Delta w_{ijsT} = \bar{w}_{ijsT} - \bar{w}_{ijsT}$$

where $\bar{w}_{ijsT}$ is the average of the predicted wages that all individuals $k$ in sector $s$ who actually lived in $i$ five years ago would earn in state $i$ in year $t$. $\bar{w}_{ijsT}$ uses the same set of individuals $k$ and is defined as the average of the predicted wages in year $t$ that all workers in sector $s$ coming from state $i$ would have potentially earned in state $j$ (regardless of whether in year $t$ they actually live in $j$ or not).

This aggregation method keeps the composition of the labor force constant across the states, since the same individuals are used for computing the regional

---

18. This approach has become standard in the recent migration literature. For a most recent study see Bertoli et al. (2013). For a detailed description of the methodology see Dahl (2002).

19. We regress individual hourly wages over the standard wage determinants age, age squared, education, gender, ethnic group, and sector dummies plus an individual correction term. The correction terms are the individuals’ migration probabilities as proposed by Dahl (2002). The individual probability of moving from $i$ to $j$ is constructed using only observed personal characteristics (educational attainment, age group, gender, family status, and state of origin). By adding a polynomial of these migration probabilities to the wage equations, we get consistent estimates of the coefficients on the wage determinants. Estimation results of the wage equations are available upon request.
wage at the origin and at the destination. Thus, differences in regional wage levels are only due to variations in the estimated parameters of the wage equations and not to the composition of the labor force.\(^{20}\)

There is however one remaining source of potential reverse causality, which results from the possibility that with more sizeable immigration levels, migrants may exert a negative impact on the local wage level. But so far, studies concerning the impact of migration on wages are not conclusive and indicate either a weak positive or neutral effect.\(^{21}\) Moreover, bilateral flows, compared to total immigration, can be considered of small magnitudes, which justifies the assumption that general equilibrium effects are of second order. Therefore, we are confident that our wage variable is not subject to important endogeneity concerns, even though it is not directly addressing all general equilibrium issues.

In table 3, we use migration rates that are constructed separately for highly educated and low-educated workers. Here, the wage variable takes different values for the different educational groups. \(\Delta w_{ijt} \) is constructed as in equation 13 but takes the average of the predicted wages only for the relevant group of workers.

**Main Results**

*Sector-Specific Market Access and Migration Rates*

In column 1 of table 1, we start by estimating a standard model of migration with a reduced set of fixed effects. Instead, next to sector-specific wage gaps, we also take into account the regional differences in unemployment rates, \(\Delta u_{ijt} \), population size, \(\Delta pop_{ijt} \), and homicide rates, \(\Delta death_{ijt} \).\(^{22}\)

Homicide rates are considered as a proxy for crime and security. For both the unemployment gap and the difference in homicide rates, we expect a negative impact. The expected sign of population is ambiguous. Although there are more

---

\(^{20}\) Note that \(\Delta w_{ijt} \) is constructed using predicted wages in levels and not in log, as do Grogger and Hanson (2011). When repeating our main estimations with the wage variables in log, wages are not significant and market access shows a higher coefficient. Overall, this would not affect our general conclusions on market access. However, given the highly significant results for wages in levels, we believe that wages in this form are the relevant variable for the estimation of the location decision of workers in Brazil.

\(^{21}\) In order to explain these findings, more complex models have been proposed that take into account investment reactions or other adjustment channels to migration (Dustmann et al. 2013; Moretti 2011). Accounting for all of these general equilibrium effects would require a careful treatment of the potential interactions between wages, the housing sector, and investment, among other potential outcomes. Yet, preliminary work by Morten and Oliveira (2014) indicates that these alternative adjustment channels are of little overall importance for Brazil.

\(^{22}\) See appendix S4.3 for the sources of the additional control variables and the construction of the unemployment rate.
jobs available in large states, there are also possible congestion costs. In column 1, all coefficients have the expected sign and are significant, except for the population variable.23

Column 2 contains our preferred specification described in equation 7. Here we include destination-year and origin-year fixed effects to control for time and state varying variables like the price index or the presence of foreign owned firms. Despite the addition of these controls, the magnitude of the coefficient of market access decreases only slightly and remains significant at the 1 percent level. The observed effect here corresponds to the impact that market access has on migration beyond its indirect impact via the wage gap.

This direct effect can be interpreted as the consequence of improved job opportunities generated through several mechanisms. Notably, this direct effect of international demand could be the result of the growth in the number of vacancies, an

23. We do not adjust standard errors for the fact that the market access and wage variables are themselves estimated. Bootstrapping standard errors is prohibitive given the already considerable computational requirements for the construction of each of these variables.
increase in the tightness of the labor market or more “high quality” jobs. Due to
the lack of more detailed data, we cannot identify which is the exact channel, but
all of these would increase the utility of workers in this state and thus attract more
migrants.

In column 3, we repeat our benchmark estimate but exclude the wage variable.
This specification captures the joint effect market access has on migration via the
two possible channels: higher wages and more job opportunities. As expected,
the coefficient of market access is higher and remains highly significant, when
wages are excluded.

Since our empirical specification derives from an aggregate discrete choice
model (grouped logit model), the estimated coefficients cannot be directly inter-
preted as marginal effects. To find the partial effect of a change in a location
characteristic on the migration probability between two states, we need to differ-
entiate equation 4 with respect to the $X_{ij}$ of interest, which can be written as:

$$
\frac{\partial s_{ijst}}{\partial X_{ijst}} = \beta s_{ijst} (1 - s_{ijst})
$$

(14)

To evaluate the importance of the direct effect of market access on domestic mi-
gration, we replace $\beta$ with the estimated coefficient of market access and $s_{ijst}$ with
the observed migration probabilities. Equation 14 then tells us how the probabil-
ity of migrating from state $i$ to any state $j$ in sector $s$ in year $t$ is affected by a
change of 1 percent in the sectoral market access gap.

The values of the elasticities for the 4183 observations in our benchmark spec-
ification (column 2) range from 0.0003 to 0.14, with an average elasticity of
0.012. For an increase of 1 percent in the market access gap, this translates into a
substantial growth of 34 percent to 57 percent in the number of migrants for
each observation. Using the estimates from column 3, which consider the joint
effect via both channels, this increase reaches 44 percent to 74 percent.

The last three columns of table 1 provide robustness checks. Column 4 repli-
cates our benchmark estimation using the Poisson Pseudo-Maximum Likelihood
estimator (PPML) to deal with the high number of zero-migration flows. The co-
efficient of our key variable of interest remains highly significant, confirming the
positive impact of market access on migration rates. The large standard error in

24. Helpman et al. (2010) develop a model that may lead to another possible explanation for our
finding of a significant market access coefficient next to a significant wage gap: When firms do not react to
an increase in market access by opening more positions but with screening more intensively to obtain a
better match, this may attract suitable candidates from other regions. When this mechanism is not fully
capitalized into wages, our market access coefficient could represent the better matching between
employers and employees provoked by deeper trade integration.

25. All main results hold also when using destination-origin-year fixed effect instead of
destination-year, origin-year, and origin-destination dummies (results available upon request). Table S1.3
presents some sensitivity analyses of our benchmark equation on our market access measure, with overall
similar coefficients. Table S2.1 shows that all main results hold for a subsample of workers in
sector-specific occupations only.
column 4 indicates that even though the coefficient is higher than in the previous
estimates, the magnitude is not significantly different from the one in the benchmark equation.26

Columns 5 and 6 address the concern that the positive coefficient on market
access could reflect state j’s comparative advantage in the export supply of a particular sector s if these two are correlated. To make sure that our variable of interest is indeed capturing regional differences in access to foreign markets, column 5 includes sector-destination and sector-origin dummies, which account for sector-region specific characteristics, such as a potential comparative advantage of state i in sector s.27 Our market access coefficient remains comparable to the previous estimates. However, the parameter of the wage gap becomes very small and turns insignificant. This suggests that even though wages vary a lot between sectors and states, the yearly variation within sector-state combinations is relatively low, which makes it difficult to identify the effect of wages on migration in the presence of these additional fixed effects.

In column 6, we add an additional variable (Δsupplyijs ), which captures the
difference between regions in their capacity to supply goods in sector s. This variable is the four years average of the estimated exporter fixed effect for each Brazilian state in sector s from the gravity trade equation (equation 11) and captures the supply capacity of each exporting region. The higher the comparative advantage of a state in sector s, the higher its supply capacity. Even though the coefficient of this variable is positive, we do not want to give it a strong causal interpretation, as this measure is likely to be endogenous to domestic migration.28

A highly significant coefficient of market access also in these last two specifications gives us further confidence that the spatial structure of foreign demand matters and that our results are not driven by any local comparative advantage in a specific industry correlated with our market access variable.

Heterogeneous Impact by Industries

Workers in different industries might react differently to changes in market
access. This could arise, for example, from a different degree of dependence of the industries on foreign demand or different labor market structures across industries affecting the mobility of workers. To test empirically for heterogeneity in the role of market access in the migration pattern, we allow the coefficient of market access to vary across all eleven industries.

26. The data set in column 4 consists of all the 1748 sector-origin-destination combinations for which we observe a positive migration flow for at least one year. The panel is not entirely balanced since we exclude fifty-seven migration rates because i) they are constructed with less than six individual observations; or ii) we do not have wage data for the origin-sector combination.

27. We exclude here the origin-year and destination-year fixed effects to reduce the number of fixed effects. Including all sets of dummies would substantially reduce the variation left to explain.

28. The number of observations in column 6 is reduced since not all Brazilian states have been exporting in all sectors during our sample period. As a consequence, we cannot estimate all the sector-year specific exporter fixed effects for each state.
In column 1 of table 2, all sectors, except Electrical & Electronics, exhibit a positive and significant coefficient. This shows that the positive effect of market access that we found before is not driven by any particular sector. Column 2 also allows the coefficient on the sector-specific wage variable to vary by industry. Although this decreases the magnitudes of the market access coefficients, these estimates confirm the findings of column 1.

Table 2. Market Access Impact by Sector

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>ln(migrants_{ijst}/stayers_{iist})</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High: comparative advantage industries</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta MA_{ijs} \times \text{Agriculture}$</td>
<td></td>
<td>2.949$^a$</td>
<td>$-0.018$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.435)</td>
<td>(0.298)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta MA_{ijs} \times \text{Food}$</td>
<td></td>
<td>1.377$^a$</td>
<td>0.924$^b$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.451)</td>
<td>(0.424)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta MA_{ijs} \times \text{Wood}$</td>
<td></td>
<td>2.334$^a$</td>
<td>1.474$^a$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.433)</td>
<td>(0.383)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta MA_{ijs} \times \text{Plastic &amp; non-metallic}$</td>
<td></td>
<td>0.522$^a$</td>
<td>0.234$^c$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.136)</td>
<td>(0.134)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta MA_{ijs} \times \text{Basic metals}$</td>
<td></td>
<td>1.028$^a$</td>
<td>0.529$^a$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.197)</td>
<td>(0.183)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(\beta_{H}) \Delta MA_{ijs} \times \text{Strong Adv}$</td>
<td></td>
<td></td>
<td></td>
<td>0.810$^a$</td>
<td>0.829$^a$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.150)</td>
<td>(0.152)</td>
</tr>
<tr>
<td><strong>Medium: no comparative advantage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta MA_{ijs} \times \text{Mining}$</td>
<td></td>
<td>1.062$^b$</td>
<td>0.915$^a$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.467)</td>
<td>(0.340)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta MA_{ijs} \times \text{Textiles}$</td>
<td></td>
<td>2.008$^a$</td>
<td>1.564$^a$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.241)</td>
<td>(0.218)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta MA_{ijs} \times \text{Chemical &amp; Pharmaceuticals}$</td>
<td></td>
<td>0.440$^b$</td>
<td>0.263</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.184)</td>
<td>(0.182)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta MA_{ijs} \times \text{Machinery and others}$</td>
<td></td>
<td>0.785$^a$</td>
<td>0.378$^b$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.153)</td>
<td>(0.161)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(\beta_{M}) \Delta MA_{ijs} \times \text{Medium Adv}$</td>
<td></td>
<td></td>
<td></td>
<td>0.570$^a$</td>
<td>0.596$^a$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.118)</td>
<td>(0.118)</td>
</tr>
<tr>
<td><strong>Low: comparative disadvantage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta MA_{ijs} \times \text{Paper &amp; Printing}$</td>
<td></td>
<td>0.658$^a$</td>
<td>0.442$^d$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.119)</td>
<td>(0.113)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta MA_{ijs} \times \text{Electrical &amp; Electronics}$</td>
<td></td>
<td>0.226</td>
<td>$-0.383$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.331)</td>
<td>(0.331)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(\beta_{L}) \Delta MA_{ijs} \times \text{Low Adv}$</td>
<td></td>
<td></td>
<td></td>
<td>0.340$^a$</td>
<td>0.302$^a$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.106)</td>
<td>(0.104)</td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td></td>
<td>4183</td>
<td>4183</td>
<td>4183</td>
<td>4183</td>
</tr>
<tr>
<td><strong>$H_0: \beta_{H} = \beta_{L}$ (p-value)</strong></td>
<td></td>
<td></td>
<td></td>
<td>0.001</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Heteroskedasticity-robust standard errors clustered at the state of origin-year level appear in parentheses. $^a$, $^b$, and $^c$ indicate significance at the 1%, 5% and 10% confidence levels. All regressions include the fixed effects $FE_{ij}$, $FE_{it}$ and $FE_{st}$ & $FE_{jt}$. Columns 1 and 3 restrict the coefficient of sector-specific wage gaps to be the same across all industries. Column 2 and 4 allow the coefficient of the wage gap to vary across industries in the same way as market access. Wage coefficients are not reported for the sake of brevity. They are mostly positive and significant. For details on the industry classification see appendix S4.4.

Sources: Authors’ analysis based on data described in the text.
Magnitudes of the market access coefficient vary substantially, leading to important differences in marginal effects across sectors (from on average 0.005 for Electrical & Electronics to 0.1 for Wood). A first indication for a possible source of such a variation across sectors lies in the sector’s comparative advantage on the world market. After Brazil opened itself to foreign trade, certain sectors started to flourish, whereas others experienced a substantial decline.

The industries in table 2 are categorized into three groups (high, medium, and low) according to their comparative advantage on the world market. Sectors with an international comparative advantage have on average higher and more significant coefficients for market access. Columns 3 and 4 repeat the estimations from the first two columns, but restrict the coefficients so as to be the same for all industries within a group. The t-test in the bottom line of the table rejects the hypothesis of equality between the market access coefficient of the group with comparative advantage and that with a comparative disadvantage.

These results suggest that workers in more international competitive industries are moving to higher market access regions and taking full advantage of the positive economic prospects linked to increased exposure to exports. Our findings can thus help to explain the concentration of certain industries in specific regions.

In contrast, workers in disadvantaged industries seem less sensitive to changes in foreign market access. Since international demand for their goods is generally low, better access to foreign markets will have less additional value for workers in these industries. As a consequence, market access is expected to play a less important role in the location decision of these workers.

**Empirical Results by Sector and Education**

In this section, we distinguish between highly educated and low-educated workers. Figure A-1 displays differences in migrant shares between the two educational groups for each state for the years 1995 and 2003. Over the sample period, highly educated migrants were more likely to move to the South and Northeast, while the Center region has become a more popular destination for low-educated migrants. These differences in the location choices suggest that the utility of migrating to a specific state might vary across educational levels.

We thus investigate whether the observed differences in migration patterns can be explained partly by a heterogeneous impact of sectoral market access, depending on the educational attainment of the individuals.

However, there is no clear theoretical prediction on whether the effect of market access on migration rates should be stronger for highly educated or low-educated workers. On the one hand, a more pronounced reaction of highly qualified workers to a change in market access would be in line with the New Economic Geography model by Redding and Schott (2003). Their model

29. This classification of industries is based on the measure of revealed comparative advantage for Brazilian industries proposed by Muendler (2007) (for details see appendix S4.4.)
predicts that higher market access leads to a higher wage premium for skilled workers. Thus, we could expect that highly educated workers have a stronger incentive to go to states with high market access to benefit from the additional wage premium or a steeper wage gradient in these regions.

On the other hand, numerous theoretical and empirical studies have suggested that highly educated workers are more sensible to certain region-specific amenities. At the same time, highly educated workers might have better access to well-paid jobs. From this perspective, higher wages and career opportunities created by a higher foreign demand could play a minor role in the migration decision of these individuals.

Fally et al. (2010) show that in Brazil, the states with higher foreign market access pay low qualified workers relatively more than highly qualified workers. This finding is in line with traditional trade theory. The Stolper-Samuelson mechanism predicts that in the case of trade liberalization, there should be an increase in the relative returns of the production factor, which is relatively more abundant in the country. Thus, in the case for Brazil, we could expect a strong effect of market access on migration for low-educated workers via the indirect wage channel.

Menezes-Filho and Muendler (2011) and Corseuil et al. (2013) provide a first indication that trade liberalization could also lead to a strong adjustment via the direct channel for low-educated workers. Both studies document for Brazil that higher educational attainment contributes to increased employment durations. Low-educated workers are thus more likely to be laid off and obliged to move for new employment.

To test for a heterogeneous role of foreign demand depending on educational attainment, we adapt equation 7 to allow the coefficient of the independent variables to be different for highly educated and low-educated workers. Our second benchmark specification can then be written as

\[
\ln m_{ijst}^e = \alpha + \beta_H \Delta MA_{ijst} \times High^e + \beta_L \Delta MA_{ijst} \times Low^e \\
+ \beta_3 \Delta w_{ijst}^e \times High^e + \beta_4 \Delta w_{ijst}^e \times Low^e + FE_{st}^e + FE_{ij}^e + FE_{jt}^e + \varepsilon_{ijst}^e
\]

(15)

where \(m_{ijst}^e\) is defined as the number of migrants in sector \(s\) belonging to educational group \(e\) in year \(t\) moving from \(i\) to \(j\) divided by the number of stayers. The dummy \(High\) (\(Low\)) takes the value one when the migration rate is constructed with high (low)-educated workers. The wage gap, \(\Delta w_{ijst}^e\), is calculated using means of predicted wages that vary across states, sectors and skill groups. As before, \(\overline{t}\) indicates that independent variables are constructed as means over the

30. For example, Levy and Wadycki (1974) have shown that in Venezuela educated individuals tend to value amenities much more than low-qualified individuals. More recently, Adamson et al. (2004) find that returns to education for the higher educated workers fall with the population size in US metropolitan areas, which is also consistent with a skill-biased effect of amenities.
years $t-4$ to $t-1$. To take into account that other migration determinants might also vary according to educational attainment, all included fixed effects ($FE_e$) are allowed to differ between the two groups.\textsuperscript{31}

Table 3 reports results on the heterogeneous impact of market access across educational groups. As in table 1, we display first estimation results for a less restrictive specification. Column 1 does not include the state-year fixed effects, but instead the relative population size, the unemployment gap, and the difference in homicide rates. Column 2 contains our second benchmark specification (equation 15) and column 3 excludes the wage variable to obtain the joint effect of

\begin{table}[h]
\centering
\caption{Bilateral Migration by Education}
\begin{tabular}{@{}lcccc@{}}
\toprule
Dependent variable: & $\ln(\text{migrants}_{ijt}^{e}/\text{stayers}_{ijt}^{e})$ & (1) & (2) & (3) & (4) \\
\midrule
$(\beta_1) \Delta MA_{ijt} \times \text{High edu}$ & 0.058 & 0.041 & 0.071 & \text{benchmark II} \\
 & (0.062) & (0.078) & (0.087) & (0.091) \\
$(\beta_L) \Delta MA_{ijt} \times \text{Low edu}$ & 0.871$^a$ & 0.917$^a$ & 1.069$^a$ & 0.890$^a$ \\
 & (0.132) & (0.151) & (0.161) & (0.170) \\
$\Delta w_{ijt}^{e} \times \text{High edu}$ & 0.188$^a$ & 0.233$^a$ & \text{0.220}$^a$ \\
 & (0.025) & (0.028) & (0.030) \\
$\Delta w_{ijt}^{e} \times \text{Low edu}$ & 0.149$^b$ & 0.202$^b$ & 0.179$^c$ \\
 & (0.072) & (0.090) & (0.099) \\
$\Delta \theta_{ijt}^{e} \times \text{High edu}$ & \text{-0.148}$^c$ \\
 & (0.083) \\
$\Delta \theta_{ijt}^{e} \times \text{Low edu}$ & \text{-0.167}$^c$ \\
 & (0.095) \\
$\Delta pop_{ijt} \times \text{High edu}$ & 1.073 \\
 & (0.937) \\
$\Delta pop_{ijt} \times \text{Low edu}$ & \text{-0.184} \\
 & (0.996) \\
$\Delta death_{ijt} \times \text{High edu}$ & \text{-0.169} \\
 & (0.118) \\
$\Delta death_{ijt} \times \text{Low edu}$ & \text{-0.043} \\
 & (0.069) \\
$\Delta supply_{ijt} \times \text{High edu}$ & \text{0.034}$^a$ \\
 & (0.009) \\
$\Delta supply_{ijt} \times \text{Low edu}$ & \text{0.023}$^b$ \\
 & (0.011) \\
$FE_e$ & yes & yes & yes & yes \\
$FE_{ijt}$ & yes & yes & yes & yes \\
$FE_e \times FE_{ijt}$ & yes & yes & yes & yes \\
Observations & 4614 & 4614 & 4614 & \text{4209} \\
$H_0: \beta_{H} = \beta_L$ (p-value) & 0.000 & 0.000 & 0.000 & 0.000 \\
\bottomrule
\end{tabular}
\end{table}

Heteroskedasticity-robust standard errors clustered at the state of origin-year level appear in parentheses. $^a$, $^b$ and $^c$ indicate significance at the 1%, 5% and 10% confidence levels. Sources: Authors’ analysis based on data described in the text.

31. This specification corresponds to splitting the sample between high and low qualified workers. Migration rates of highly educated workers represent 34 percent of our final sample.
market access on migration via both channels. Column 4 adds the gap of the sector-state specific export supply capacity.

In all specifications, the coefficients of the control variables, including wages, are similar across educational groups. However, the coefficient of market access is significant at conventional levels only for low-qualified workers. The t-tests reported at the bottom of the table clearly reject the hypothesis of a uniform impact of market access across educational groups.

In light of our findings, the differences in the observed migration patterns across educational levels can be partly explained by a different sensitivity to foreign market access.

Economic opportunities associated with international trade seem most important for the location choice of low educated individuals. The strong impact of market access for low-qualified workers can be explained by the fact that Brazil is exporting mainly goods that are intensive in unskilled labor. The industries in which Brazil has a high comparative advantage on the world market exhibit a higher share of low-skilled workers. Consequently, an increase in the demand for exported goods signifies a higher demand and more jobs for low-qualified workers. Also when controlling for wage differentials, $\beta_L$ remains highly significant. This indicates that new employment opportunities created by a stronger local export activity are indeed important for the location choice of this group.

For the highly educated workers, market access remains insignificant in all specifications, even if we exclude wages in order to estimate the joint impact of foreign market access via both channels. The interpretation of this result is less straightforward since it might be driven by various forces, as explained above. One possible explanation is that these individuals have, in general, easier access to “high quality” jobs with good working conditions and career prospects. The alternative explanation of a predominant role of amenities with respect to economic considerations is, however, at odds with the fact that highly educated workers are also responsive to wage differentials.

Simulations

Before we conclude, we use the estimated coefficients of column 2 of table 3 to simulate the implied change in each observed migration rate in response to a positive shock in the foreign demand for Brazilian goods. This provides more intuition for our results and allows to identify the regions that are particularly affected by a specific demand shock.

In table 4, we simulate the effects of four different shocks to $\Delta MA_{ist}$. Column 1 reports the average share of immigrants of each state over the sample period. Columns 2 to 5 show for each state how this share would be affected by these different demand shocks.

32. In the PNAD of 1992, the share of high-skilled workers is highest in the two industries with comparative disadvantage. The share of skilled workers was lowest in agriculture and wood, which are in the highest comparative advantage group.
Marginal effects of market access for the highly educated workers being very low, the implied change in the number of migrants is driven by the low-educated workers. Note that the numbers presented in this table correspond to partial equilibrium effects since our simulation rules out any impact of market access on migration going through the indirect effect of wage differentials. Also, the model does not incorporate any potential impacts of migration on housing costs or other congestion costs.33

### Table 4. Effects of Changes in MA

<table>
<thead>
<tr>
<th>State</th>
<th>Immigrants (share in local pop.) (in %)</th>
<th>Positive demand shock in Mercosur EU NAFTA Decrease in internal distance (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) (2) (3) (4) (5)</td>
<td></td>
</tr>
<tr>
<td>North</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rondonia</td>
<td>.8</td>
<td>−6.5 1.7 5 −.2</td>
</tr>
<tr>
<td>Acre</td>
<td>1.4</td>
<td>−.8 .8 .3 1.3</td>
</tr>
<tr>
<td>Amazonas</td>
<td>.7</td>
<td>−1.7 −.6 .6 9</td>
</tr>
<tr>
<td>Roraima</td>
<td>4.2</td>
<td>−1.1 −1.1 1.5 1.2</td>
</tr>
<tr>
<td>Para</td>
<td>1</td>
<td>−1.3 −.6 2.2 −.3</td>
</tr>
<tr>
<td>Amapa</td>
<td>.4</td>
<td>−.6 −1.4 1.8 −.2</td>
</tr>
<tr>
<td>Tocantins</td>
<td>6.7</td>
<td>−1.8 .9 1.2 1</td>
</tr>
<tr>
<td>Northeast</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maranhao</td>
<td>.3</td>
<td>−1.6 .8 1.3 −.2</td>
</tr>
<tr>
<td>Piaui</td>
<td>1.3</td>
<td>−3 1.7 2 .1</td>
</tr>
<tr>
<td>Ceara</td>
<td>1</td>
<td>−3.4 2.9 1.1 −.1</td>
</tr>
<tr>
<td>Rio Grande do N.</td>
<td>1.2</td>
<td>−4.4 2.9 1.8 −.6</td>
</tr>
<tr>
<td>Paraiba</td>
<td>.8</td>
<td>−4.1 3.5 .6 0</td>
</tr>
<tr>
<td>Pernambuco</td>
<td>.6</td>
<td>−3.6 3.1 .4 0</td>
</tr>
<tr>
<td>Alagoas</td>
<td>.4</td>
<td>−1 .8 .1 −.1</td>
</tr>
<tr>
<td>Sergipe</td>
<td>1.5</td>
<td>−3.2 1.9 1.1 .1</td>
</tr>
<tr>
<td>Bahia</td>
<td>.5</td>
<td>−2.7 1.8 .6 −.3</td>
</tr>
<tr>
<td>Southeast</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minas Gerais</td>
<td>.8</td>
<td>−.7 .3 0 −.2</td>
</tr>
<tr>
<td>Espirito Santo</td>
<td>1.2</td>
<td>−.7 .5 −.2 −.2</td>
</tr>
<tr>
<td>Rio de Janeiro</td>
<td>.2</td>
<td>1.3 −.9 −.6 −.4</td>
</tr>
<tr>
<td>Sao Paulo</td>
<td>.6</td>
<td>3.3 −1.9 −1.4 −1</td>
</tr>
<tr>
<td>South</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parana</td>
<td>1.2</td>
<td>5.6 −2.1 −2 6.4</td>
</tr>
<tr>
<td>Santa Catarina</td>
<td>1.1</td>
<td>2.8 −.9 −1.2 −3.9</td>
</tr>
<tr>
<td>Rio Grande do S.</td>
<td>.7</td>
<td>11.4 −5.1 −5.1 −2.2</td>
</tr>
<tr>
<td>Center</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mato Grosso do S.</td>
<td>3.8</td>
<td>−1.9 .4 .4 −.3</td>
</tr>
<tr>
<td>Mato Grosso</td>
<td>3.9</td>
<td>−1.2 .1 −.5 .3</td>
</tr>
<tr>
<td>Goias</td>
<td>1.6</td>
<td>.8 −.5 −1.2 .2</td>
</tr>
<tr>
<td>Distrito Federal</td>
<td>4</td>
<td>1.4 −.8 −1.5 .6</td>
</tr>
</tbody>
</table>

**Sources:** own calculations. Immigrant shares in column 1 are the observed shares in the PNAD, constructed based on the sectors included in our analysis. The changes in the immigration shares in columns 2 to 5 are obtained with help of the estimates of column 2 of table 3. Column 2 to 4 simulate the consequences of an increase by 3% in the market capacity of the corresponding group of countries. Column 5 assumes a decrease in the internal distance by 10%. Authors’ analysis based on data described in the text.

However, the additional effects mentioned here should play only a minor role. Notably, Morten and Oliveira (2014) show that congestion costs associated with housing would be negligible in the case of Brazil.
The first scenario supposes an increase of 3 percent in the demand coming from the Mercosur members (Argentina, Uruguay and Paraguay). This increase in the relative importance of the Mercosur countries affects states differently. Notably the increase in market access of the Southern states which are closer to the Mercosur partners will be higher than for the Northern states. The resulting change in the market access gap between two states impacts directly the bilateral migration rates. When summing over all sending states, sectors, and the two educational groups, we can calculate the total number of additional immigrants a state will receive. Column 2 shows that states in the South will see the most important increase in their share of immigrants, whereas the North and Northeast are relatively less well connected to these markets and will attract less migrants.

Columns 3 and 4 repeat the exercise for an increase in 3 percent of the demand coming from one of the other two main destinations of Brazil’s exports, respectively the European Union and the NAFTA countries.

Results are different here: for these two scenarios it is the Southern states that will see the strongest decrease in the share of immigrants. In contrast, the geographic proximity of the Northern and the Northeastern states to the European Union and NAFTA countries leads to an important increase in their market access and hence in the immigration share of these states.

These changes in migration patterns in response to a change in the access to foreign demand illustrates well how much the spatial structure of the domestic economy is influenced by what happens abroad.

In the last scenario (column 5), we consider a decrease in the internal distance to the next harbor by 10 percent, i.e., a reduction in bilateral trade costs ($\tilde{f}_{ij}$). This improves relatively more foreign market access of inland states.

This last finding also has implications for domestic policies: the reduction in internal distance can also be interpreted as an improved domestic infrastructure that facilitates the access to the sea for the inland states. However, as our results point out, policy makers aiming at regional development need to be aware that due to the country’s integration into the world economy the effects of their measures can be reinforced or opposed by events happening outside of the country.

**Conclusion**

This paper shows that workers move away from states with low market access and prefer states with higher market access. By controlling for region and sector-specific wages, we can identify the direct impact that market access has on the migration decision beyond the wage channel. We further find differences in the sensitivity of migration rates to changes in foreign demand across sectors and educational levels. This heterogeneity can reinforce the industrial specialization of regions and explain differences in migration patterns between groups of workers.

---

34. This positive demand shock is modeled as an increase by 3 percent of the market capacity (the estimated importer fixed effect $FM_{ijt}$ in equation 12) of these countries. An increase of 3 percent corresponds to an increase by one standard deviation of the estimated market capacities.
Our findings highlight the importance of interactions with foreign countries in shaping the internal spatial distribution of the labor force. This aspect is generally excluded from regional migration studies, which rely only on purely domestic migration determinants. This paper employs household survey data, which has the advantage of considering the informal sector that represents over one third of the Brazilian workforce. However, our data doesn’t allow us to identify the main driving force behind the observed direct effect of foreign market access. Linked employer-employee data could be used to study the evolution of the wage profile after migration, a potential improvement of matching between firms and workers, or to assess nonpecuniary aspects of the jobs (e.g., job tenure) and how they are linked to the export activity of a region.

Supplementary Material

A supplemental appendix to this article is available at http://wber.oxfordjournals.org/.

Appendix

Figure A-1. Differences in Migration Patterns Across Educational Groups

Sources: Own calculations. Migrant shares for each state are calculated as migrants from $i$ to $j$ with educational level $e$ over the total number of migrants in the respective educational group.
Table A-1. Estimation Results of the Trade Equation (Equation 11): Averages of Coefficients by Industry

<table>
<thead>
<tr>
<th>Industry</th>
<th>Distance</th>
<th>Border</th>
<th>RTA</th>
<th>WTO</th>
<th>Market capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>−1.134</td>
<td>1.058</td>
<td>0.549</td>
<td>0.417</td>
<td>26.44</td>
</tr>
<tr>
<td></td>
<td>[0.0455]</td>
<td>[0.245]</td>
<td>[0.127]</td>
<td>[0.248]</td>
<td>[1.777]</td>
</tr>
<tr>
<td>Mining</td>
<td>−1.457</td>
<td>1.040</td>
<td>0.317</td>
<td>0.230</td>
<td>27.19</td>
</tr>
<tr>
<td></td>
<td>[0.0393]</td>
<td>[0.0885]</td>
<td>[0.0972]</td>
<td>[0.146]</td>
<td>[1.944]</td>
</tr>
<tr>
<td>Food</td>
<td>−0.739</td>
<td>0.726</td>
<td>0.289</td>
<td>0.183</td>
<td>18.66</td>
</tr>
<tr>
<td></td>
<td>[0.0899]</td>
<td>[0.171]</td>
<td>[0.199]</td>
<td>[0.373]</td>
<td>[2.017]</td>
</tr>
<tr>
<td>Textiles</td>
<td>−1.334</td>
<td>1.078</td>
<td>0.392</td>
<td>0.417</td>
<td>28.82</td>
</tr>
<tr>
<td></td>
<td>[0.0155]</td>
<td>[0.229]</td>
<td>[0.119]</td>
<td>[0.211]</td>
<td>[1.760]</td>
</tr>
<tr>
<td>Wood</td>
<td>−1.401</td>
<td>0.584</td>
<td>0.778</td>
<td>0.312</td>
<td>28.97</td>
</tr>
<tr>
<td></td>
<td>[0.0315]</td>
<td>[0.187]</td>
<td>[0.167]</td>
<td>[0.178]</td>
<td>[2.021]</td>
</tr>
<tr>
<td>Paper &amp; Printing</td>
<td>−1.689</td>
<td>0.735</td>
<td>0.727</td>
<td>0.503</td>
<td>30.74</td>
</tr>
<tr>
<td></td>
<td>[0.0631]</td>
<td>[0.151]</td>
<td>[0.0801]</td>
<td>[0.205]</td>
<td>[1.866]</td>
</tr>
<tr>
<td>Chemical &amp; Pharmaceuticals</td>
<td>−1.520</td>
<td>0.577</td>
<td>0.608</td>
<td>0.346</td>
<td>30.78</td>
</tr>
<tr>
<td></td>
<td>[0.0191]</td>
<td>[0.149]</td>
<td>[0.0944]</td>
<td>[0.119]</td>
<td>[1.861]</td>
</tr>
<tr>
<td>Plastic &amp; non-metallic</td>
<td>−1.616</td>
<td>0.768</td>
<td>0.619</td>
<td>0.560</td>
<td>30.75</td>
</tr>
<tr>
<td></td>
<td>[0.0309]</td>
<td>[0.144]</td>
<td>[0.171]</td>
<td>[0.180]</td>
<td>[1.687]</td>
</tr>
<tr>
<td>Basic metals</td>
<td>−1.572</td>
<td>0.589</td>
<td>0.558</td>
<td>0.375</td>
<td>31.04</td>
</tr>
<tr>
<td></td>
<td>[0.0330]</td>
<td>[0.187]</td>
<td>[0.142]</td>
<td>[0.166]</td>
<td>[2.028]</td>
</tr>
<tr>
<td>Electrical &amp; Electronics</td>
<td>−1.398</td>
<td>0.596</td>
<td>0.580</td>
<td>0.734</td>
<td>30.14</td>
</tr>
<tr>
<td></td>
<td>[0.0287]</td>
<td>[0.200]</td>
<td>[0.186]</td>
<td>[0.212]</td>
<td>[1.914]</td>
</tr>
<tr>
<td>Machinery</td>
<td>−1.430</td>
<td>0.799</td>
<td>0.546</td>
<td>0.660</td>
<td>31.25</td>
</tr>
<tr>
<td></td>
<td>[0.0319]</td>
<td>[0.124]</td>
<td>[0.157]</td>
<td>[0.201]</td>
<td>[1.740]</td>
</tr>
</tbody>
</table>

Equation 11 is run separately for every industry-year combination. This corresponds to 12 regressions for each industry. This table shows averages of coefficients by industry. Standard deviations of the coefficients are indicated in parentheses. 

Sources: Authors’ analysis based on data described in the text.

---

Table A-2. Migration Rates by Sectors

<table>
<thead>
<tr>
<th>Industry</th>
<th>Migration rates Averages over all years</th>
<th>Nb of individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>2.66</td>
<td>16026.50</td>
</tr>
<tr>
<td>Mining</td>
<td>3.37</td>
<td>414.00</td>
</tr>
<tr>
<td>Food</td>
<td>3.54</td>
<td>3402.88</td>
</tr>
<tr>
<td>Textiles</td>
<td>2.80</td>
<td>5376.13</td>
</tr>
<tr>
<td>Wood</td>
<td>4.14</td>
<td>1043.00</td>
</tr>
<tr>
<td>Paper &amp; Printing</td>
<td>2.78</td>
<td>966.13</td>
</tr>
<tr>
<td>Chemical &amp; Pharmaceuticals</td>
<td>3.54</td>
<td>1075.25</td>
</tr>
<tr>
<td>Plastic &amp; non-metallic</td>
<td>3.13</td>
<td>1621.63</td>
</tr>
<tr>
<td>Basic metals</td>
<td>2.87</td>
<td>3414.13</td>
</tr>
<tr>
<td>Electrical &amp; Electronics</td>
<td>3.04</td>
<td>583.50</td>
</tr>
<tr>
<td>Machinery</td>
<td>2.89</td>
<td>1567.50</td>
</tr>
</tbody>
</table>

Sources: Own calculations. Data are from the PNAD (1995–2003).
REFERENCES


mimeo.

———. 2013. “Regional effects of trade reform: What is the correct measure of liberalization?”  

Letouzé, E., M. Purser, F. Rodríguez, and M. Cummins. 2009. “Revisiting the Migration-Development  

of Migration in Venezuela.”  


Planned Capital City.” mimeo.

2001.” mimeo.

———. 2008. “Trade and Workforce Changeover in Brazil.” In Lane Bender, Andersson Shaw,  
and Wachter von, eds., The Analysis of Firms and Employees: Quantitative and Qualitative  

Muriel, B., and C. Terra. 2009. “Sources of Comparative Advantage in Brazil.”  


The Decision to Invest in Child Quality over Quantity: Household Size and Household Investment in Education in Vietnam

Hai-Anh H. Dang and F. Halsey Rogers

During Vietnam’s two decades of rapid economic growth, its fertility rate has fallen sharply at the same time that its educational attainment has risen rapidly—macro trends that are consistent with the hypothesis of a quantity-quality tradeoff in child-rearing. We investigate whether the micro-level evidence supports the hypothesis that Vietnamese parents are in fact making a tradeoff between quantity and “quality” of children. We present private tutoring—a widespread education phenomenon in Vietnam—as a new measure of household investment in children’s quality, combining it with traditional measures of household education investments. To assess the quantity-quality tradeoff, we instrument for family size using the commune distance to the nearest family planning center. Our IV estimation results based on data from the Vietnam Household Living Standards Surveys (VHLSSs) and other sources show that rural families do indeed invest less in the education of school-age children who have larger numbers of siblings. This effect holds for several different indicators of educational investment and is robust to different definitions of family size, identification strategies, and model specifications that control for community characteristics as well as the distance to the city center. Finally, our estimation results suggest that private tutoring may be a better measure of quality-oriented household investments in education than traditional measures like enrollment, which are arguably less nuanced and less household-driven. JEL: I22, I28, J13, O15, O53, P36

Over the past four decades, there has been considerable study of the relationship between household choices on the quantity and quality of children, starting with the seminal studies by Becker (1960) and Becker and Lewis (1973). The

Hai-Anh H. Dang (corresponding author) is an economist with the Poverty and Inequality Unit, Development Research Group, World Bank; his email address is hdang@worldbank.org. F. Halsey Rogers is lead economist with the Global Education Practice, World Bank; his email address is hrogers@worldbank.org. We would like to thank the editor Andrew Foster, three anonymous referees, Mark Bray, Miriam Bruhn, Hanan Jacoby, Shahidur Khandker, Stuti Khemani, David McKenzie, Cem Mete, Cong Pham, Paul Schultz, and colleagues participating in the World Bank’s Hewlett grant research program, and participants at the Population Association of America Meeting for helpful comments on earlier drafts of this paper. We would also like to thank the Hewlett Foundation for its generous support of this research (grant number 2005-6791). A supplemental appendix to this article is available at http://wber.oxfordjournals.org/.
hypothesis driving the literature is that parents make tradeoffs between the number of children they bear and the “quality” of those children, which is shorthand for the amount of investment that parents make in their children’s human capital. If this hypothesis is true, it has considerable implications for policies aimed at increasing economic growth and reducing poverty. For example, this can motivate policy makers to work on policies that assist couples to avoid unwanted births or to subsidize birth control (Schultz 2008).

We investigate a different measure of household investment in their children in this paper, which is private tutoring—or extra classes—in mainstream subjects at schools that children are tested in. Private tutoring is now widespread in many countries, especially but not solely in East Asia, and evidence indicates that it improves students’ academic performance in some countries, including Germany, Israel, Japan, and Vietnam (Dang and Rogers 2008). There has been considerable debate about tutoring among policymakers. One crucial question is whether widespread availability and use of private tutoring exacerbates or helps equalize social and income inequality (Bray 2009; Bray and Lykins 2012), a question that is relevant to both developing and developed countries. Here, the link with demography is important: if use of tutoring is correlated with both smaller family size and higher family income, this heightens the risk that it could exacerbate inequality.

We make several conceptual and empirical contributions in this paper. Our conceptual contribution is to propose private tutoring as a new measure of household investment in their children’s education quality in the context of the child quantity-quality tradeoff literature. Private tutoring may be an especially good measure of a household’s decision to invest voluntarily in children’s human capital—compared with enrollment, for example, which may also reflect exogenous factors such as compulsory schooling laws. Put differently, private tutoring

1. The empirical evidence on the correlation between household size and poverty appears inconclusive. For example, Lanjouw et al. (2004) argue that the common view that larger-sized households are poorer is sensitive to assumptions made about economies of scale in consumption.

2. Private tutoring (or supplementary education) is a widespread phenomenon, found in countries as diverse economically and geographically as Cambodia, the Arab Republic of Egypt, Japan, Kenya, Romania, Singapore, the United States, and the United Kingdom. A recent survey of the prevalence of tutoring in twenty-two developed and developing countries finds that in most of these countries, 25–90 percent of students at various levels of education are receiving or recently received private tutoring, and spending by households on private tutoring even rivals public sector education expenditures in some countries such as the Republic of Korea and Turkey (Dang and Rogers 2008).

3. Other recent studies that find tutoring to have positive on different measures of student academic performance include student test scores and academic performance in India (Banerjee et al. 2010) and the United States (Zimmer et al. 2010); but see Zhang (2013) for recent evidence that tutoring may benefit only certain student groups in China.

4. Given the rapid expansion of educational attainment around the developing world, the tradeoffs that households make between the quantity and quality of children may increasingly manifest themselves outside of the formal education system. For example, in a recent opinion piece in the New York Times on the widening inequality in the United States, the Nobel laureate Joseph Stiglitz (2013) calls for more “summer and extracurricular programs that enrich low-income students’ skills” to help level the playing field between these students and their richer peers.
can capture the household’s extra efforts to increase their children’s human capital. In particular, in countries where the private-school sector is almost non-existent (at least at the pre-tertiary school level) such as Vietnam, private tutoring represents a type of flexible household education investment, which is most likely to be the equivalent of household investment in private education in other contexts. Very few, if any, existing studies offer such study of private tutoring seen in this light.

Furthermore, the existing literature on private tutoring focuses on examining this phenomenon on its own, rather than exploring its intertwined connection with regular school. We attempt to improve on this with an explicit investigation of this nexus. Theoretically, we (slightly) extend the standard Becker-Lewis quantity-quality tradeoff framework to provide further insights that can then guide our empirical analysis; empirically, we propose new measures that exploit both the absolute and relative differences between household investments in regular school and private tutoring. This combined approach thus provides new and original interpretations that appear not to have been attempted elsewhere.

We further make a threefold contribution with our empirical analysis. First, we improve on previous studies by providing the most comprehensive empirical investigation to date of different aspects of household investment in private tutoring for each child (i.e., at the child level). These include participation in tutoring, household monetary investment in tutoring, and time spent both in the short term (i.e., frequency of attending tutoring classes in one year) and in the long term (i.e., number of years attending tutoring classes) on tutoring. We also go one step beyond just looking at household investment in tutoring by considering the situation where households can make a joint decision on whether to enroll their children in school and to send them to tutoring classes.

Second, to identify the impacts of family size on household investment in private tutoring, we use as an instrument the distance from the household’s commune to the nearest family planning center. In contrast to those used in most previous studies, this instrumental variable allows us to study the effects of family size for families with one child or more. Our results provide considerable support for the quantity-quality tradeoff in the Vietnamese context. Furthermore, the IV estimates of the impacts of family size are larger in magnitude than the uninstrumented results. These estimation results hold for several different measures of tutoring and are generally robust to different model specifications, identification strategies, and definitions of family size.

5. In this paper we focus on households’ investment in their children rather than children’s outcomes because doing so may provide a more direct test of the quantity-quality tradeoff hypothesis (see, for example, Caceres-Delpiano (2006) and Rosenzweig and Zhang (2009) for a similar approach). In the context of Vietnam, private tutoring as a new measure of the households’ investment in the quality of their children appears more appropriate than traditional measures (such as education expenditures or private school attainment) for two reasons. First, Vietnam’s education system is mostly public with more or less uniform tuition, and second, the market for private tutoring is well developed, with approximately 42 percent of children age 6–18 attending private tutoring in the past twelve months.
Finally, we explore the hypothesized child quantity-quality tradeoff in the context of rural Vietnam, a country that has undergone rapid change in fertility and educational attainment. The total fertility rate decreased steadily from 6 births per woman in the 1970s to 4 births per woman in the late 1980s and to just under 2 births per woman currently (World Bank 2014). Over the past two decades, the average number of years of schooling for the adult population has increased rapidly, from 4 in 1990 (Barro and Lee 2012) to 6.6 in 1998 and 8.1 in 2010 (VLSS 1998; VHLSS 2010). The Government of Vietnam has paid much attention to family planning and has promulgated policies over the past fifty years encouraging (and in the case of government employees, requiring) families to restrict their number of children to one or two, but to our knowledge, our study is the first to investigate rigorously the quantity-quality tradeoff for this country.

Our estimation results indicate that each additional sibling reduces the rural household’s investments in a child’s schooling as measured through a variety of indicators: it reduces education expenditure and tutoring expenditure by 0.4 and 0.5 standard deviations, respectively; it decreases the child’s probability of being enrolled in tutoring by 32 percentage points; it reduces the child’s enrollment and tutoring index and tutoring attendance frequency by 0.34 and 0.49, respectively; and it cuts the average time spent on tutoring by 74 hours and 1.4 years of tutoring. With regard to the differences between tutoring and regular school, one more sibling reduces by 31 percentage points the probability of attending tutoring (unconditionally on whether the child is enrolled in school or not); reduces by D 243,000 the amount spent on education expenditure net of tutoring expenditure; and reduces by 8 percentage points and 20 percentage points, respectively, the share of tutoring expenditure in education expenditure and the share of years attending tutoring over completed years of schooling.

This paper has five sections. We provide a review of the literature in the next section, followed in section II by the data description and a description of family planning policies and the private tutoring context in Vietnam. Section III presents our theoretical and empirical framework of analysis and the instrumental variable, which is then followed by the estimation results in section IV and the conclusion in section V.

I. Empirical Literature: Testing the Quantity-Quality Tradeoff

Our paper straddles two strands of literature: the more established literature on the quantity-quality tradeoff and a smaller but growing number of studies on private tutoring. We briefly review the most relevant studies in this section.

One central and empirical challenge among the first literature, on the hypothesized quantity-quality tradeoff, is to address the endogeneity of family size

6. Unless otherwise noted, all estimates from the Vietnam Living Standards Surveys (VLSSs) and Vietnam Household Living Standards Surveys (VHLSSs) are authors’ estimates.
convincingly in the data, since unobserved factors can affect both fertility and child human development outcomes. Different instrumental variables have been used and include unplanned (multiple) births (Rosenzweig and Wolpin 1980; Li, Zhang, and Zhu 2008), the gender mix of children combined with parental sex preference (Angrist and Evans 1998; Angrist, Lavy, and Schlosser 2010), and relaxation of government regulation on family size (Qian 2013). Despite these (and other) studies, the existing evidence on the quantity-quality tradeoff appears far from conclusive;7 furthermore, while these identification strategies are useful, they cannot be applied in all contexts.

In the quantity-quality tradeoff framework proposed by Becker and Lewis (1973), a reduction in the costs of maternity care leads to changes in the relative price of quality and quantity of children and in the amount that parents choose to invest in their children. While no studies on the quantity-quality tradeoff appear to have used this insight to construct instruments, several studies in labor economics use variables related to family planning as instruments to identify the causal impacts of family size on female labor supply.8 Instrumenting for fertility with state- and county-level indicators of abortion and family planning facilities and other variables, Klepinger, Lundberg, and Plotnick (1999) find that teenage childbearing has substantial negative effects on women’s human capital and future labor market opportunities in the United States. Another US study by Bailey (2006) employs state-level variations in legislation on access to the contraceptive pill to instrument for fertility, and it also provides strong evidence for the impact of fertility on female labor force participation. More recently, Bloom et al. (2009) instrument for fertility with country-level abortion legislation in a panel of 97 countries over the period 1960–2000; they find that removing legal restrictions on abortion significantly reduces fertility and that a birth reduces a woman’s labor supply by almost two years during her reproductive life.

We follow an identification strategy that is similar in spirit to that literature: we use the availability of family planning services as our instrument, which can reduce the cost of maternity care as well as the cost of controlling the quantity of children in general.9 Specifically, in our test of the quantity-quality tradeoff

7. For example, Angrist, Lavy and Schlosser (2010) find no tradeoff in Israel; Lee (2008) finds a weak tradeoff in Korea that gets stronger with more children. In addition, conflicting results have been found for different countries including Brazil (e.g., Ponczek and Souza (2012) and Marteleto and de Souza (2012)), China (e.g., Li et al. (2008) and Qian (2013)), and Norway (Black, Devereux, and Salvanes (2005) and Black, Devereux, and Salvanes (2010)). See also Steelman et al. (2002) and Schultz (2008) for recent reviews.

8. Another thread of the quantity-quality tradeoff literature estimates the reduced-form impacts of family planning services instead (see, for example, Rosenzweig and Schultz (1985) and Joshi and Schultz (2013)). Recent studies that find that family planning-related variables have important impacts on fertility include DeGraff, Bilsborrow, and Guilkey (1997) for the Philippines, Miller (2010) for Columbia, and Portner, Beegle, and Christiaensen (2011) for Ethiopia.

9. Throughout this paper, we follow the literature by using the term “quality” of children to refer to the amount of human capital invested in them. Needless to say, this should not be taken as a value judgment about their worth as individuals. As noted earlier, however, higher human capital is associated with a host of other desirable development outcomes, at both the individual and societal levels.
hypothesis, we use the distance to the nearest family planning center at the commune level as an instrumental variable for the quantity of children.\(^\text{10}\) Perhaps the greatest advantage of this instrument over other commonly used instruments such as twins and sibling sex composition is that the family-planning instrument allows us to analyze the impacts of family size on all of the children in the household (or the single child, if there is only one), while using either twins or children sex composition restricts analysis to a subset of these children.\(^\text{11}\) We discuss this instrument further in section III.

Turning now to the second strand of literature, on private tutoring, few papers have investigated the correlation between household size and household educational investment in their children through private tutoring. To our knowledge, the exceptions are the two papers on Korea by Lee (2008) and Kang (2011), and the former touches only briefly on tutoring. Both of these papers share the same identification strategy, in that they use the sex of the first-born child as an instrument for family size,\(^\text{12}\) but the former implements this analysis at the household level, while the latter does so at the level of the child. Lee (2008) finds a negative impact of larger family size on household investment in education in general and tutoring in particular, but Kang (2011) finds these negative impacts to be significant only for girls.

II. DATA DESCRIPTION, FAMILY PLANNING AND TUTORING IN VIETNAM

Data Description

In this paper, we analyze data from three rounds (2002, 2006, and 2008) of the Vietnam Household Living Standards Surveys (VHLSSs). The VHLSSs are implemented by Vietnam’s General Statistical Office (GSO) with technical assistance from the World Bank and cover around 9,200 households in approximately

10. Distance to services is often used as an instrument in the literature. For example, distance to college is used to identify the returns to education (Card 1995), distance to the tax registration office is used to identify the impact of tax registration on business profitability (McKenzie and Sakho 2010), and distance to the origins of the virus is used to estimate the response of sexual behavior to HIV prevalence rates in Africa (Oster 2012). Gibson and McKenzie (2007) provide a related review of household surveys’ use of distances measured via global positioning systems (GPS).

11. Using twins as the instrument also requires a much larger estimation sample size; as a result, most previous studies that took this strategy have had to rely on population censuses.

12. The use of the sex of the first-born child as an IV has some limitations. First, it requires the assumption of son preference—which appears to be a weak IV, so that Kang (2011) has to rely on bound analysis to identify bounds of impacts of family size in the case of boys. Second, the assumption of son preference in turn requires the assumption that parents do not abort girls at their first childbirth; if they do, the sex of the first-born child is clearly not valid as an exogenous instrument. This concern is especially relevant to Vietnam, which has one of the highest abortion rates in the world (Henshaw, Singh, and Haas 1999). And finally, this identification approach may only work for families with more than one child; our study makes no such restriction on family size, investigating families with between one and seven children.
3,000 communes across the country in each round. The surveys provide detailed information on household demographics, consumption, and education. The surveys also collect data on community infrastructure and facilities such as distances to schools or family planning facilities. Since 2002, the VHLSSs have been implemented biannually and have collected more data for rotating themes for each survey round; for example, the 2006 round focused on educational activities and tutoring. These surveys are widely used for education analysis by the government and the donor community in Vietnam.

Since only the 2002 round collected data on the distance to family planning for rural communes, we restrict our analysis to rural households in Vietnam. The VHLSSs’ commune sample frame remains almost the same during the period 2002–08, which allows us to match the commune information from the 2002 survey round to most of the households in the 2006 and 2008 survey rounds. However, we focus on the 2006 round of the VHLSSs for the outcome variables, since this round has the most detailed information on household investment in tutoring activities. We also supplement our analysis with data from another nationally representative survey (VHTS) focused on private tutoring that we fielded in 2008, as well as data on teacher qualifications in the community from the primary school census (DFA) database.

Since most children start their first grade at six years old, we restrict our analysis to children who are between six and eighteen years old. To address concerns about grown-up children that have already moved away from home, we consider only children who are living at home and households where the total number of children born of the same mother is equal to the number of children living in the household. We define family size as consisting of children born of the same mother, but we also experiment with a more relaxed definition of family size that

13. A commune in Vietnam is roughly equal to a town and is the third administratively largest level (i.e., below the province and district levels) and higher than the village level. There are approximately 9,100 communes in the country (GSO 2012). The respondents for the community module of the VHLSSs are mostly the (deputy) head of the commune.

14. This matching process is complicated by the fact that there were administrative changes resulting in changes to administrative commune codes between 2002, 2006, and 2008. For around 150 communes, we have to rely on both commune and district names (in addition to province and district codes) for matching. We can match 96 percent of all of the communes in 2002 to those in 2006 and 2008 (i.e., we can match 2,808 communes out of 2,933 communes in 2002).

15. For details on this survey, see Dang and Glewwe (2009). We collaborated on designing the survey with other researchers, including Paul Glewwe (University of Minnesota), Seema Jayachandran (Northwestern University), and Jeffrey Waite (World Bank). The survey was administered by Vietnam’s Government Statistics Office, using funding from the World Bank’s Research Support Budget and the Hewlett Foundation.

16. This database is initiated and maintained by World Bank-supported projects. For a brief description on the history and objectives for the primary school census database, see Attfield and Vu (2013).

17. We also experimented with other age ranges such as ages 10–18 and 12–18. Estimation results (available upon request) are qualitatively very similar and even more statistically significant than those for the age range 6–18.
considers all children living together in the households, as well as other stricter definitions to be discussed later.

**Overview of Family Planning in Vietnam**

Vietnam’s family planning policy dates back to 1961 in the North of Vietnam, but it initially had limited success. Following the unification of Vietnam in 1975, policymakers responded to the faster growth of the population than the economy by setting a goal of lowering population growth rates to less than 2 percent. Subsequently, in 1988 the government adopted a policy restricting families to one to two children, which has largely remained in effect until now. The highlights of this policy include the universal and free provision of contraceptives and abortion services, incentives for families, and strict penalties for families with more than two children. Vietnam’s approach to family planning policy closely follows that of one-child-per-family in China, but it is administered less rigorously (Goodkind 1995). This lack of rigor contributes to our analysis of the quantity-quality tradeoff, in fact, by expanding the range of variation of family size.19

An important administrative landmark for family planning—and one that is quite relevant to the discussion below of our instrument’s validity—was the establishment of the ministry-level National Council of Population and Family Planning (NCPFP) in 1984. By the late 1980s, the NCPFP had established administrative offices and staff down to the commune level to ensure that their activities reached the whole population. Together with the official administrative apparatus, the NCPFP also built up a wide-reaching network of family planning volunteers, both at the village level and in most government agencies, to promote family planning policies.20

**Background on Tutoring in Vietnam**

The current education system in Vietnam has three levels: primary (grades one to five), secondary (grades six to nine for lower secondary sublevel and grades ten to twelve for upper secondary sublevel), and tertiary (post-secondary). Almost all schools in rural Vietnam are public schools and provided by the government. Vietnam has almost achieved universal primary education with 94 percent of Vietnamese children age 15–19 having completed primary education (VHLSS 2006). High-stakes examinations are widely used in the education system for

18. This section is mostly based on GDPFP (2011). See also Vu (1994) for discussion of family planning policies in earlier periods.

19. The family size penalties include fines, restrictions on promotion (or even demotions) for government employees, and denial of urban registration status. We attempted in an earlier draft to use households’ exposure to the two-child-per-family policy as an instrument since the strictness with which it is applied varies with certain characteristics that can be largely exogenous to the family. However, it turned out that the policy was not implemented rigorously enough to make it a viable instrument.

20. In 2007, the NCPFP was merged into the Ministry of Health and renamed the General Department of Population and Family Planning (GDPFP).
performance evaluation, and performance on the exams determines whether students can obtain secondary-school degrees and gain admission to colleges/universities. The strict rationing at the tertiary level results in strong competition among high school students, which helps fuel the demand for private tutoring. Private tutoring is such a major feature of the Vietnamese educational landscape that it is hotly debated, both in the media and during the Minister of Education’s presentations to the National Assembly. Policymakers, educators, and parents fall into two main opinion camps—one arguing that private tutoring worsens educational outcomes and harms children, and the other that tutoring can improve the quality of education. The former group calls for a total ban on private tutoring, while the latter supports the (controlled) development of tutoring.21

Table 1 lists the reasons that students take private tutoring classes, according to data from the VHTS. Tutoring classes are divided into two categories: tutoring classes organized by the student’s own school, and other tutoring classes. Across the two types of tutoring, the most important reason for taking tutoring is to prepare for examinations, which accounts for almost half of all responses (42–47 percent). Other commonly cited reasons given include to catch up with the class (13–14 percent), to acquire better skills for future employment (13 percent), and to pursue a subject that the student enjoys (6–11 percent). Other reasons, such as to get childcare, to compensate for poor-quality lessons in school, or to study subjects not taught in mainstream classes, account for a smaller proportion of all responses (1–6 percent each). The preeminence of exam preparation over other performance on the exams determines whether students can obtain secondary-school degrees and gain admission to colleges/universities. The strict rationing at the tertiary level results in strong competition among high school students, which helps fuel the demand for private tutoring.

Private tutoring is such a major feature of the Vietnamese educational landscape that it is hotly debated, both in the media and during the Minister of Education’s presentations to the National Assembly. Policymakers, educators, and parents fall into two main opinion camps—one arguing that private tutoring worsens educational outcomes and harms children, and the other that tutoring can improve the quality of education. The former group calls for a total ban on private tutoring, while the latter supports the (controlled) development of tutoring.21

Table 1 lists the reasons that students take private tutoring classes, according to data from the VHTS. Tutoring classes are divided into two categories: tutoring classes organized by the student’s own school, and other tutoring classes. Across the two types of tutoring, the most important reason for taking tutoring is to prepare for examinations, which accounts for almost half of all responses (42–47 percent). Other commonly cited reasons given include to catch up with the class (13–14 percent), to acquire better skills for future employment (13 percent), and to pursue a subject that the student enjoys (6–11 percent). Other reasons, such as to get childcare, to compensate for poor-quality lessons in school, or to study subjects not taught in mainstream classes, account for a smaller proportion of all responses (1–6 percent each). The preeminence of exam preparation over other

21. See also Dang (2011, 2013) for more detailed discussions of the private tutoring phenomenon in Vietnam.

<table>
<thead>
<tr>
<th>Reason</th>
<th>Tutoring organized by school</th>
<th>Tutoring not organized by school</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prepare for examinations</td>
<td>47.2%</td>
<td>41.7%</td>
</tr>
<tr>
<td>Do not catch up with the class</td>
<td>12.9%</td>
<td>14.4%</td>
</tr>
<tr>
<td>Acquire skills for future employment</td>
<td>12.2%</td>
<td>12.7%</td>
</tr>
<tr>
<td>Like this subject</td>
<td>6.4%</td>
<td>11.3%</td>
</tr>
<tr>
<td>Parents too busy to take care</td>
<td>2.7%*</td>
<td>1.6%*</td>
</tr>
<tr>
<td>Poor quality lessons in school</td>
<td>2.7%*</td>
<td>6.0%*</td>
</tr>
<tr>
<td>Subjects not taught in mainstream classes</td>
<td>0.5%*</td>
<td>1.5%*</td>
</tr>
<tr>
<td>Others</td>
<td>15.4%</td>
<td>10.9%</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>N</td>
<td>376</td>
<td>301</td>
</tr>
</tbody>
</table>

Note: *Fewer than 20 observations.
Source: Authors’ analysis based on data from Vietnam Household and Tutoring Survey 2007–08.
reasons for taking tutoring classes reflects the importance of examinations in the school system in Vietnam.22

Richer households in Vietnam spend more on tutoring classes than do poorer households, as shown in table 2. Currently about 40 percent (¼ 0.4) of households in Vietnam send their children to private lessons, and the majority of them (90 percent) spend between 1 percent and 5 percent of household expenditure on tutoring classes. The percentage of households with positive expenditures on tutoring classes is only 21 percent in the poorest (1st) consumption quintile but nearly doubles to 38 percent in the next richer quintile (2nd) and hovers around 35 percent in the top three quintiles (3rd to 5th). In terms of actual expenditure, the mean expenditure on tutoring classes by the wealthiest 20 percent of households is fifteen times higher than expenditure by the poorest 20 percent of households. And more expenditure on tutoring is found to increase student grade point average (GPA) ranking in Vietnam, with a larger influence for lower secondary students (Dang 2007, 2008).

Our calculation (not shown) using the 2006 VHLSS shows that the majority of children age 6–18 have at most three siblings, with 10 percent having no sibling, 48 percent having one sibling, 27 percent having two siblings, and 10 percent having three siblings; only five percent of these children have four siblings or more. Table 3 provides a first look at children age 6–18 that are currently enrolled in school that comprise our estimation sample, of whom 42 percent attended private tutoring in the past twelve months. They spent on average

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poorest Quintile 2</td>
<td>Quintile 3</td>
<td>Quintile 4</td>
<td>Richest Quintile</td>
</tr>
<tr>
<td>Average household expenditure on tutoring in 2006 (D'000)</td>
<td>54.2</td>
<td>126.4</td>
<td>222.8</td>
<td>325.0</td>
</tr>
<tr>
<td>Distribution of household with exp. on private tutoring as percent of total expenditure in 2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0%</td>
<td>78.8</td>
<td>61.8</td>
<td>55.1</td>
<td>56.3</td>
</tr>
<tr>
<td>1%–5%</td>
<td>20.0</td>
<td>36.4</td>
<td>41.6</td>
<td>38.7</td>
</tr>
<tr>
<td>5%–10%</td>
<td>1.0*</td>
<td>1.5*</td>
<td>3.0</td>
<td>4.4</td>
</tr>
<tr>
<td>10% or higher</td>
<td>0.1*</td>
<td>0.3*</td>
<td>0.2*</td>
<td>0.6*</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>No. of households</td>
<td>1,278</td>
<td>1,269</td>
<td>1,263</td>
<td>1,290</td>
</tr>
</tbody>
</table>

Note: *Fewer than 20 observations.
Source: Authors’ analysis based on data from Vietnam Household Living Standards Survey 2006.

22. For examining our hypothesis of the quantity-quality tradeoff, we are in fact assuming that sending children to tutoring classes are completely determined by parents. If corrupt teachers force tutoring on their own students beyond parental control (see, e.g., Bray 2009; Jayachandran 2014), household investment in tutoring would not provide valid evidence for this tradeoff. However, the results in table 1 suggest this concern is a minor one in the context of Vietnam.
Table 3. Summary Statistics for Children age 6–18, Vietnam 2006

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs.</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrollment in past 12 months</td>
<td>5012</td>
<td>0.87</td>
<td>0.33</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total education expenditure in past 12 months</td>
<td>4248</td>
<td>583.83</td>
<td>745.71</td>
<td>0</td>
<td>20165</td>
</tr>
<tr>
<td>(D’000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completed years of schooling</td>
<td>5012</td>
<td>5.80</td>
<td>3.25</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Private tutoring attendance in past 12 months</td>
<td>4125</td>
<td>0.42</td>
<td>0.49</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Enrollment and private tutoring attendance in past</td>
<td>5012</td>
<td>1.22</td>
<td>0.65</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>12 months (0 = not enrolled in school,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 = enrolled in school but have no tutoring,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 = enrolled in school and have tutoring)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expenditure on private tutoring in past 12 months</td>
<td>4125</td>
<td>104.15</td>
<td>465.35</td>
<td>0</td>
<td>18000</td>
</tr>
<tr>
<td>(D’000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expenditure on private tutoring in past 12 months</td>
<td>1614</td>
<td>246.59</td>
<td>691.19</td>
<td>6</td>
<td>18000</td>
</tr>
<tr>
<td>for those attending private tutoring (D’000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of hours spent on private tutoring in past</td>
<td>4247</td>
<td>89.06</td>
<td>158.71</td>
<td>0</td>
<td>1728</td>
</tr>
<tr>
<td>12 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of hours spent on private tutoring in past</td>
<td>1624</td>
<td>215.43</td>
<td>183.61</td>
<td>2</td>
<td>1728</td>
</tr>
<tr>
<td>12 months for those attending private tutoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tutoring attendance frequency (0 = no tutoring,</td>
<td>4248</td>
<td>0.65</td>
<td>0.77</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1 = tutoring either during school year or</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>holidays/ break, 2 = tutoring during both school</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year and holidays/ break)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years attending private tutoring to date</td>
<td>4248</td>
<td>1.90</td>
<td>2.58</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Number of siblings age 0–18</td>
<td>4248</td>
<td>1.58</td>
<td>1.04</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Distance to family planning center</td>
<td>4248</td>
<td>8.56</td>
<td>9.78</td>
<td>0</td>
<td>80.5</td>
</tr>
<tr>
<td>Age</td>
<td>4248</td>
<td>11.90</td>
<td>3.20</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>Male</td>
<td>4248</td>
<td>0.50</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Years before last grade in current school level</td>
<td>4248</td>
<td>1.67</td>
<td>1.23</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Secondary school</td>
<td>4248</td>
<td>0.58</td>
<td>0.49</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Mother age</td>
<td>4248</td>
<td>37.38</td>
<td>6.00</td>
<td>21</td>
<td>68</td>
</tr>
<tr>
<td>Female-headed household</td>
<td>4248</td>
<td>0.12</td>
<td>0.32</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Head’s years of schooling</td>
<td>4248</td>
<td>7.36</td>
<td>3.39</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>Ethnic majority group</td>
<td>4248</td>
<td>0.83</td>
<td>0.37</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total household expenditures</td>
<td>4248</td>
<td>19222</td>
<td>10209</td>
<td>2145</td>
<td>175393</td>
</tr>
<tr>
<td>Distance to primary school</td>
<td>4248</td>
<td>0.82</td>
<td>1.25</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Distance to secondary school</td>
<td>4248</td>
<td>2.78</td>
<td>2.81</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>North East and West region</td>
<td>4248</td>
<td>0.16</td>
<td>0.37</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>North Central region</td>
<td>4248</td>
<td>0.19</td>
<td>0.39</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>South Central region</td>
<td>4248</td>
<td>0.09</td>
<td>0.29</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Central Highlands region</td>
<td>4248</td>
<td>0.06</td>
<td>0.24</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>South East region</td>
<td>4248</td>
<td>0.09</td>
<td>0.29</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Mekong River Delta region</td>
<td>4248</td>
<td>0.16</td>
<td>0.37</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: All numbers are weighted using population weights.
Source: Authors’ analysis based on data from Vietnam Household Living Standards Survey 2006.

D 104,150 (equivalent to $US 6)\textsuperscript{23} and eighty-nine hours on these tutoring classes also in the past twelve months, and had attended tutoring for 1.9 years; for those that attended tutoring in the past twelve months, the corresponding

\textsuperscript{23} The exchange rate was D 15,994 for $US 1 in 2006 (World Bank 2014).
expenditure and hours spent on tutoring are D 246,590 and 215 hours. Most tutoring attendees (80 percent) take these classes organized by their school (VHLSS 2006).24 Table 3 also shows that the children in our estimation sample have 1.6 siblings on average, are mostly in secondary school (58 percent), and live an average of 8.6 kilometers away from the nearest family planning center.

III. FRAMEWORK OF ANALYSIS

Family Size, Private Tutoring, and Regular school

We present a simple theoretical model that builds on the standard quantity-quality tradeoff framework (Becker and Lewis 1973) for interpreting the interwoven connection between private tutoring and regular school. We note three main specific features with private tutoring, which provide the underlying assumptions behind our model. First, the existence of private tutoring depends on the mainstream education system and it does not stand alone as an independent educational activity; second, it can offer lessons that are often much more flexible and informal than regular school; and third, compared to the public-subsidized regular school, private tutoring is more costly for the average household.

The household maximizes its utility function U(n, q, y)

\[
\max U(n, q, y)
\]

subject to its budget constraint

\[
y + n(p_u e_u + p_r e_r) = I
\]

where \( n \) is the number of children, \( q \) is their quality, \( y \) is the other (numeraire) good with its price set to 1, \( p_k \) is the price of household investment in (or expenditure on) their children’s quality, for \( k = u \) or \( r \), and \( I \) is household income. A child’s quality is assumed to be equivalent to the total amount of public education \( e_u \) and private tutoring \( e_r \) that the household invests in the child:

\[
q = e_u + e_r
\]

We also assume further that regardless of consumer demand, there is a limit \( \bar{e}_u \) on the capacity of public schools to provide the quality of education desired by the household.26

24. See also table S1.1 in the online appendix for a breakdown of tutoring prevalence and expenditure by urban/rural areas.

25. This supplementary aspect of private tutoring helps explain why it has been referred to as “shadow education” (Bray 2009) or “supplementary education” (Aurini et al. 2013).

26. Particularly in developing countries, the public education system is well known for its rigidity, lack of teacher incentives and accountability, and inadequate infrastructure (see Glewwe and Kremer (2006) for a recent review). In our model, this inelasticity of supply should hold at least in the short run.
Examples of this limit can be the inability of public schools to provide more than, say, the basic reading skills in primary grades or a fixed number of hours of instruction, given short-run constraints on resources and capacities. We then make the standard assumptions that the number of children and the goods are nonnegative—that is, $n \geq 0$, $q \geq 0$, $y \geq 0$. Our model extends the standard quantity-quality framework by introducing household tutoring consumption into the household utility function (1), the budget constraint (2), and the limit on public education consumption. Without these extensions (i.e., with $e_r = 0$ and $e_u \leq \infty$), the standard Becker-Lewis model results.

Assuming the marginal utilities of income ($\lambda_1$) is positive, the Kuhn-Tucker conditions for maximizing the utility function subject to the child quality function, the budget constraint, and the public education constraint yield the following results:

\begin{align}
U_n - \lambda_1 (p_u e_u + p_r e_r) &= 0 \\
U e_u - \lambda_1 n p_u - \lambda_2 &= 0 \\
U e_r - \lambda_1 n p_r &= 0 \\
U_y - \lambda_1 &= 0 \\
I - y - n(p_u e_u + p_r e_r) &= 0 \\
\lambda_2 (\hat{e}_u - e_u) &= 0
\end{align}

Equations (5) to (9) thus yield the same result as under the standard Becker-Lewis model: the shadow prices of the quality of children for either public education ($n p_u$) or private tutoring ($n p_r$) are proportional to the quantity of children; or, put differently, an increase in quality is more expensive if there are more children. Under this standard model, a reduction in quantity-related costs such as contraception costs would increase the shadow prices of quantity relative to quality and other goods, leading to smaller household size and better-quality children.

Furthermore, the different values of the marginal utility of relaxing the public education constraint ($\lambda_2$) offer the following results:

(i) If $\lambda_2 = 0$, then the typical household does not consume the maximum available quality of public education (i.e., $e_u < \hat{e}_u$). However, this case is likely to be the exception rather than the norm, since a Vietnamese child
that is currently in school typically has more than a 40 percent chance of attending private tutoring in the past year (table 3) and around half of these children resort to private tutoring besides their regular classes to better prepare for examinations (table 1).

(ii) If \( \lambda_2 > 0 \), then the household consumes the maximum available quality of public education (i.e., \( e_u = \bar{e}_u \)), which has several important implications. First, to improve the quality of its children, the household’s only option is to invest in tutoring; equivalently, since \( e_u = \bar{e}_u \), private tutoring is the only choice variable for maximizing the household’s utility function.\(^{27}\) Second, when coupled with the standard result of quantity-quality tradeoff, this result leads to household demand for private tutoring that is more elastic to household size than the household’s demand for public education is. The model can thus better capture the tradeoff of household investment in their children’s education. In other words, our model indicates that households would cut down on tutoring consumption and increasingly shift their education expenses to the public subsidies as their family size grows. Finally, since private tutoring is more costly than regular education, relaxing the capacity constraint of public education—for example by providing more teacher time with students—can help reduce the demand for tutoring. This result comes from equation (9) where, given a fixed budget constraint, increasing \( e_u (=\bar{e}_u) \) would ceteris paribus result in a lower value of \( e_r \).

Analogously, for a better and fuller picture on the quantity-quality tradeoff, household investment in private tutoring should be examined together with investment in the regular school.

Figure 1 provides a graphical illustration for a typical household in case (ii) discussed above. The supply of education is represented by the supply curves S\(_1\) (solid line) for public education and S\(_2\) (dashed line) for private tutoring. The gradient of S\(_2\) is flatter than the vertical segment of S\(_1\) but steeper than the upward-sloping segment of S\(_1\); these relationships represent, respectively, the fact that private tutoring can fill in the demand for education where the public education system cannot and that private tutoring is more expensive than public schooling. Since private tutoring is prevalent in Vietnam (as shown with tables 1 to 3), the average household would consume the maximum available quality of public education and also some private tutoring. Household demand for tutoring can be represented by a demand curve that lies higher and to the right of point A and that cuts across both the public education supply S\(_1\) and private tutoring supply S\(_2\).\(^{28}\)

\(^{27}\) This result can generally apply to contexts where the household has no other choice besides public education, and already consumes the maximum available quality of public education. In such cases, household investment in public education would not respond to changes in family size.

\(^{28}\) For case (ii), households consume the maximal available quality of public education (Q\(_1\)), and therefore we do not show the demand curve for public education in Figure 1.
This graphical model helps illustrate our theoretical results. First, other things equal, since public education supply is inelastic after point A, family size would have little or no impact on the household’s consumption of public education; consequently, household investment in private tutoring is a better measure of household quantity-quality tradeoff. Second, compared to a representative household with the demand curve $D_1$, the demand curve $D_2$ represents another household that is assumed to have stronger education preferences, which can be represented by a smaller family size according to our theoretical model. Thus, the household with smaller family size would consume more private tutoring ($Q_2^*$) than the household with larger family size ($Q_1^*$). Finally, focusing on investigating private tutoring on its own rather than examining its intertwined relationship with regular school is equivalent to studying the dashed line $S_2$ in Figure 1 alone without taking into consideration its connection with the solid line $S_1$. This can result in an incomplete—or even potentially misleading—picture of private tutoring.

29. Other factors that shift the demand curve include household income, the price of substitute goods or the number of buyers on the market, or expectations about future returns to education.
These findings offer new interpretations of private tutoring as a new measure of household education investment. We will validate these theoretical predictions empirically in later sections, after first discussing the empirical framework and the instrument.

**Empirical Framework**

Our basic estimation equations are for child $j$, $j = 1,\ldots,J$ in household $I$, $i = 1,\ldots,N$

$$E_{ij} = \alpha + \beta \text{FamSize}_i + \gamma X_{ij} + \epsilon_{ij}$$ (11)

$$\text{FamSize}_i = \delta + \lambda \text{DisFam} + \phi X_{ij} + \eta_{ij},$$ (12)

where, for the first equation, the dependent variable $E_{ij}$ includes household education investment. The traditional measures for $E_{ij}$ include school enrollment, educational expenditure, and completed years of schooling. The new measures include private tutoring attendance, a combined school enrollment/tutoring index (which takes a value of 2 if enrolled in both school and tutoring, 1 if school only, and 0 if neither), frequency of tutoring attendance (which takes a value of 2 if enrolled in tutoring during both school year and holidays, 1 if either school year or holidays, and 0 if neither), expenditure on tutoring, and the number of hours in the past year and the number of years to date spent on tutoring. Of these measures, only tutoring attendance and expenditure appear to have been used in previous studies on tutoring.

If some parents decide to choose fewer children and greater investment in each child, a smaller family size will be strongly correlated with unobserved parental devotion to their children, thus biasing estimates upward; however, the opposite holds if parents decide to choose both more children and greater investment in them at the same time. Thus, estimating equation (11) alone would provide biased estimates of the relationship between family size and household investment. The direction of bias appears to be an empirical issue and depends on parental

30. Some further extensions can be added to our theoretical model. For example, we can generalize by assuming a child endowment component in equation (3) as in Becker and Tomes (1976), or another extension is to assume that, instead of prices being fixed, the price of tutoring is a function of the price of regular school. These extensions, however, do not change the main results. Another extension is to assume that $\epsilon_e$ and $\epsilon_\ell$ are multiplicative up to $\bar{\epsilon}_a$ (the constraint on public education), and are additive beyond this value. This would correspond to private tutoring being complementary up to this value, and being substitute after this value. The latter case, however, appears to be the dominant case in Vietnam as discussed above.

31. For children that are currently in school, completed years of schooling is right-censored since we do not observe the final years of schooling for these children. Thus for such children (and our estimation sample), this variable represents a lower-bound estimate only.

32. For easier interpretation of results and because of the large number of zero observations, in our preferred specification we do not transform variables such as expenditures and hours spent on tutoring to logarithmic scale. Estimation results with the transformed variables are similar, however, and coefficients are slightly more statistically significant.
heterogeneity of preference; the IV model would help remove this bias and uncover the true impacts of family size on household investment. Thus, we jointly estimate equations (11) and (12) in an IV model using the commune-level distance to the nearest family planning center (DisFam) as the instrumental variable.

$X_{ij}$ is a vector of child, household, community and school characteristics that include age, gender, school level, mother’s age, mother’s age squared, gender of the household head, head’s years of schooling, ethnicity, household expenditure, and distances to the nearest primary and secondary schools. A variable indicating the number of years that remains before the last grade in the current school level is also added, since this variable can capture the increasing intensity of tutoring investment as students progress through school (Dang 2007), but this variable is left out in the regression for the enrollment/tutoring index since it applies only to children currently enrolled in school.

For easier interpretation of results, we jointly estimate equations (11) and (12) for all the outcomes above using a 2SLS model, except for expenditure and hours spent on tutoring, where we use an IV-Tobit model instead and subsequently provide separate estimates for the marginal effects since a large number of children have zero values for these variables. Let $E_{ij}$ be the latent variable that represents the household’s potential spending (or hours) on tutoring, the Tobit model for equation (11) has the form

$$E_{ij}^* = \alpha + \beta \text{FamSize}_i + \gamma X_{ij} + \epsilon_{ij},$$

where the relationship of the actual ($E_{ij}$) and latent ($E_{ij}^*$) spending on tutoring is given by $E_{ij} = 0$ if $E_{ij}^* \leq 0$ and $E_{ij} = E_{ij}^*$ if $E_{ij}^* > 0$.

Similarly, we can examine the marginal impacts of family size (or other explanatory variables) on either households’ propensity to spend or households’ actual (observed) spending on tutoring classes. While the former interpretation (shown in table 5) may be more relevant for forecasting the future, the latter (shown in table S1.3 in the online appendix S1, available at http://wber.oxfordjournals.org/) is more commonly used and focuses on household spending at present. For our purposes, we will use the latter interpretation of the marginal effects.

33. There are more missing observations with father’s age so we omit this variable.
34. While the number of years of tutoring can also be fitted in a Tobit model, we prefer to use the OLS model for better interpretation. Estimation results using an IV-Tobit provide very similar results.
35. The marginal impacts for household propensity to spend can be calculated as

$$\frac{\partial E(E_{ij}^*|\text{FamSize}_i, Z_{ij})}{\partial \text{FamSize}_i} = \beta,$$

and the marginal impacts for household actual spending can be calculated as

$$\frac{\partial E(E_{ij}|\text{FamSize}_i, Z_{ij})}{\partial \text{FamSize}_i} = \beta \Phi \left( \frac{\alpha + \beta \text{FamSize}_i + \gamma Z_{ij}}{\sigma} \right),$$

where we also assume $\epsilon_{ij} \sim N(0, \sigma^2)$ as in the OLS models. See, for example, Greene (2012) for more discussion on the marginal effects with the Tobit model.
Distance to Family Planning Center as Instrument

Our instrumental variable for family size is the distance to the nearest family planning center since it meets the exogeneity, relevance, and exclusion restriction conditions. In this section, we consider these three criteria in turn.

A major exogeneity-related concern with using public programs, including placement of family planning centers, as instruments is that these programs may have been established in response to local demand (Rosenzweig and Wolpin 1986). The evidence suggests, however, that such demand response is not an issue in Vietnam, where family planning services were already offered at the commune level and reached virtually the whole population by the late 1980s (Goodkind 1995; GDPFP 2011). While little data exist on the local conditions when family planning centers were set up, it is generally the case with most policy implementation in Vietnam that the central government sets the national policies but it is the local governments that ultimately decide exactly how these policies will be implemented.36

Indeed, the provincial governments were observed to be responsible for all work related to family planning and for mothers and children’s health in general (Vu 1994), which should include the establishment of family planning centers. This is corroborated by an analysis of a survey of local governments’ family planning efforts in fifteen provinces across Vietnam by San et al. (1999), which finds that effort strength is mostly driven by the quality of local governments’ leadership and implementation ability, rather than local conditions such as geographical terrain or the level of economic development.37

Still, some variation of the location (and timing) of family planning center may stem from differences in local governments’ resources: communes with more resources might have been more likely to build a family planning center earlier. We argue, however, that once this channel is controlled for in the regressions (as proxied for by commune infrastructure in several model specifications we examine later), the location of the family planning center is exogenous to each household’s decision on number of children. While it is impossible to test directly for the instrument’s exogeneity, we use a three-pronged approach as an extra precaution to ensure its validity.

First, we use the distance to family planning centers in 2002 to instrument for the impacts of family size on household investments in education four years later, in 2006. This approach can help reduce any contemporaneous correlation between the former and the latter.

Second, in one of the robustness checks, we will restrict our analysis to a subsample of cases in which the family planning centers had already been established

36. Scornet (2001) observes that local governments’ strong autonomy in implementing family planning policies takes its root in the traditional decentralization of monarchical governments in the past. Kaufman et al. (1992) note that the local governments in China—which had a similar although stricter regulation on family size—were similarly responsible for setting up family planning clinics.

37. San et al. (1999) also provide some evidence that their selected 15 provinces share many characteristics of the overall functioning of the national family planning program.
earlier. If family planning centers were more likely to be established first in locations with stronger demand for family planning, older centers would be more effective in reducing family size and would consequently allow households to increase investment in their children’s education. Thus an analysis showing similar impacts of family size for the sample with older centers compared to those for the overall sample would provide evidence for the instrument’s exogeneity.\(^{38}\)

Finally, if it were true that family planning centers were more likely to be first established in locations where households have larger family size, assuming a negative relationship between family size and household investment in their children, we would expect this endogenous placement of these centers to weaken the impacts of the instrument and thus bias estimates upward toward zero. Thus, our estimation results would provide conservative estimates of the extent of the tradeoff.\(^{39}\)

In terms of the relevance criterion for the instrument, our review of the literature from other countries suggests that access to family planning facilities is highly relevant to household decisions on family size. Previous studies for Vietnam using data from the 1997 Demographic and Health Survey offer similar findings that increased access to family planning services increases contraceptive use (Thang and Anh 2002; Thang and Huong 2003) and reduces unintended pregnancy (Le et al. 2004). Our first-stage estimates turn out to show a consistently strong and negative impact of the distance to family planning center on family size.

For the exclusion restriction, there may be concerns that family planning centers directly affect the investment in children by explicitly promoting the idea of a quantity-quality tradeoff. But given the uniform presence in every commune of family planning workers (GDPFP 2011) who can provide interested households with detailed information on the benefits of family planning, family planning centers mostly serve as facilities that provide options for restricting family size to the desired number of children.\(^{40}\) These centers focus on services related to providing contraceptives—such as insertion of intrauterine devices (IUDs),

38. This check does not hold in the opposite direction since older centers may also be effective through other channels that are uncorrelated with endogeneity of location (e.g., longer existence simply increases the chances families know about and use the services at these centers). Larger impacts for family size in the sample of older centers thus would not necessarily indicate violation of exogeneity.

39. An additional concern related to exogeneity is that families could have immigrated to their current commune, meaning that they were not necessarily constrained by the current distance to family planning center when making their decision on giving birth. However, this concern does not apply in our context: we restrict our analysis to rural families only, and fewer than 3 percent of the total population over five years of age move within or to rural areas in Vietnam between 1994 and 1999 (Dang, Tacoli, and Hoang 2003).

40. A reviewer pointed out that family planning centers’ services may also possibly operate through family planning workers/volunteers. However, since these workers were already present in all the communes by 2001 (and most of the communes well before that in the late 1980s), any additional impacts brought about by the new workers that are associated with these centers are likely to be small. This is consistent with Do and Koenig (2007)’s finding that family planning outreach programs (including visits by family planning workers) do not have statistically significant impact on women’s continued use of contraceptive methods. Other programs such as communications campaigns or economic incentives were most often employed by the government through channels (e.g., administrative measures as discussed earlier) that are not typically associated with the activities of family planning centers.
provision of condoms and oral contraceptives, menstrual regulation, and advice on family planning—as well as birth-related medical services and abortions (MOH 2001). In 2002, around one third of the population lived in communes that were within one kilometer of such a center. Thus, access to family planning facilities should affect the educational outcomes of interest only through family size, which satisfies the exclusion restriction.

Another possible objection to the validity of the exclusion restriction is that the distance to the nearest family planning center may be correlated with unobserved commune characteristics that also affect household investment in their children. For example, more remote, less developed communes may also be farther away from any family planning center. In such cases, any negative impacts of household sizes on the outcome variables as instrumented by availability of family planning might be caused by the negative correlation between the general development level of the commune and these outcomes (e.g., poorer communes may spend less on their children’s tutoring classes).

We use two strategies to address this concern. The first is to consider a number of different specifications that test for the strength of this instrument as different commune characteristics are included in the regressions. If the instrument becomes weaker or loses its statistical significance, this means that it is strongly correlated with these commune characteristics (or other unobserved characteristics proxied for by these variables) and concerns about the exclusion restriction are justified. Our second strategy is to use an alternative identification that relies on the heteroskedasticity of the error terms (Lewbel 2012) rather than a regular instrumental variable.41 Heteroskedasticity-based identification has been used for some time (see, e.g., Klein and Vella 2010). In particular, the Lewbel identification approach has been applied in various settings to examine the impacts of body weight on academic performance (Sabia 2007) or the effects of access to domestic and international markets on household consumption (Emran and Hou 2013). Due to its reliance on higher moments, this identification strategy is less reliable than the standard IV approach, but it can provide a qualitative robustness check on our estimation results.

We show estimation results for the first strategy in table 4, which tests for the strength of this instrument using several different specifications sequentially. (Full estimation results are shown in table S1.2 in the online appendix S1.) Model 1, the most basic model, includes only the instrument and the regional dummy variables. Model 2 adds the children’s characteristics and their household characteristics, while model 3 adds to model 2 the distances to the nearest

41. Our standard IV identification strategy comes from the exclusion restriction that the coefficient on the distance to family planning center be zero in equation (1). However, Lewbel (2012) shows that, given the standard regularity condition on the data, we do not need to use this restriction for identification if the error terms are uncorrelated with the right-hand side variables and we can find a variable (or vector of variables) Z that is uncorrelated with the product of the two error terms, that is, cov(Z, εiηi) = 0. In other words, we can use (Z – Z)ηi as the instrument for family size in equation (1), where the distance to family planning center is Z.
primary and secondary schools and includes the variables we use for the subsequent second-stage regressions. Model 4 then adds to model 3 basic commune characteristics such as distances to the nearest paved road, public transportation, and the post office, which are expected to proxy for the general level of economic development of the commune.

Next, to net out any effects that access to community health care has on family size (for example, inadequate health care may reduce family size through high child mortality rates), model 5 adds to model 3 the distance to the nearest health facilities.

Given the low-technology production techniques typically used in agriculture, rural farming households in Vietnam have had to rely for the most part on manpower for their farm work, giving them an incentive to have more children. Furthermore, government employees may be subject to a stricter enforcement of the one-to-two children rule than are farming households. Thus, in the IV terminology (see, e.g., Angrist and Pischke (2009)), farming households may be the population subgroup that is affected differently by the distance to the family planning center than other population subgroups.

### Table 4. Impacts of Distance to Family Planning Center on Number of Siblings Age 6–18, Vietnam 2006 (First-Stage Regressions)

<table>
<thead>
<tr>
<th>Model</th>
<th>Distance to Family Planning Center</th>
<th>Additional Control Variables</th>
<th>Regional Dummy Variables</th>
<th>Individual &amp; Household Characteristics</th>
<th>Distances to School</th>
<th>Community Infrastructure</th>
<th>Distance to Health Facilities</th>
<th>Share of Commune Population Working in Agriculture</th>
<th>R2</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.009***</td>
<td></td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>0.12</td>
<td>6309</td>
</tr>
<tr>
<td></td>
<td>(3.60)</td>
<td></td>
<td>(2.95)</td>
<td>(2.87)</td>
<td>(2.85)</td>
<td>(2.86)</td>
<td>(2.58)</td>
<td>(2.61)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.007***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
<td>0.23</td>
<td>5413</td>
</tr>
<tr>
<td></td>
<td>(2.85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
<td>0.25</td>
<td>4248</td>
</tr>
<tr>
<td></td>
<td>0.007***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
<td>0.23</td>
<td>4178</td>
</tr>
<tr>
<td></td>
<td>(2.85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
<td>0.23</td>
<td>4294</td>
</tr>
<tr>
<td></td>
<td>0.007***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
<td>0.23</td>
<td>4294</td>
</tr>
<tr>
<td></td>
<td>(2.86)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
<td>0.24</td>
<td>4168</td>
</tr>
</tbody>
</table>

Notes: *p < .1, **p < .05, ***p < .01; robust t statistics in parentheses account for clustering at the household level. All regressions control for regional dummy variables, which include the following regions: Northeast and Northwest, North Central, South Central Coast, Central Highlands, South East, and Mekong River Delta. The reference category is the Red River Delta. All household expenditures are in million Vietnamese dong.

Source: Authors’ analysis based on data from Vietnam Household Living Standards Surveys 2002 and 2006.
To address this issue, in model 6 we add to model 3 a variable indicating the share of the commune population working in agriculture. If this addition changes significantly the estimated coefficient on the instrument, this result would suggest that the estimated impact of the distance to the family planning center on family size in model 3 is influenced by the farming-oriented occupation structure in the commune rather than the costs of family planning. Finally, model 7 includes all the variables from models 1 through 6.

The results in table 4 show that the distance to family planning center has a positive and strongly statistically significant impact on family size, as expected.\(^{42}\) Importantly, except in the case of model 1 (which is clearly too simplistic), the magnitude of the estimated coefficient on the distance to the family planning center is almost identical in all the models at around 0.007; this magnitude indicates that a child living 10 kilometers further away from a family planning center would have 0.07 more siblings on average. The consistency of the point estimates suggests both the strong relevance and robustness of this instrument. Since most of the additional variables in models 4 to 6 are statistically insignificant, to keep our models parsimonious, we will use the variables in model 3 in subsequent regressions. In a later section on robustness analysis, we explore different specifications to further assess the validity of this instrument.

IV. Estimation Results

We investigate the impacts of family size on private tutoring alone in the next section, before turning to examining these impacts in the intertwined relationship with regular school.

Impacts of Family Size on Household Education Investment

Table 5 provides the instrumented regressions of the impacts of family size on household education investment; the uninstrumented coefficients on family size are also provided at the bottom of this table for comparison. The instrumented regressions shown in table 5 indicate that a quantity-quality tradeoff exists in Vietnam: all of the instrumented estimated coefficients on family size have a negative sign (as do all the uninstrumented estimated coefficients). While the instrumented coefficients on family size are not statistically significant for school enrollment and completed years of schooling, we can use the point estimates for a rough comparison with the results of previous studies. For example, the ratios for the instrumented coefficient over the uninstrumented coefficient in the regression for these variables (specifications 1 and 3) are around two and fall within a range of corresponding estimates by Li et al. (2008) and Qian (2013) for China;

\(^{42}\) The t-statistics for model 3 are equivalent to an F-statistic of 8.6, which is slightly below the value of 8.96 for a strong IV suggested by Stock and Yogo (2005). Note, however, that Stock and Yogo’s critical values rely on the assumption of independently and identically distributed (iid) errors, whereas our F-statistic is obtained from a cluster-robust regression that is robust to heteroskedastic errors. Without this cluster-robust option, the F-statistic for model 3 is much higher at 22.6.
<table>
<thead>
<tr>
<th>Instrumented Regressions</th>
<th>Spec. 1</th>
<th>Spec. 2</th>
<th>Spec. 3</th>
<th>Spec. 4</th>
<th>Spec. 5</th>
<th>Spec. 6</th>
<th>Spec. 7</th>
<th>Spec. 8</th>
<th>Spec. 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of siblings age 0–18</td>
<td>−0.072</td>
<td>−0.308**</td>
<td>−0.589</td>
<td>−0.318**</td>
<td>−0.337**</td>
<td>−0.488**</td>
<td>−573.957*</td>
<td>−188.425</td>
<td>−1.424**</td>
</tr>
<tr>
<td>Age</td>
<td>−0.033***</td>
<td>0.118**</td>
<td>0.783***</td>
<td>0.010**</td>
<td>−0.026***</td>
<td>0.017**</td>
<td>28.066***</td>
<td>9.061***</td>
<td>0.245***</td>
</tr>
<tr>
<td>Male</td>
<td>−0.038**</td>
<td>−0.084***</td>
<td>−0.236***</td>
<td>−0.085***</td>
<td>−0.124***</td>
<td>−0.138***</td>
<td>−166.057**</td>
<td>−56.664**</td>
<td>−0.365***</td>
</tr>
<tr>
<td>Years before last grade in current school level</td>
<td>0.045***</td>
<td>(0.06)</td>
<td>−0.006</td>
<td>(−0.71)</td>
<td>−0.023*</td>
<td>−6.165</td>
<td>−12.973**</td>
<td>−0.016</td>
<td></td>
</tr>
<tr>
<td>Secondary school</td>
<td>−0.359***</td>
<td>(−7.30)</td>
<td>0.018</td>
<td>(0.61)</td>
<td>0.063</td>
<td>−41.929</td>
<td>−3.793</td>
<td>−0.176</td>
<td></td>
</tr>
<tr>
<td>Mother age</td>
<td>0.048**</td>
<td>0.084</td>
<td>0.334**</td>
<td>0.111**</td>
<td>0.148***</td>
<td>0.148**</td>
<td>203.311*</td>
<td>54.731</td>
<td>0.433*</td>
</tr>
<tr>
<td>Mother age squared</td>
<td>−0.001**</td>
<td>−0.001</td>
<td>−0.004**</td>
<td>−0.001**</td>
<td>−0.002***</td>
<td>−0.002**</td>
<td>−2.563*</td>
<td>−0.680</td>
<td>−0.006**</td>
</tr>
<tr>
<td>Female-headed household</td>
<td>−0.038</td>
<td>−0.018</td>
<td>−0.137</td>
<td>−0.044</td>
<td>−0.076</td>
<td>−0.069</td>
<td>−91.760</td>
<td>−27.225</td>
<td>−0.150</td>
</tr>
<tr>
<td>Head’s years of schooling</td>
<td>0.009*</td>
<td>0.005</td>
<td>0.062**</td>
<td>−0.006</td>
<td>0.007</td>
<td>−0.000</td>
<td>−9.533</td>
<td>0.974</td>
<td>−0.012</td>
</tr>
<tr>
<td>Ethnic majority group</td>
<td>0.010</td>
<td>0.069</td>
<td>0.200</td>
<td>0.091</td>
<td>0.080</td>
<td>0.096</td>
<td>189.625</td>
<td>127.766**</td>
<td>0.218</td>
</tr>
<tr>
<td>Total household expenditures</td>
<td>0.004***</td>
<td>0.016***</td>
<td>0.022***</td>
<td>0.007**</td>
<td>0.010**</td>
<td>0.012***</td>
<td>0.017**</td>
<td>4.112***</td>
<td>0.034***</td>
</tr>
<tr>
<td>Distance to primary school</td>
<td>0.003</td>
<td>0.006</td>
<td>0.055**</td>
<td>0.012</td>
<td>0.009</td>
<td>0.006</td>
<td>26.266</td>
<td>9.945</td>
<td>0.028</td>
</tr>
</tbody>
</table>

**Note:** The table presents the impacts of family size on educational investment for children age 6–18 in Vietnam 2006. The table includes regressions for different educational outcomes such as total education expenditure, completed years of schooling, tutoring attendance, and enrollment & tutoring attendance. The table also includes regressions for various demographic and economic factors, such as the number of siblings, age, gender, years before last grade in current school level, secondary school, mother's age, mother's age squared, female-headed household, head's years of schooling, and ethnic majority group. The table also includes regressions for total household expenditures and distance to primary school, which are important factors in educational investment. The table includes standard errors in parentheses and statistical significance levels (*** p<0.01, ** p<0.05, * p<0.1).
<table>
<thead>
<tr>
<th>Model</th>
<th>F/Chi² test</th>
<th>Log likelihood</th>
<th>N</th>
<th>Number of left-censored obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2SLS</td>
<td>32.49</td>
<td>19019</td>
<td>5012</td>
<td>2511</td>
</tr>
<tr>
<td>2SLS</td>
<td>46.85</td>
<td>18262</td>
<td>4125</td>
<td>4247</td>
</tr>
<tr>
<td>2SLS</td>
<td>862.88</td>
<td></td>
<td>5012</td>
<td>4248</td>
</tr>
<tr>
<td>2SLS</td>
<td>44.50</td>
<td></td>
<td>4125</td>
<td></td>
</tr>
<tr>
<td>2SLS</td>
<td>46.67</td>
<td></td>
<td>4125</td>
<td></td>
</tr>
<tr>
<td>2SLS</td>
<td>47.28</td>
<td></td>
<td>4125</td>
<td></td>
</tr>
<tr>
<td>IV-Tobit</td>
<td>40.31</td>
<td>46.051</td>
<td>4248</td>
<td>2623</td>
</tr>
<tr>
<td>IV-Tobit</td>
<td>501.57</td>
<td></td>
<td>4248</td>
<td></td>
</tr>
<tr>
<td>2SLS</td>
<td>50.25</td>
<td></td>
<td>4248</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** *p<.1, **p<0.05, ***p<0.01; robust t statistics in parentheses account for clustering at the household level. All regressions control for regional dummy variables, which include the following regions: Northeast and Northwest, North Central, South Central Coast, Central Highlands, South East, and Mekong River Delta. The reference category is the Red River Delta. Total household expenditure is net of education expenditure and tutoring expenditure respectively for the specifications of these outcomes. All household expenditures are in million Vietnamese dong, except for the expenditure variable s in the Tutoring specification. For instrumented regressions, the instrumental variable is the distance from the commune to the nearest family planning center.

**Source:** Authors’ analysis based on data from Vietnam Household Living Standards Surveys 2002 and 2006.
the former study finds the instrumented coefficients to range from 0 to 1.5 times the uninstrumented coefficients, but the latter study finds this ratio to be as large as 15 times.

The instrumented coefficients on family size are, however, statistically significant for all the tutoring variables except for tutoring hours. The instrumented coefficients on number of siblings have much larger absolute magnitude than the uninstrumented coefficients, ranging from four (enrollment and tutoring attendance index) to seven times (tutoring expenditure or attendance) as large as their uninstrumented counterparts, which points to the downward bias (in absolute magnitude) of the latter. Thus, both the stronger statistical significance and larger magnitudes for the former are consistent with our earlier theoretical discussion of private tutoring as a more elastic and refined measure of household educational investment than traditional measures.43

Controlling for other characteristics, each additional sibling results in reduced investments in a child’s schooling: reductions in education expenditure and tutoring expenditure respectively by 0.4 standard deviations (or equivalently, a reduction of D 308,246) and 0.5 standard deviations (or D 211,087; see the online appendix S1 table S1.3); a decrease of 32 percentage points in his or her probability of being enrolled in tutoring; and a drop of 0.34 in the child’s enrollment and tutoring index and 0.49 in the tutoring attendance frequency. One more sibling also leads to the child spending seventy-four fewer hours and 1.4 fewer years on tutoring, although the estimated coefficient on tutoring hours is no longer statistically significant.

Estimation results also indicate that, ceteris paribus, older children are less likely to enroll in school but more likely to attend tutoring, while boys are less likely either to enroll in school or attend tutoring.44 Children that are farther

43. Since we control for the commune-level distances to school, the uninstrumented regression results that we presented (at the bottom of table 5) are identical to estimates using an OLS model with commune random effects. As suggested by a reviewer, we also estimate an OLS model with commune fixed effects and between-commune OLS (with variables aggregated at the commune level) for comparison. Estimation results are provided in tables S1.4 and S1.5 in the online appendix, where the former’s estimated coefficients are smaller in magnitudes than the latter’s, which are in turn smaller than those of the IV estimates. This suggests that the between-commune OLS estimates are less biased than the FE estimates, and appears consistent with the bias caused by the endogeneity of family size—which occurs at the household level. In particular, the FE estimates are the commune-fixed effects estimates, which rely on the variation of a small number (at most three) households in a commune for identification. Thus, the FE estimates can be severely biased. On the other hand, the between-commune OLS would first average out this variation (bias) in a commune in constructing the commune-aggregated variables, then rely on the variation between different communes (more than 1500) for identification. Thus, while estimates are still biased, these would be to a lesser extent than those from the FE estimates.

44. We also experiment with using the distance to family center as the instrument for the number of male or female siblings, however, this instrument is statistically significant only in the first-stage regressions for the number of brothers, with qualitatively similar second-stage estimation results (not shown). While this result may indicate a degree of son preference in Vietnam, and it is consistent with previous studies (see, e.g., Phai et al. 1996; Belanger 2002), it may also suggest sex-selective abortion at the same time. Deeper analysis for intra-household gender differences would require better (and more than one) instruments than currently available. Thus, we leave this to further research.
from the last grade in their current school level are, as expected, less likely to have tutoring, but the coefficient on this variable is mostly statistically insignificant except in the case of tutoring hours. Older mothers and richer households invest more in their children’s tutoring, but the quadratic term on mothers’ age is negative, indicating that the marginal effect of age declines and eventually turns negative.

Robustness Checks

We further test the robustness of estimation results and provide them in table S1.6 in the online appendix S1. In the previous section (table 4), we have provided evidence against the concern that distance to the nearest family planning center may be proxying for other important unobserved commune characteristics. However, we test for this possibility again by including as control variables in the equation of interest some commune-level variables such as commune infrastructure, the distance to health facilities, and the share of the commune population working in agriculture. Since our estimation sample is restricted to rural households, to examine the hypothesis—albeit in an indirect way—that urban households spend more on tutoring, we also include the distance from the commune to the nearest major city in Vietnam. Estimation results are largely qualitatively similar.

Our previous study (Dang 2007) shows that communes with higher levels of education spend more on tutoring and argues that this impact can come from both the demand side (e.g., children have peer pressure to study harder or beneficial interaction with well-educated adults) and the supply side (e.g., communities with higher educational levels may be able to supply more tutors). We thus add to our equation of interest either the share of the commune adult population with upper secondary education or higher or a set of commune-averaged variables calculated from the primary school census (DFA) database including the shares of teachers with upper secondary education, upper secondary education plus two more years of additional training, two-year teacher training college education, four-year teacher training college education, and student-teacher ratios. These variables are expected to capture respectively the levels of commune education and the teacher and school quality in the commune. Again, the estimation results are similar to those in our base specification.

45. These cities are Hanoi and Haiphong in northern Vietnam, Danang in central Vietnam, and Cantho and Ho Chi Minh in southern Vietnam. We also experiment with using the distance to the provincial city instead of the distance to these major cities and obtain similar, albeit slightly statistically weaker, results.

46. The only exception is the model specification with all the commune infrastructure and distances variables (row 1), but even in that case, magnitudes are similar but the coefficients have less statistical significance. This is perhaps unsurprising: the model is over-fitted, with all the distance variables statistically insignificant in both the first-stage regressions (as shown in table S1.2 in the online appendix) and second-stage regressions (not shown).

47. Detailed estimation results are provided in tables S1.7 and S1.8 in the online appendix.
While we have reduced some contemporaneous correlation between the distance to the nearest family planning center and household investment in their children by using values for the former in 2002 and the latter in 2006 in our regressions, this gap of four years may not be enough, given that households make their tutoring investments only when children at least six years old. While a family planning center built in 2002 will have had no impact on parents’ decision to give birth to the children who are at least six years old in 2006, the impact of the family planning center on family size in this case will come through the household decision on the number of younger siblings for these children and, subsequently, on total family size. Nevertheless, to examine this case, we restrict our estimation sample to the cases where the family planning center was already operating by 1997, which reduces the estimation sample by more than half. Our results are for the most part qualitatively similar, except that the effects on education and tutoring expenditure now lose their statistical significance (though they keep their negative signs), while the effects on hours and years spent on tutoring become even more statistically significant.

In addition, we also implement other robustness checks including using the Lewbel heteroskedasticity-based IV model, and experiment with dropping out the outliers in the distance to the family planning center. Estimation results are, however, qualitatively similar. More detailed discussion of these results and other checks is provided in the working version of this paper (Dang and Rogers 2013).

**Further/Heterogeneity Analysis**

Estimation results thus far support the negative relationship between family size and household investment in tutoring classes. This subsection delves deeper into this result to provide heterogeneity analysis with, among other factors, different definitions of family size as well as subsets of the population. Estimation results are shown in table 6.

**Different definitions of family size.** Could our estimation results be sensitive to how we define family size? We provide further analysis based on different definitions of family size. First, we restrict the number of siblings to not more than three (row 1, table 6), to test whether the main result is driven by unusually large family sizes. Second, we extend the definition of family size from the children born of the same mother to all the children living in an extended family (row 2), which would perhaps be more consistent with an altruistic model in which

48. As predicted by the Becker-Lewis model, it is total family size that affects the quality-quantity tradeoff. Thus, the distance to the family planning center is still a relevant instrument as long as it can predict total family size.

49. There are a number of missing observations for the year a family planning center was set up, and the distances to school variables are not significant in these specifications, thus we left them out for larger sample sizes and more accurate estimates. As discussed in the previous section, the similarity in impacts of household size for the full sample and the sample with older family planning centers indicates that the locations of family planning centers are effectively independent of household size.
### Table 6. Further/ Heterogeneity Analysis

<table>
<thead>
<tr>
<th>No</th>
<th>Various definitions for family size</th>
<th>Spec. 1</th>
<th>Spec. 2</th>
<th>Spec. 3</th>
<th>Spec. 4</th>
<th>Spec. 5</th>
<th>Spec. 6</th>
<th>Spec. 7</th>
<th>Spec. 8</th>
<th>Spec. 9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total education expenditure</td>
<td>Enrollment</td>
<td>Total education expenditure</td>
<td>Completed years of schooling</td>
<td>Tutoring attendance</td>
<td>Enrollment &amp; Tutoring attendance frequency</td>
<td>Tutoring expenditure</td>
<td>Tutoring hours</td>
<td>Years attending tutoring</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Number of siblings age 0–18 less than or equal to 3</td>
<td>4750</td>
<td>3934</td>
<td>4750</td>
<td>3937</td>
<td>4750</td>
<td>4054</td>
<td>3937</td>
<td>4053</td>
<td>4054</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.110</td>
<td>-520.659</td>
<td>-1.001</td>
<td>-0.630*</td>
<td>-0.609**</td>
<td>-0.876**</td>
<td>-1158.703*</td>
<td>-321.754</td>
<td>-2.450*</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>(-0.89)</td>
<td>(-1.63)</td>
<td>(-1.33)</td>
<td>(-1.88)</td>
<td>(-2.03)</td>
<td>(-2.05)</td>
<td>(-1.75)</td>
<td>(-1.43)</td>
<td>(-1.92)</td>
</tr>
<tr>
<td>2</td>
<td>Number of siblings age 0–18, relaxed definition</td>
<td>7000</td>
<td>5540</td>
<td>7000</td>
<td>5550</td>
<td>7000</td>
<td>5704</td>
<td>5550</td>
<td>5703</td>
<td>5704</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.136</td>
<td>-436.722**</td>
<td>-0.745</td>
<td>-0.474**</td>
<td>-0.541**</td>
<td>-0.767**</td>
<td>-902.347**</td>
<td>-347.998**</td>
<td>-2.283**</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>(-1.37)</td>
<td>(-1.96)</td>
<td>(-1.41)</td>
<td>(-2.13)</td>
<td>(-2.15)</td>
<td>(-2.33)</td>
<td>(-2.03)</td>
<td>(-1.96)</td>
<td>(-2.28)</td>
</tr>
<tr>
<td>3</td>
<td>Number of siblings age 6–18</td>
<td>5015</td>
<td>4125</td>
<td>5015</td>
<td>4128</td>
<td>5015</td>
<td>4251</td>
<td>4128</td>
<td>4250</td>
<td>4251</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.115</td>
<td>-457.807*</td>
<td>-0.914</td>
<td>-0.461**</td>
<td>-0.523**</td>
<td>-0.729**</td>
<td>-846.219*</td>
<td>-283.496</td>
<td>-2.132**</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>(-1.04)</td>
<td>(-1.89)</td>
<td>(-1.41)</td>
<td>(-2.08)</td>
<td>(-2.06)</td>
<td>(-2.27)</td>
<td>(-1.91)</td>
<td>(-1.54)</td>
<td>(-2.22)</td>
</tr>
<tr>
<td>4</td>
<td>Birth order index added to the control variables</td>
<td>3880</td>
<td>3289</td>
<td>3880</td>
<td>3292</td>
<td>3880</td>
<td>3396</td>
<td>3292</td>
<td>3395</td>
<td>3396</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>No</th>
<th>Spec. 1</th>
<th>Spec. 2</th>
<th>Spec. 3</th>
<th>Spec. 4</th>
<th>Spec. 5</th>
<th>Spec. 6</th>
<th>Spec. 7</th>
<th>Spec. 8</th>
<th>Spec. 9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Enrollment</td>
<td>Total education</td>
<td>Completed years of</td>
<td>Tutoring attendance</td>
<td>Enrollment &amp; Tutoring</td>
<td>Tutoring attendance</td>
<td>Tutoring attendance</td>
<td>Tutoring expenditure</td>
<td>Tutoring hours</td>
</tr>
<tr>
<td>5</td>
<td>Estimation sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>being restricted to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>the school considered to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>have good or excellent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>quality by parents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N/A</td>
<td>-177.341</td>
<td>-0.790**</td>
<td>-0.306*</td>
<td>-0.288**</td>
<td>-0.565**</td>
<td>-602.280*</td>
<td>-150.720</td>
<td>-1.293**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-1.17)</td>
<td>(-2.38)</td>
<td>(-1.94)</td>
<td>(-1.97)</td>
<td>(-2.34)</td>
<td>(-1.85)</td>
<td>(-1.23)</td>
<td>(-1.97)</td>
</tr>
<tr>
<td>N</td>
<td>Outcomes in 2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>All outcome variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>in 2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.215*</td>
<td>-413.753</td>
<td>-0.073</td>
<td>-0.519*</td>
<td>-0.576**</td>
<td>N/A</td>
<td>-1222.416**</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>(-1.90)</td>
<td>(-1.13)</td>
<td>(-0.15)</td>
<td>(-1.91)</td>
<td>(-2.28)</td>
<td></td>
<td>(-2.10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>6030</td>
<td>4678</td>
<td>6030</td>
<td>4678</td>
<td>6030</td>
<td>4678</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Model</td>
<td>2SLS</td>
<td>2SLS</td>
<td>2SLS</td>
<td>2SLS</td>
<td>2SLS</td>
<td>2SLS</td>
<td>IV-Tobit</td>
<td>IV-Tobit</td>
</tr>
</tbody>
</table>

**Notes:** *p < .1, **p < 0.05, ***p < 0.01; robust t statistics in parentheses account for clustering at the household level. Unless otherwise noted, each cell provides the estimated coefficient on the number of siblings age 0–18 from a separation regression that controls for the same explanatory variables in the corresponding specification in table 5. All regressions control for regional dummy variables, which include the following regions: Northeast and Northwest, North Central, South Central Coast, Central Highlands, South East, and Mekong River Delta. The reference category is the Red River Delta. Total household expenditure is net of education expenditure and tutoring expenditure respectively for the specifications of these outcomes. All household expenditures are in million Vietnamese dong, except for the expenditure variables in the Tutoring specification. All regressions are estimated with IV method, where the instrumental variable is the distance from the commune to the nearest family planning center.

**Source:** Authors’ analysis based on data from Vietnam Household Living Standards Surveys 2002, 2006, and 2008.
resources are shared within the extended family (e.g., Alger and Weibull 2010; Schwarze and Winkelmann 2011). An altruistic model may be an equally valid model in the context of Vietnam, where Confucian culture remains strong (Huu Ngoc 1996; Tran 2001). Third, we restrict the number of siblings to age 6–18 only (row 3), hypothesizing that the quantity-quality tradeoff will be stronger because households have to invest more in school-age children than in younger ones. Reassuringly, estimates are both larger in magnitude and have slightly stronger statistical significance when we use the more general definition of family size (row 2) and restrict the analysis to school-age siblings (row 3).

**Birth Order.** Beyond the impacts of family size, the birth order of a child can also influence his or her parents’ resource allocation in different directions. For example, first-born children may enjoy more parental time and investment due to their unique timing position (Price 2008; de Haan 2010), but younger siblings may benefit more if parents’ earnings (Parish and Willis 1993) or child-rearing experience increase over the life cycle. Since birth order is closely related with family size (e.g., a child in a higher birth order is more likely to be in a larger family), we construct a birth order index suggested by Booth and Kee (2009) that is purged of family-size effect. This index is defined as $p/((n + 1)/2)$, where $p$ is the child’s birth order, and $n$ the total number of children in the family. We add this birth-order index to our equation of interest (row 4) and find that coefficients become larger (in absolute value) but estimation results are qualitatively similar.50

**Perception of School Quality.** We turn next to the role of school quality in influencing parents to send their children to tutoring lessons. Only a small proportion of households in Vietnam cite poor school quality as the reason for enrolling their children in tutoring classes (table 1), but other studies suggest that the opposite holds in other countries (Kim and Lee 2010; Bray and Lykins 2012). To examine the hypothesis that the negative impacts of family size may possibly not hold for children enrolled in high-quality schools, we restrict our estimation sample to children going to schools perceived by their parents as being of high quality, and we find estimation results for tutoring outcomes to be very similar, except that the impact of household size on education expenditure now loses its statistical significance (row 5).

50. The Pearson correlation coefficient with family size decreases from 0.49 for birth order to $-0.08$ with this index, which indicates that family size effect is largely netted out. We also try another birth order index suggested by Ejrnaes and Portner (2004) but find similar results. Because certain cultures, especially in Asia, may prefer sons over daughters, older sons may be more favored than their younger female siblings. We also try interacting this birth-order index with the male variable, but this interaction variable is not significant either. However, we do not have census data, and the birth order we have is for those children that are currently living in the household only. Thus we do not rule out the possibility that birth orders may have a (weak) impact on our results.
Outcomes for Younger Cohorts in 2008. Recent studies find that the quantity-quality tradeoff holds for younger but not older cohorts in Norway (Black, Devereux, and Salvanes 2010), turns from positive to no effect and then negative during the 1977–2009 period in Brazil (Marteleto and de Souza 2012), and changes from positive for older cohorts to negative for younger cohorts in urban areas in Indonesia (Maralani 2008). To investigate whether this tradeoff applies to younger cohorts in Vietnam, we rerun the same regressions using the 2008 round of the VHLSSs for children in the same age range (6–18). While the 2008 data collect fewer variables on tutoring, our estimation results on the available indicators provide broadly qualitatively similar results (row 6), except that the effect on education expenditure is no longer statistically significant, while the effect on enrollment is statistically significant at the 10 percent level, and the effect on tutoring expenditure becomes stronger both in magnitude and statistical significance.52

Impacts of Family Size on Tutoring Investment Versus Traditional Measures

The regressions in tables 5 and 6 consider measures of household investment in tutoring only using equations (11) and (12). As discussed with our theoretical model, private tutoring should also be examined in its relationship with regular school. To operationalize this hypothesis, we can rewrite equation (11) slightly differently

\[
E_{ijk} = \alpha_k + \beta_k \text{FamSize}_i + \gamma_k X_{ij} + \mu_{ik} + \theta_{ijk},
\]

where \(k\) indexes the different types of household investment in education such as education expenditure or private tutoring expenditure. The error term \(\hat{e}_{ij}\) is broken into two components that vary by household investment type: \(\mu_{ik}\) and \(\theta_{ijk}\), which respectively represent unobserved household effects (e.g., household tastes for their children’s education across different types of education investments) and the child idiosyncratic error term.

If we assume that households have the same preference over investment in regular school and tutoring (i.e., \(\mu_{ik}\) being the same for these two investment

51. As in a previous robustness check regression (table 1.6, row 5), because the distances to school variables are not significant in this specification, we left them out to allow larger sample sizes and greater precision of estimates.

52. Since IV estimates may refer to the unobserved subset of the population that reacts to distance to the family planning center—which is known as the Local Average Treatment Effects (LATE) (see, for example, Imbens and Angrist 1994; Angrist and Pischke 2009)—one concern arises that our previous estimation results may apply only to these households, which may comprise a small share of the total. However, various additional estimation results such as restricting the estimation sample to better-off households in the richer three consumption quintiles and others (see table S1.9 in the online appendix) indicate that a substantial share of the population (i.e., half or more) appears to be influenced by this IV. Restricting the estimation sample to households in the poorest three consumption quintiles provides qualitatively similar but less statistically significant results (see table S1.10 in the online appendix). Also see Dang and Rogers (2013) for further discussion.
types), we can in fact difference out the unobserved household effects by considering the *absolute* difference of these two investments

$$\Delta E_{ij,ln} = \Delta \alpha_{ln} + \Delta \beta_{ln} FamSize_{i,ln} + \Delta \gamma_{ln} X_{ij,ln} + \Delta \delta_{ij,ln}$$  \hspace{1cm} (15a)

or, equivalently,

$$\Delta E_{ij,ln} = \alpha_a + \beta_a FamSize_{i} + \gamma_a X_{ij} + \delta_{a,ij},$$  \hspace{1cm} (15b)

where $\Delta E_{ij,ln} = E_{ijl} - E_{ijn}$, with l and n being, respectively, the investment in regular school and tutoring, and the coefficients in equations (15a) rewritten for convenience of presentation (e.g., $\Delta \beta_{ln} \equiv \beta_a$). Similarly, we can consider the *relative* difference of these two investments

$$\frac{E_{ijl}}{E_{ijn}} = \Delta \alpha_{ln} + \Delta \beta_{ln} FamSize_{i,ln} + \Delta \gamma_{ln} X_{ij,ln} + \Delta \delta_{ij,ln}$$  \hspace{1cm} (16a)

or

$$\frac{E_{ijl}}{E_{ijn}} = \alpha_r + \beta_r FamSize_{i} + \gamma_r X_{ij} + \delta_{r,ij},$$  \hspace{1cm} (16b)

where we have $\frac{\beta_l}{\beta_n} = \Delta \beta_{ln} = \beta_r$ instead.53

Given this assumption of similar household preference over education investment types, we can simply estimate the impacts of family size on the difference between household investment in private tutoring and regular school with OLS method. However, if households have different preferences between tutoring and regular school, the unobserved household effects $\mu_{ik}$ cannot be differenced out and we would need to instrument for family size with the distance to the family planning center in estimating these equations. While it may not seem unreasonable to think that the assumption of similar preference can hold in certain contexts, we believe that this assumption may not hold for the average household in Vietnam given the diverse opinions frequently raised on tutoring in the local media. It thus appears that the uninstrumented regressions would, similar to the results shown in table 5, offer estimates of the impacts of family size that are biased upward toward zero.

Still, for comparison purposes we estimate equations (14) and (15) by both OLS and IV methods and provide estimation results in table 7, where the OLS results are shown at the bottom of this table. For the absolute differences, we

53. We derive equation (16a) by rewriting the dependent variable in equation (1) in log format before taking the ratios of the two investments, and then removing the log format of the ratio of the two investments for easier interpretation. Another way to think about this ratio is tutoring investment standardized by investment in regular schooling.
<table>
<thead>
<tr>
<th>Instrumented Regressions</th>
<th>Spec. 1</th>
<th>Spec. 2</th>
<th>Spec. 3</th>
<th>Spec. 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of siblings age 0–18</td>
<td>-0.311**</td>
<td>-0.243**</td>
<td>-0.077*</td>
<td>-0.203**</td>
</tr>
<tr>
<td></td>
<td>(-2.23)</td>
<td>(-2.02)</td>
<td>(-1.84)</td>
<td>(-2.12)</td>
</tr>
<tr>
<td>Age</td>
<td>0.009*</td>
<td>0.101***</td>
<td>0.002</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(1.74)</td>
<td>(12.49)</td>
<td>(1.06)</td>
<td>(-0.48)</td>
</tr>
<tr>
<td>Male</td>
<td>-0.087***</td>
<td>-0.050*</td>
<td>-0.027***</td>
<td>-0.061***</td>
</tr>
<tr>
<td></td>
<td>(-2.84)</td>
<td>(-1.93)</td>
<td>(-2.87)</td>
<td>(-2.83)</td>
</tr>
<tr>
<td>Years before last grade in current school level</td>
<td>-0.008</td>
<td>0.044***</td>
<td>-0.003</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(-0.95)</td>
<td>(6.74)</td>
<td>(-1.12)</td>
<td>(0.48)</td>
</tr>
<tr>
<td>Secondary school</td>
<td>0.020</td>
<td>-0.315***</td>
<td>-0.012</td>
<td>-0.044**</td>
</tr>
<tr>
<td></td>
<td>(0.70)</td>
<td>(-6.79)</td>
<td>(-1.39)</td>
<td>(-2.30)</td>
</tr>
<tr>
<td>Mother age</td>
<td>0.112**</td>
<td>0.056</td>
<td>0.027*</td>
<td>0.064*</td>
</tr>
<tr>
<td></td>
<td>(2.20)</td>
<td>(1.25)</td>
<td>(1.79)</td>
<td>(1.82)</td>
</tr>
<tr>
<td>Mother age squared</td>
<td>-0.001**</td>
<td>-0.001</td>
<td>-0.000*</td>
<td>-0.001*</td>
</tr>
<tr>
<td></td>
<td>(-2.20)</td>
<td>(-1.28)</td>
<td>(-1.82)</td>
<td>(-1.88)</td>
</tr>
<tr>
<td>Female-headed household</td>
<td>-0.041</td>
<td>-0.026</td>
<td>0.007</td>
<td>-0.027</td>
</tr>
<tr>
<td></td>
<td>(-0.71)</td>
<td>(-0.54)</td>
<td>(0.38)</td>
<td>(-0.69)</td>
</tr>
<tr>
<td>Head’s years of schooling</td>
<td>-0.005</td>
<td>0.001</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(-0.64)</td>
<td>(0.19)</td>
<td>(-0.47)</td>
<td>(-0.24)</td>
</tr>
<tr>
<td>Ethnic majority group</td>
<td>0.085</td>
<td>0.063</td>
<td>0.026</td>
<td>0.073*</td>
</tr>
<tr>
<td></td>
<td>(1.42)</td>
<td>(1.26)</td>
<td>(1.52)</td>
<td>(1.71)</td>
</tr>
<tr>
<td>Total household expenditures</td>
<td>0.007***</td>
<td>0.012***</td>
<td>0.003***</td>
<td>0.005***</td>
</tr>
<tr>
<td></td>
<td>(3.59)</td>
<td>(6.12)</td>
<td>(4.23)</td>
<td>(3.74)</td>
</tr>
<tr>
<td>Distance to primary school</td>
<td>0.010</td>
<td>0.004</td>
<td>0.005*</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(1.02)</td>
<td>(0.45)</td>
<td>(1.81)</td>
<td>(0.85)</td>
</tr>
<tr>
<td>Distance to secondary school</td>
<td>-0.003</td>
<td>0.001</td>
<td>-0.002**</td>
<td>-0.005**</td>
</tr>
<tr>
<td></td>
<td>(-0.71)</td>
<td>(0.21)</td>
<td>(-2.45)</td>
<td>(-2.21)</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.273*</td>
<td>-1.585***</td>
<td>-0.264</td>
<td>-0.445</td>
</tr>
</tbody>
</table>
### Non-Instrumented Regressions

<table>
<thead>
<tr>
<th>Model</th>
<th>F test</th>
<th>N</th>
<th>Mean of dependent variable</th>
<th>Non-Instrumented Regressions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2SLS</td>
<td>2SLS</td>
<td>2SLS</td>
<td></td>
</tr>
<tr>
<td>F test</td>
<td>39.98</td>
<td>45.59</td>
<td>31.68</td>
<td>−0.045***</td>
</tr>
<tr>
<td>N</td>
<td>4248</td>
<td>4125</td>
<td>4091</td>
<td>−0.050***</td>
</tr>
<tr>
<td>Mean of dependent variable</td>
<td>0.41</td>
<td>0.47</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>Non-Instrumented Regressions</td>
<td>−0.014***</td>
<td>−0.034***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Source: Authors’ analysis based on data from Vietnam Household Living Standards Surveys 2002 and 2006. |

Notes: *p<.1, **p<0.05, ***p<0.01; robust t statistics in parentheses account for clustering at the household level. All regressions control for regional dummy variables, which include the following regions: Northeast and Northwest, North Central, South Central Coast, Central Highlands, South East, and Mekong River Delta. The reference category is the Red River Delta. Total household expenditure is net of education expenditure and tutoring expenditure, respectively, for the specifications of these outcomes. All household expenditures are in million Vietnamese dong, except for the expenditure variables in the Tutoring specification. All regressions are estimated with IV method, where the instrumental variable is the distance from the commune to the nearest family planning center.
consider a dummy variable that is 1 if the child attended tutoring in the past twelve months and 0 otherwise (i.e., equivalent to subtracting the school enrollment variable from the enrollment and tutoring attendance variable), and education expenditure net of tutoring expenditure (i.e., equivalent to subtracting tutoring expenditure from total education expenditure). For the relative differences, we consider two share variables: tutoring expenditure over total education expenditure and years of tutoring over completed years of schooling.

The IV estimated coefficients on family size are negative and statistically significant at the 5 percent level for all these variables, except for the share of tutoring expenditure over education expenditure, which is significant at the 10 percent level. These results indicate that one more sibling reduces the probability of attending tutoring (unconditional on whether the child is enrolled in school or not) by 31 percentage points; reduces education expenditure net of tutoring expenditure by $243,000; and reduces the two share variables by 8 percentage points and 20 percentage points, respectively. These estimated coefficients are roughly five or six times larger in absolute magnitude than the uninstrumented regression coefficients.

These estimation results thus validate our theoretical discussion that household demand for tutoring is more elastic to changes in family size than are other traditional measures and that tutoring investment merits more attention as a new measure of household education investment.

V. Conclusion

We find in this paper that families invest less in the education of school-age children who have larger numbers of siblings. Using the distance to the nearest family planning center as the instrument to identify the impacts of family size on household investment, the instrumented number of siblings has a strongly negative effect on education investment, and the estimated coefficient is much larger (in absolute value) than in the original uninstrumented regressions. This effect is robust across different indicators of educational investment—including the general education expenditure on the child, frequency of tutoring attendance, and expenditure and hours spent on tutoring—as well as with different specifications and definitions of family size.

Our results provide evidence that parents in Vietnam are indeed making a child quality-quantity tradeoff. The results suggest further that by lowering the relative cost of child quality and encouraging families to invest in quality, the availability of family planning services has increased investment in education in Vietnam. Finally, the analysis suggests that, compared with traditional indicators like enrollment, data on tutoring may be a more illuminating indicator of parents’ willingness to invest in the quality of education of their children. Indeed, the hypothesized quantity-quality tradeoff appears much more strongly in the tutoring-based measures than in the simple enrollment decision, which may be a coarser indicator of the household’s desire to invest in human capital. These results suggest the need for more research into these quality-oriented measures of
schooling investment, which could be examined in other contexts—besides the quantity-quality tradeoff model—that are broadly related to education efficiency and human capital enrichment.

**REFERENCES**


World Bank. 2014. World Bank Development Indicators Online.


The Impact of Vocational Schooling on Human Capital Development in Developing Countries: Evidence from China

Prashant Loyalka, Xiaoting Huang, Linxiu Zhang, Jianguo Wei, Hongmei Yi, Yingquan Song, Yaojiang Shi, and James Chu

A number of developing countries are currently promoting vocational education and training (VET) as a way to build human capital and strengthen economic growth. The primary aim of this study is to understand whether VET at the high school level contributes to human capital development in one of those countries—China. To fulfill this aim, we draw on longitudinal data on more than 10,000 students in vocational high school (in the most popular major, computing) and academic high school from two provinces of China. First, estimates from instrumental variables and matching analyses show that attending vocational high school (relative to academic high school) substantially reduces math skills and does not improve computing skills. Second, heterogeneous effect estimates also show that attending vocational high school increases dropout, especially among disadvantaged (low-income or low-ability) students. Third, we use vertically scaled (equated) baseline and follow-up test scores to measure gains in math and computing skills among the students. We find that students who attend vocational high school experience absolute reductions in math skills. Taken together, our findings suggest that the rapid expansion of vocational schooling as a substitute for academic schooling can have detrimental consequences for building human capital in developing countries such as China. JEL codes: I25, J24, O15

Prashant Loyalka is a research fellow at Freeman Spogli Institute for International Studies and a faculty member of the Rural Education Action Program (REAP) at Stanford University; his email address is loyalka@stanford.edu. Xiaoting Huang is an associate professor at China Institute for Educational Finance Research (CIEFR), Peking University, China; her email address is xthuang@ciefr.pku.edu.cn. Linxiu Zhang is professor and deputy director of Center for Chinese Agricultural Policy (CCAP), Institute for Geographical Sciences and Natural Resources Research (IGSNRR), Chinese Academy of Sciences (CAS), China; her email address is lxzhang.ccap@igsnrr.ac.cn. Jianguo Wei is an associate professor at CIEFR, Peking University, China; his email address is jgwei@ciefr.pku.edu.cn. Hongmei Yi (corresponding author) is an associate professor of CCAP, IGSNRR, CAS, China; her email address is yihm.ccap@igsnrr.ac.cn. Yingquan Song is an associate professor at CIEFR, Peking University, China; his email address is yqsong@ciefr.pku.edu.cn. Yaojiang Shi is professor and director of Center for Experimental Economics of Education, Shaanxi Normal University, China. James Chu is project manager of REAP at Stanford University, his email address is jchu1225@stanford.com. The authors gratefully acknowledge the financial assistance of the National Natural Science Foundation of China (No. 71110107028 and 71573246), 3ie, and the Ford Foundation. A supplemental appendix to this article is available at http://wber.oxfordjournals.org/.
As the economies of developing countries shift from lower value-added to higher value-added industries and experience technological change, their need for human capital also increases (Heckman and Yi 2012). Higher value-added jobs must be staffed with employees who are equipped with greater skills (Bresnahan et al. 2002). Without a labor force with sufficient skills, developing economies could ultimately stagnate (Hanushek and Woessman 2012).

A number of developing countries identify vocational education and training (VET) as a key approach to building human capital. For example, the promotion of VET at the high school level (or “vocational high school”) has become a policy priority among emerging economies such as Brazil, Indonesia, and China (Newhouse and Suryadarma 2011; National Congress of Brazil 2011; China State Council 2010). Over the past decade, these countries have increased funding and enrollments in vocational high school (often in lieu of academic high school—e.g., Indonesia, see Newhouse and Suryadarma 2011). The rationale underlying these policies is that increases in the proportion of vocational—as opposed to academic—high school enrollments can more effectively build human capital.

For VET to successfully build human capital in these countries, however, it must meet two prerequisites. The first prerequisite is that VET must help students learn specific (vocational) skills that can either directly be used in the labor market after graduation or serve as a foundation for vocational college (Kuczera et al. 2008). Second, VET must help students acquire general skills (e.g., in math, reading, and/or science—Chiswick, Lee, and Miller 2003). The international literature shows that a solid foundation of general skills has a significant and long-term impact on the wages of high school graduates (Levy and Murnane 2004). Research also suggests that job stability for individuals (as well as economic stability for countries) requires lifelong learning, which is contingent on a foundation in general skills (Kezdi 2006). For these reasons, almost all countries require vocational high schools to teach general skills (Kuczera et al. 2008).

Surprisingly, there is little evidence from developing countries as to whether vocational high school helps students acquire specific and general skills, especially in comparison to academic high school. Cross-national studies based on international tests such as the PISA show that students in vocational high school have lower levels of general skills than students in academic high school (by almost half a standard deviation, see Altinok 2011). However, since the PISA data do not contain detailed information on student background characteristics (such as prior test scores) that are necessary to adjust for selection bias, the PISA data are not suitable for measuring the causal impacts of attending vocational versus academic high school. Furthermore, because the PISA data are cross-sectional and not longitudinal, they cannot show how much vocational high school contributes to gains in student learning.

One exception uses longitudinal data from Indonesia in the 1990s to show that attending vocational school has little impact on students’ general skills (Chen 2009). Unfortunately, the Chen study relies on a sample of students
smaller than 1,000. Because this sample does not have sufficient vocational and academic high school students that share a common set of characteristics, the OLS regressions used in the study may give biased results (as they are based on linear extrapolations away from a common support—King and Zeng 2006).

In this paper, we examine whether vocational high school students are learning specific and/or general skills. We seek to accomplish three goals. First, we seek to assess the impact of attending vocational versus academic high school on the dropout rates, math, and computing skills of the average student. Second, we seek to estimate the heterogeneous impacts of attending vocational versus academic high school on the dropout rates and skill levels of disadvantaged (low-income or low-ability) students. Third, we aim to establish whether vocational high school leads to any absolute gains in math and computing skills.

To accomplish these aims, we conduct analyses using longitudinal data on more than 10,000 students in China. Estimates from instrumental variables and matching analyses show that attending the most popular major in Chinese vocational schools (computers) relative to attending academic high school substantially reduces math skills without improving computing skills. Attending vocational high school also increases dropout, especially among disadvantaged (low-income and low-ability) students. We also use comparable (equated or scaled) baseline and follow-up test scores to measure students’ absolute gains in math and computing skills. We find that computing major students who attend vocational high school experience absolute reductions in math. Taken together, our findings indicate that the promotion of vocational schooling as a substitute for academic schooling may be detrimental to building human capital in developing countries such as China.

I. Background

Like many other developing countries, policymakers in China have a strong interest in using VET to build human capital and drive economic growth (China State Council 2010). This interest has resulted in the expansion of vocational high school enrollments from 11.7 to 22.1 million students between 2001 and 2011 and annual investments of more than 21 billion dollars (NBS various years; MOF and NBS 2011). Policymakers in China also hope to use VET to help disadvantaged (low-income or low-ability) students gain employment (China State Council 2010). It is for this reason that policymakers have provided financial aid to all vocational high school students and waived tuition for low-income vocational high school students in particular (China State Council 2010; MOF and MOE 2006).

What are vocational high schools supposed to accomplish? Vocational high school students are trained to become mid-level skilled workers. By policy design, the computer major in China is set up to train workers for entry level jobs in database management, website administration, software engineering, advertising (layout, photo-editing), or computer animation (Chinese Ministry of
This differs from academic high school, which trains students in academic or general skills, mainly for entry into higher education.

In terms of curriculum, vocational high school students in the first year of the computing major are supposed to spend roughly equal amounts of time on academic and computing skills. In their second year, students spend the majority of their time on computing skills. Students spend the third year in internships.

Academic high schools, by contrast, are focused on academic subjects tested on the college entrance examination, with roughly only 10% of time spent on subjects like music, computers, or physical exercise.

How do students choose to attend vocational high school? After graduating from junior high, students decide between entering the labor market, vocational high school, or academic high school. In China, the level of a student’s high school entrance examination (HSEE) score is the primary determinant for entry into academic high school. Every county ostensibly has a cutoff for whether a student’s score makes him/her eligible to enter academic high school (based on the number of positions in academic high school available that year). Those who test below the cutoff are unable to attend academic high school and must choose whether to enter the labor market or vocational high school. Those students that test just above the cutoff sometimes waver between whether to attend vocational or academic high school. Those who test far above the cutoff almost always attend academic high school.

II. RESEARCH DESIGN

2.1 Sampling

This paper draws on longitudinal survey data collected by the authors in October 2011 and May 2012. The sample for the longitudinal survey was chosen in several steps and covers vocational and academic high schools in different regions of China. First, we sampled two provinces in China: Shaanxi and Zhejiang. Shaanxi province is an inland province in Northwest China and ranks fifteenth out of thirty-one provinces in terms of GDP per capita (NBS 2012). Zhejiang is a coastal province that ranks fifth in terms of GDP per capita (NBS 2012). After selecting the two provinces, we sampled the most populous prefectures within each province (three in Shaanxi and four in Zhejiang) and all the counties in those prefectures. In sum, we sampled two provinces, seven prefectures in those provinces, and the seventy-five counties in those prefectures.

1. All VET curricula are based on national standards (such as those from the Ministry of Education, which publishes a detailed list of standards for all facilities, textbooks, and software required for instruction—Chinese Ministry of Education 2008).

2. In contemporary China each province is split into multiple prefectures, which are in turn split into counties.

3. Note that the seven prefectures in our sample contain seventy-five counties. Since not every one of these counties had a vocational school or a vocational school with the computing major, the schools in our sample were distributed in thirty-five of the seventy-five total counties.
We next sampled vocational high schools from the seven prefectures. According to administrative records, there were 204 and 285 vocational schools in the sample prefectures in Shaanxi and Zhejiang, respectively. Using administrative records, we included all vocational high schools that offered a computer major in our sample. We focused on the computer major for two reasons. First, computing is studied in academic high schools (albeit to a lesser degree), which allows us to compare learning gains in specific skills (i.e., computers) across vocational and academic high schools. Second, the computer major is the major with the largest number of enrollments in the two provinces. Over half of all vocational high schools had computing majors, and we only had to exclude 101 schools in Shaanxi and 133 schools in Zhejiang due to the fact that they did not offer computer majors.

After selecting vocational high schools with computing majors, we called these schools to ask how many new (grade ten) students enrolled in autumn 2011. Schools that reported fewer than fifty grade ten students enrolled in the computer major were excluded from our sampling frame. This criterion meant that we excluded fifty-six schools in Shaanxi and seventy-eight schools in Zhejiang. Although the number of excluded schools was higher than we expected, these small schools comprised less than 15% of the share of computing students in Shaanxi and Zhejiang. We then enrolled the remaining forty-six schools in Shaanxi and fifty-five schools in Zhejiang in our sample.

We concurrently sampled academic high schools in the seven prefectures. We found 104 and 155 academic high schools in the sample prefectures in Shaanxi and Zhejiang, respectively. Because we planned to match vocational and academic high school students, we needed a sample of academic high school students that might have considerable overlap in basic student characteristics across the two types of high schools. To achieve this goal, we excluded elite academic high schools from our sample. In China, elite academic high schools select students of much higher ability than nonelite academic high schools. Few (if any) students that are eligible for elite academic high schools would ever consider going to vocational high school. Because students currently enrolled in nonelite academic high school were more likely to have considered attending vocational high schools, we only sampled nonelite academic high schools.

Given these criteria for academic high school, we then selected our sample. Within the seven prefectures, there were sixty-two and eighty-eight nonelite academic high schools in Shaanxi and Zhejiang (about 60% of all academic high schools). From these schools, we randomly sampled fifteen eligible nonelite academic high schools from each province (thirty schools in total).

The next step was to choose which students would be surveyed within the sample schools. In each vocational high school, we randomly sampled two first-year computer major classes (one class if the school only had one computer major class) and surveyed all students in these classes. In each nonelite academic

---

4. We excluded these small schools because policymakers informed us that such schools were at high risk of being closed or merged during the school year.
high school, we randomly sampled two first-year classes and surveyed all students in these classes.

2.2 Data Collection

Our data collection started with a baseline (October 2011) survey. The baseline survey collected data from students, students’ homeroom teachers, and school principals. Among vocational high schools, 7,114 first-year students in 184 classes filled out the baseline survey. Among academic high schools, 2,957 students in fifty-nine classes filled out the baseline survey. In total, we surveyed 10,071 students (7114 + 2957).

We followed up with the sample vocational and academic high school students in May 2012 (hereafter known as the endline survey). The survey forms used in the endline survey were similar to those used in the baseline survey. Most importantly, our data allowed us to create three primary outcome variables: (a) student dropout (whether a student was enrolled in a high school as of May 2012); (b) student gains in computing skills (according to a standardized exam); and (c) student gains in math skills (according to a standardized exam).

Our first outcome was whether a student (who had started high school in October 2011) had dropped out by May 2012. To identify dropouts, our enumerators filled in a student-tracking form for each class during the endline survey. This form contained a list of all the students who completed our baseline survey. Our enumerators marked each student on the baseline list as present, absent, transferred, on leave, or dropped out, according to information provided by class monitors. Moreover, after the field survey was over, our enumerators called the parents or guardians of the students to further ascertain whether the students marked as dropped out on our tracking form had in fact dropped out.

A multi-step procedure was used to ensure that the computing and math tests were valid (and represented the types of skills that students were expected to acquire in high schools in China). First, we collected a pool of over 200 computer and math exam items (questions) from official sources. Because the test items were based on national standards, they are (according to policy) supposed to be reflected in the content actually taught in school. Second, to further verify the content validity of the items, we asked vocational high school teachers to ensure that the items were relevant to what computer majors would actually be learning in vocational high school. Third, after piloting the large pool of exam items with more than 300 students, we designed vertically scaled (equated) baseline and endline exams using item response theory (IRT). By using the IRT procedure suggested by Kolen and Brennan (2004), we were able to ensure that baseline and

5. Because of low enrollments, there was one academic high school (out of the 30) that only had one class (instead of 2). This explains why there are fifty-nine classes as opposed to sixty.

6. Specifically, the computer exam items were taken from the previous year’s National Computer Rank Examination and the National Applied Information Technology Certificate exams. The math exam items were provided by the National Examination Center and closely matched the current curricular requirements of high school students in China.
endline exam scores could be compared on a common scale. Placing the baseline and endline exam scores on a common scale allows us to measure absolute gains (or losses) in learning from the start of grade ten until the end of grade ten.

We administered and closely proctored the standardized computer and math exams during the baseline (October 2011) and endline surveys (May 2012). The exam scores were then normalized into z-scores (for computers and math separately and for the baseline and endline exams separately) by subtracting the mean and dividing by the standard deviation (SD) of the exam score distribution.7

In addition to gathering data on our outcome variables, our survey included three blocks pertaining to student background characteristics. The first block asked students to report their gender, age, whether their household registration (urban) status was rural or urban, and whether they had migrated before. As a part of this block, we also asked students to report their HSEE scores, the year they took the examination, and the prefecture where they took the examination.8

The second block gathered information on students’ families. This block included parental education level (a dummy indicator equal to 1 if neither parent finished junior high and 0 otherwise), parental migration status (whether both parents stayed at home between January 2011 to August 2011), and whether the student had any siblings.

The third block was used to identify whether students were from low-income backgrounds. Students were asked to fill out a checklist of household durable assets. We used principal components analysis, adjusting for the fact that the variables are dichotomous and not continuous, to calculate a single metric of the “family asset value” for each student (see Kolenikov and Angeles 2009).9 Low-income students are defined as those students whose family asset value was in the bottom 33% of the sample.

Before conducting any analyses, we trim observations that, for substantive reasons, clearly lie outside the common support shared by academic and vocational high school students. As detailed in our analysis section below, students

7. Although it is standard in education studies, we did not implement a reading test because principals and local education administrators were concerned that the time of the survey would be too long. Likewise, our Institutional Review Board was concerned that we would take too much instructional time away from our respondents by adding a third test.

8. Students are unlikely to suffer from recall bias when reporting on their HSEE scores because they received their scores only two months before the time of the survey. Granted, it is possible that vocational students remembered their scores as being lower than they actually were precisely because they ended up in vocational school. However, this would bias the results toward finding a positive effect of vocational school. As such, it would not challenge the findings of the paper.

9. We conduct standard robustness tests to see whether the use of polychoric PCA results in a viable family wealth metric. First, we find that the first principal component explains a large proportion of the variance in the family asset variables. The second and remaining principal components explain little of the variance. This indicates that the poverty metric reflects a common relationship underlying the inputs (wealth). Second, the scoring coefficients on the first principal component for each asset indicator all run in the anticipated directions. This means that the possession of assets indicates a higher first principal component score (wealth). Third, we find no evidence of clumping or truncation in our family wealth metric.
with extreme ages and test scores or that did not take the HSEE are dropped from the analyses. Thus, of the original 10,071 students, we first trimmed 263 students who scored in the bottom and top 1% of the baseline math and computer score distributions. Second, we trimmed away another 137 students whose age is outside the normal range for high school (roughly fourteen to nineteen years old). We further trimmed students that did not take the HSEE (1,279 students) or those who took the HSEE in years other than 2010 or 2011 (748 students). These students made schooling choices (whether or not to take the HSEE and thus apply for academic high school; whether or not to take the HSEE “on time” like “regular” students aspiring to go to academic high school) that were clearly different from academic high school students. By trimming away these students, we effectively controlled for school choice (between vocational and academic high school) before conducting our matching analyses. In total, there were 7,644 students remaining after our trimming procedure.

The attrition rate in our analytical sample was low. Of the 7,644 in our analytical sample, 367 students (4.8% of the sample) were absent (305) or on long-term sick leave (sixty-two). Another group of 583 students (or roughly 8% of the analytical sample) dropped out. For the students who dropped out, we recorded their dropout status and thus include them in our analyses of the impacts of attending vocational (versus academic) high school on dropout. However, measures of the computing and math skills of dropouts are missing for such students. We also test for attrition in appendix table S1 (appendix table S1 in the supplemental appendix, available at http://wber.oxfordjournals.org/). From appendix table S1, we can see that the attriting students (the majority of which are dropouts) differ from non-attriting students in baseline characteristics. Specifically, attritors were more likely to be male (column 2, significant at the 10% level), older (column 3, significant at the 5% level), have parents who are not at home (column 7, significant at the 1% level), have lower math scores (column 9, significant at the 1% level), and have lower computer scores (column 10, significant at the 10% level). Although there is imbalance between attriting and non-attriting students, the imbalance does not appear to bias our results. In particular, we estimate Lee Bounds (that account for problematic attrition) for the endline math and endline computer achievement outcomes (see our robustness check subsection 3.1.2 below).

As our study did not randomly assign students (to academic high school and vocational high school), we do not expect to see balance between the students that attended vocational high school and those that attended academic high school. Indeed, the groups differ substantially in terms of baseline characteristics (table 1). Vocational high school students are less likely to be among students with the lowest incomes (row 4), tend to be older (row 6), and have parents that tend to have migrated in the past (row 8). Moreover, their parents are less likely to have completed junior high (row 11). Although their math scores are much lower than academic high school students at the baseline (row 12), their computer scores are slightly higher (row 13). Because of these differences, outcomes such
as dropout rates or learning in vocational high schools could be due to the kinds of students who attend rather than the low quality of vocational high schools compared to (nonelite) academic high schools. Our analytical approach focuses on addressing this type of selection bias.

### 2.3 Analytical Approach

To assess the impact of attending vocational versus academic high school on student dropout rates, computing skills and math skills, we conduct three types of analyses: (a) ordinary least squares (OLS); (b) instrumental variable (or IV); and (c) matching analyses. Note that, in all three types of analyses, we estimate Huber-White standard errors that correct for prefecture-level clustering.¹⁰

#### TABLE 1. Differences between Vocational High School and Academic High School Students

<table>
<thead>
<tr>
<th>Variable</th>
<th>Academic high school</th>
<th>Vocational high school</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-income¹</td>
<td>0.40</td>
<td>0.31</td>
<td>-0.09**</td>
</tr>
<tr>
<td>Male</td>
<td>0.50</td>
<td>0.57</td>
<td>0.06</td>
</tr>
<tr>
<td>Age</td>
<td>15.97</td>
<td>16.14</td>
<td>0.17***</td>
</tr>
<tr>
<td>Urban</td>
<td>0.88</td>
<td>0.90</td>
<td>0.01</td>
</tr>
<tr>
<td>Student migrated</td>
<td>0.14</td>
<td>0.17</td>
<td>0.04***</td>
</tr>
<tr>
<td>Siblings</td>
<td>0.72</td>
<td>0.68</td>
<td>-0.03</td>
</tr>
<tr>
<td>Parents home</td>
<td>0.87</td>
<td>0.89</td>
<td>0.02</td>
</tr>
<tr>
<td>Parents no junior high</td>
<td>0.29</td>
<td>0.40</td>
<td>0.11***</td>
</tr>
<tr>
<td>Math baseline (z-score)</td>
<td>2.13</td>
<td>1.16</td>
<td>-0.97***</td>
</tr>
<tr>
<td>Computer baseline (z-score)</td>
<td>-0.33</td>
<td>-0.13</td>
<td>0.19***</td>
</tr>
</tbody>
</table>

¹The academic high school students appear to be economically poorer than the vocational high school students in our sample. This is likely because we sampled second tier, nonelite academic high schools that enrolled students with characteristics more comparable to those in vocational high schools. In addition, students who do not qualify for academic high schools have two choices: they may enter the labor market or attend vocational high schools. Students entering the labor market are typically from poorer backgrounds than those going into vocational high (Song et al. 2013). That is, a number of vocational high school students were children who were unable to test well enough to enter academic high schools but came from richer families that did not need to send their children to the labor market.

Note: Cluster-robust SEs in parentheses; ***p<0.01, **p<0.05, *p<0.1.

Source: Authors’ analysis based on data described in the text.

¹⁰ Although it would be most appropriate to adjust standard errors for clustering at the school level, we conservatively adjust for clustering at the higher levels of aggregation. In particular, we adjust for clustering at the prefecture level for the OLS and matching analyses (and the county level for the IV analyses) because we add in prefecture (county) fixed effects when estimating differences across treatment and control groups. In fact, the results of the paper are substantively the same whether we adjust the standard errors for clustering at the school level (without prefecture/county fixed effects) or prefecture/county level (with prefecture/county fixed effects). Results are available from the authors on request.
treatment (attending vocational versus academic high school) and student outcomes, while controlling for observable covariates that may confound that relationship. The basic specification for the OLS analysis is:

$$Y_{ij} = \alpha_0 + \alpha_1 V_{ij} + X_{ij}\alpha + \tau_p + \epsilon_{ij}$$

(1)

where $Y_{ij}$ represents the outcome variable of interest (dropout, computing, or math skills) of student $i$ in school $j$. $V_{ij}$ is a dummy variable for whether or not student $i$ attended vocational high school at the time of the baseline survey. In the absence of omitted variables bias, $\alpha_1$ would be the treatment impact of attending vocational (versus academic) high school on $Y_{ij}$.

The term $X_{ij}$ in equation (1) represents a vector of observable baseline covariates for student $i$ in school $j$. It includes student and family covariates such as male (equals 1 if the student is male and 0 if female), age (in days), urban (equals 1 if the student has urban residential permit status and 0 if rural), student migrated (equals 1 if the student has migrated prior to the baseline survey and 0 otherwise), siblings (equals 1 if the student has siblings and 0 otherwise), parents at home (equals 1 if both parents stayed at home between January 2011 to August 2011 and 0 otherwise), parents did not finish junior high (equals 1 if neither parent finished junior high school and 0 otherwise), and low-income (equals 1 if students are in the bottom 33% of the distribution of our family asset value variable and 0 otherwise). Importantly, we also control for baseline computer and math scores. Finally, we control for social, economic, and political differences in local context by adding a fixed effect term $\tau_p$ to indicate the prefecture where the student went to high school.11

2.3.2 Instrumental variables. For our second type of analysis, we conduct an instrumental variables (IV) analysis. We conduct the IV analysis because, in contrast to OLS, it can in theory produce causal estimates of the impact of vocational versus academic high school on student outcomes. In particular, whereas OLS fundamentally relies on the assumption of ignorability (that after controlling for observable pretreatment covariates, treatment assignment is independent of the outcome of interest), the IV analysis relies on two different assumptions (Murnane and Willett 2010). The first assumption is that of exogeneity: the IV should influence student outcomes only through the treatment variable (attending vocational versus academic high school) and not through any other channel. The second assumption is that the IV should be strongly correlated with the treatment variable in order to produce consistent treatment effect estimates. We discuss whether these two assumptions are met in our IV analysis immediately below.

11. The estimates of the impact of attending vocational schools on dropout in tables 2 and 3 are from a linear probability (OLS) model. Since dropout is a binary outcome, as a robustness check, we also estimated the impact using a logit model. The results from the logit model are substantively identical to the OLS results and are available upon request.
Our IV analysis exploits variation in student HSEE scores relative to an HSEE score cutoff. In China, HSEE scores determine entry into academic high school. Every county has a different cutoff for whether a student’s score makes him/her eligible to enter academic high school. Students with HSEE scores that are equal to or higher than the HSEE score cutoff in their county can go to academic high school. By contrast, students with HSEE scores that are lower than the cutoff can only go to vocational high school (or enter the labor market).

Significantly, while our approach is similar in spirit to a regression discontinuity design, we apply an IV strategy because standard sensitivity tests show that the typical RD design is not valid for our situation. In particular, due to the fact that we sampled (academic and vocational) high school students, the density of students that score just below the HSEE cutoff (to enter academic high school) is significantly less than the density of students that score just above the HSEE cutoff. This difference is not due to students’ ability to manipulate their HSEE scores. Rather, the difference arises because a large proportion of students that scored under the HSEE cutoff and did not get into academic high school chose to enter the unskilled labor market (where wages are relatively high—see Cai et al. 2008) instead of vocational high school. Since our sample does not include students that chose to enter the unskilled labor market, the RD design is not strictly valid for our situation.

We instead rely on an IV estimation strategy that still leverages the HSEE cutoff for academic (versus vocational) high school. Namely, our IV estimation strategy takes advantage of the strict assignment rule associated with the HSEE cutoff and yet controls for a number of important baseline covariates that, because of sample selection bias, may be correlated with the treatment and the outcome variables of interest (see appendix S1 for a full discussion, appendix S1 in the supplemental appendix, available at http://wber.oxfordjournals.org/). We further check the robustness of our IV results to sample selection bias (see subsection 3.1.2).

To apply the IV analysis, we first create an instrumental variable called below cutoff. Below cutoff equals 1 if a student scored below the HSEE cutoff in the county in which he/she took the HSEE and 0 if otherwise. We attempted to collect information on HSEE score cutoffs from each county in our sample prefectures for 2011 (the year in which the vast majority of students in our sample took the HSEE). In the end, we were able to collect HSEE score cutoffs from twenty-one of the seventy-five total sample counties, and the IV analysis is only among these twenty-one counties. Note that the HSEE test scores are not comparable across different prefectures because each prefecture administers a different test. In addition, although students within the same prefecture take the same

12. To attempt to keep information from the other counties, we did in fact try to infer the cutoff points by looking at the distribution of HSEE scores and vocational versus academic high school entrants in each county. Unfortunately, we were unable to identify large jumps in entry at particular HSEE score values. Failing to identify these jumps, we surmise that these other counties did not use a strict cutoff rule to determine entry into academic high school. This may be the reason why officials in these counties did not publicly publish their HSEE cutoffs (and the reason that we could not obtain information about these cutoffs from the county officials themselves).
HSEE, different counties within the same prefecture may use slightly different rules for grading the (same) HSEE test forms. For this reason, we always control for county fixed effects (using the county that each student took the HSEE in) when using our instrumental variable. By using below cutoff as an instrument for \( V_{ij} \) in equation 1, we assume that, conditional on baseline covariates, whether a student is below or above the HSEE cutoff exclusively affects his/her outcomes (dropout, specific skills, general skills) through his/her decision to attend vocational or academic high school. This is the exogeneity assumption of IV analysis.

We provide justification for why below cutoff may be an appropriate IV. Figures 1 map the relationship between each student’s HSEE score (centered at the HSEE score cutoff in the county he/she took the HSEE, x-axis) and the probability of attending vocational versus academic high school (\( V_{ij} \), y-axis). Figure 1a shows that the probability of attending vocational high school drops by over 50% at the HSEE cutoff. By contrast, the probability of attending vocational high school only drops by 10% or less at ten points to the right or left of the HSEE cutoff (figures 1b and 1c respectively). The probability of attending vocational high school hardly drops at all at twenty points to the right or left of the HSEE cutoff (figures 1d and 1e respectively). Figures 1, taken together with the fact that county officials set HSEE cutoffs after the HSEE is administered and scored, lend support to the idea that (in the absence of sample selection bias), the HSEE cutoff rule is likely exogenous. If controlling for baseline covariates can appropriately adjust for sample bias, the HSEE cutoff variable should be uncorrelated with (observable and unobservable) factors that influence the relationship between vocational high school attendance (\( V_{ij} \)) and student outcomes. In an attempt to control for possible sources of endogeneity, we control for \( X_{ij} \), HSEE score, and county fixed effects in all of our IV analyses.\(^{13}\)

Below cutoff also fulfills the second important assumption of IV analyses (Murnane and Willett 2010). Namely, the below cutoff variable is strongly correlated with \( V_{ij} \) in the first stage of the IV regression (see appendix table S2 and appendix table S2 in the supplemental appendix, available at http://wber.oxfordjournals.org/). Specifically, the first stage results show that the instrument has a strong and statistically significant (at the 1% level) relationship with the endogenous regressor. The weak identification tests (using the Craig-Donald Wald F Statistic) all reject the null hypotheses that the equations are weakly identified (with a p-value < 0.01).

2.3.3 Coarsened exact matching (CEM) analyses. As a robustness check on our IV analyses and also to see whether our IV analyses hold over a broader range of data (i.e., because the IV analyses were only for students from twenty-one counties with HSEE cutoff data), our third analysis is a matching

\(^{13}\) It is true that a small number of students (eighty-six out of 1754 or less than 5% of students) scored below the cutoff and yet managed to enter academic high school. This small number of students may have (unofficially) paid high fees to enter academic high school. Nevertheless, this is the exception and not the rule. As such, this phenomenon should not substantively change our analyses.
Figure 1. Graphs Showing the Discontinuity at the HSEE Cutoff (Between Attending Academic and Vocational High School) Figure 1a. At the HSEE Cutoff Figure 1b. 10 Points above the Cutoff Figure 1c. 10 Points below the Cutoff Figure 1d. 20 Points above the Cutoff Figure 1e. 20 Points below the Cutoff

Source: Authors’ analysis based on data described in the text.
exercise. This third analysis isolates the sample of vocational and academic high school students that are similar on baseline characteristics by using coarsened exact matching or CEM. The CEM procedure is comprised of three steps. In step one, each variable is recoded (or “coarsened”) so that substantively similar values of the variable are grouped and assigned the same numerical value. In step
two, students are matched “exactly” on the coarsened data: if either a vocational high school student or an academic high school student does not find one or more matches on the coarsened data, that student is dropped from the sample. In step three, the data are “uncoarsened” or returned to their original values for the students that were not dropped from the sample. The post-matching estimation procedure is conducted on the data from step three.

Why do we choose CEM over traditional matching methods like propensity score matching? First and most importantly, compared to propensity score and Mahalanobis distance matching (which belong to the Equal Percent Bias Reducing or EPBR class of matching methods—Rosenbaum and Rubin 1985), CEM can obtain unbiased estimates with fewer restrictions on the data (see appendix S2 for a detailed discussion and appendix S2 in the supplemental appendix, available at http://wber.oxfordjournals.org/). As a result and as shown across a wide variety of datasets (see Iacus et al. 2011), CEM typically finds better balance in baseline characteristics across treatment and control groups than matching methods from the EPBR class. Second, CEM automatically eliminates the extrapolation region and thus (unlike matching methods in the EPBR class) does not require a separate procedure by which to restrict the data to a common support. Third, CEM is robust to measurement error. Fourth, CEM works with multiply imputed data. Fifth, CEM is computationally fast. A full discussion for how CEM outperforms propensity score and Mahalanobis distance matching is in appendix S2.

Given our choice to apply CEM, we make two substantive choices. First, we choose to match students from vocational (treatment) and academic (control)
high schools on the baseline covariates $X_{ij}$ in equation (1). To ensure that we are comparing students who face similar educational choices within a similar local context, we also choose to match students (exactly) within the prefecture and year in which they took the high school entrance exam (HSEE).

Second, we also had to choose how much to coarsen each covariate (see appendix S2 for an explanation of coarsening). By way of example, we can choose to coarsen baseline math scores into quintiles, meaning that we can choose to create five equally sized bins of students based on the quintile of their baseline math score. It is by choosing how much to coarsen or bin each covariate (such as baseline math scores) that we can decide ex ante on the maximum amount of imbalance in covariates between the treatment and control groups. In our actual CEM analysis, we choose to coarsen the distributions of each of our baseline exam score variables (computing, math) into six equally spaced bins. We next coarsen age by year (where a year is defined by the calendar of a typical school year, e.g., from Sept. 1, 1985 to Aug. 31 1986). We also configure the CEM procedure to match students within (and not across) prefectures. All of the other covariates in $X_{ij}$ are dummy variables. As with exact matching, the CEM procedure uses the two values of each dummy variable to help create the bins on which we match treatment and control students.

The CEM procedure produces balance across the observable covariates. After applying the matching procedure, the vocational and academic high school students look similar on all of the baseline characteristics in equation 1 (appendix table S3 versus appendix table S4, and appendix table S3 and S4 in the supplemental appendix, available at http://wber.oxfordjournals.org/). As a robustness check, we also coarsen the baseline math and computer exam distributions into finer bins (from six up to fifteen bins each). Although the size of the matched sample decreases with the finer coarsening, we obtain similar results across the various matching specifications. The balance in baseline covariates is not just at the mean but also at different parts of the distribution of each covariate (see appendix table S2). Furthermore, as explained above, the use of CEM automatically ensures that the matched data share a common support. As such, we do not have to check to make sure that the matched data share a common support.

After matching the data using CEM, we run the same regression analyses as in equation (1) on the matched set of students. Like Iacus et al. (2011), we use doubly robust methods to estimate the causal effects: we use linear regressions (that adjust for baseline covariates) to estimate the impacts of attending vocational high school on student outcomes after matching the data. Our causal estimators are doubly robust in the sense that the estimators are unbiased if either the matching procedure or the regression specification is correctly specified (Ho et al. 2007). We call the regression analyses on the matched set of students our CEM analyses.

Granted, matching methods like CEM rely on the assumption of ignorability. That is, after controlling for observable covariates, no unobservable covariate is significantly correlated with both the treatment and outcome(s) of interest. While we cannot claim that CEM accounts for all possible confounding covariates, we
believe that the CEM model presented here controls for the main confounding influences for why students attend vocational high school even when they are eligible to attend academic high school (see appendix S2 for more details).

III. Results

3.1 What is the Impact of Attending Vocational (Versus Academic) High School?

3.1.1 Main Results. According to the results from the OLS analysis, students in vocational schools have different dropout rates and learn both computing and math skills at different rates than students from academic high schools. Specifically, students in vocational schools are 4 percentage points (or about 78 percent) more likely to drop out compared to students in academic high schools (table 2, row 1, column 1). The difference is statistically significant at the 1% level. The OLS regressions also show that students attending vocational high school do not improve computing skills more than students attending academic high school (row 1, column 2). Students in vocational high school scored only 0.02 SDs higher than academic high school students in computing skills (not statistically different from zero). Finally, in terms of math skills, students in vocational high school score far lower (0.44 SDs) than students in academic high school (row 1, column 3). The difference is significant at the 1% level. In summary, students attending vocational versus academic high schools drop out more, learn fewer general skills, and have no measurable advantage in learning specific skills.

The results from our IV analysis generally support the story that vocational high schools do not build human capital (table 3). Vocational high school students are 1.1 percentage points more likely to drop out (although this finding on differences in the dropout rate—unlike the OLS finding—is no longer statistically significant). The IV estimates of vocational schooling on computing and math skills remain consistent with the findings from the OLS analysis. Vocational schooling reduces math skills by 0.30 SDs (a finding significant at the 1% level). Moreover, there is no statistically significant evidence that attending VET improves computing skills (an increase of 0.12 SDs, p = 0.16). The magnitude of the point estimate, even if it were statistically significant, is not large given the much greater number of class hours spent on learning computing in vocational schools compared to academic schools.

The results of the CEM analysis also tell the same story (table 4). Attending vocational high school increases dropout rates by 3 percentage points (over academic high school students—row 1, column 1). This finding is significant at the 1% level. Attending vocational high school has a negligible effect on computing skills. Although vocational high school students appear to do slightly worse than their academic high school peers on the computer skills exams (by 0.05 SDs), the estimated coefficient is not statistically significant (row 1, column 2). The CEM analysis—which matches similar students from vocational high school and...
academic high schools—demonstrates that attending vocational high school decreases math skills by 0.42 SDs (significant at the 1% level—row 1, column 3).

Taken together, our findings demonstrate that attending vocational high school actually hurts students relative to attending academic high school. First, vocational high school encourages drop out (or at least does not encourage students to stay in school). Second, vocational high schools are failing to equip students with computing skills relative to academic high schools (which spend little class time teaching computing). Third, attending vocational versus academic high school results in a loss of math skills.

3.1.2 Robustness Checks. To test the sensitivity of our estimates, we conduct six sets of robustness checks. Our first set of robustness checks tests whether our IV analyses are robust when we adjust our IV estimation strategy in four ways: (a) add nonlinear controls of the running variable; (b) allow slopes to be different on either side of the cutoff; (c) limit the sample to students that are closer to (on either side of) the cutoff; and (d) relax the assumption of linearity by using a probit model. Our second set of robustness checks involves defining our sample

Table 2. Impact of Attending Vocational High School (versus Academic High School) on Student Outcomes—OLS Regressions with Fixed Effects (on Unmatched Data)

<table>
<thead>
<tr>
<th></th>
<th>(1) Dropout</th>
<th>(2) Computer endline</th>
<th>(3) Math endline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Went to VET</td>
<td>0.04***</td>
<td>0.02</td>
<td>−0.44***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.05)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Low-income</td>
<td>0.00</td>
<td>0.01</td>
<td>0.05***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Male</td>
<td>0.03***</td>
<td>0.01</td>
<td>−0.05**</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Age</td>
<td>0.01***</td>
<td>−0.02***</td>
<td>−0.05***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Urban</td>
<td>0.01</td>
<td>−0.03</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Student migrated</td>
<td>0.01</td>
<td>0.02</td>
<td>0.06*</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Siblings</td>
<td>0.01***</td>
<td>−0.01</td>
<td>−0.02</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.02)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Parents home</td>
<td>−0.04***</td>
<td>0.02</td>
<td>−0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Parents no junior high school</td>
<td>0.02***</td>
<td>−0.01</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Math baseline</td>
<td>−0.01***</td>
<td>0.05***</td>
<td>0.26***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Computer baseline</td>
<td>0.00</td>
<td>0.33***</td>
<td>0.19***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.04)</td>
</tr>
</tbody>
</table>

Observations: 7,299  6,395  6,395

Note: Cluster-robust SEs in parentheses; ***p<0.01, **p<0.05, *p<0.1.

Source: Authors’ analysis based on data described in the text.
differently: (a) by excluding dropouts from analyses of the impact of vocational (versus academic) high school on skills; and (b) by using a multiple imputation procedure to fill in (or predict) the missing outcome values of the dropout students and thereafter including these students in our analyses. A third set of checks involves ascertaining whether our estimates are sensitive to attrition, which we test for using Lee Bounds. A fourth set of robustness checks involves checking whether defining our variables differently (as continuous rather than binary variables, for example) would change our results. A fifth set of robustness checks tests whether sample selection could bias our IV estimates. Sixth (and related to the sample selection issue), we use a procedure suggested by Conley et al. (2012) to test the sensitivity of our IV estimates to deviations from the exogeneity assumption. In all cases, the results are not substantively different from the results from the models presented in the paper. While we do not display these results in the body of the text for the sake of brevity, they are presented in the

<table>
<thead>
<tr>
<th></th>
<th>(1) Dropout</th>
<th>(2) Computer Endline</th>
<th>(3) Math endline</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Went to VET</strong></td>
<td>0.01</td>
<td>0.12</td>
<td>−0.30***</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.08)</td>
<td>(0.11)</td>
</tr>
<tr>
<td><strong>Low-income</strong></td>
<td>0.01</td>
<td>−0.01</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.03)</td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td>0.01</td>
<td>0.06***</td>
<td>0.09***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>0.01**</td>
<td>−0.01</td>
<td>−0.08***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td><strong>Urban</strong></td>
<td>−0.003</td>
<td>−0.07***</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.06)</td>
</tr>
<tr>
<td><strong>Student migrated</strong></td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.05)</td>
</tr>
<tr>
<td><strong>Siblings</strong></td>
<td>0.01</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.04)</td>
</tr>
<tr>
<td><strong>Parents home</strong></td>
<td>−0.03***</td>
<td>0.02</td>
<td>−0.04</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.05)</td>
</tr>
<tr>
<td><strong>Parents no junior high school</strong></td>
<td>0.00</td>
<td>−0.02</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td><strong>Math baseline</strong></td>
<td>0.00</td>
<td>0.02**</td>
<td>0.17***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td><strong>Computer baseline</strong></td>
<td>0.01</td>
<td>0.30***</td>
<td>0.13***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.04)</td>
<td>(0.03)</td>
</tr>
<tr>
<td><strong>Centered HSEE score</strong></td>
<td>−0.00**</td>
<td>0.00***</td>
<td>0.00***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Observations</td>
<td>3,600</td>
<td>3,303</td>
<td>3,303</td>
</tr>
</tbody>
</table>

*Note:* Cluster-robust SEs in parentheses; **p < 0.01, *p < 0.05, p < 0.1.

*Source:* Authors’ analysis based on data described in the text.

### 3.2 The Impact of Vocational High Schools on Low-Income and Low-Ability Students

According to some policy documents (e.g., MOF and MOE 2006), vocational high schools are meant to benefit low-income and low-ability students. In this section, we examine the heterogeneous impacts of attending vocational (versus academic) high school on dropout rates and skills by income (poverty) level and ability. To do so, we rerun two additional versions of the IV analyses (one with an additional treatment-low-income interaction term; and one with an additional treatment-low-ability interaction term).

Our results show that low-income students not only fail to benefit from attending vocational high school, they actually perform worse (table 5). Low-income students who attend vocational versus academic high school are 5.9 percentage points more likely than higher income students to drop out (significant at the 10% level—column 1). Furthermore, like the average student (as

---

**Table 4. Impact of Attending Vocational High School (versus Academic High School) on Student Outcomes—OLS Regressions on Matched Data**

<table>
<thead>
<tr>
<th></th>
<th>(1) Dropout</th>
<th>(2) Computer endline</th>
<th>(3) Math endline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Went to VET</td>
<td>0.03***</td>
<td>-0.05</td>
<td>-0.42***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.08)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Low-income</td>
<td>-0.00</td>
<td>0.05</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Male</td>
<td>0.04***</td>
<td>0.05</td>
<td>-0.07**</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.06)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Age</td>
<td>0.01</td>
<td>-0.04</td>
<td>-0.07**</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Urban</td>
<td>0.02**</td>
<td>-0.12</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.15)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Student migrated</td>
<td>0.04**</td>
<td>-0.02</td>
<td>0.24***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Siblings</td>
<td>0.01</td>
<td>0.10***</td>
<td>-0.11**</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.04)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Parents home</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.10)</td>
</tr>
<tr>
<td>Parents no junior high school</td>
<td>0.03***</td>
<td>-0.06</td>
<td>-0.00</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.06)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Math baseline</td>
<td>-0.01**</td>
<td>0.04***</td>
<td>0.22***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Computer baseline</td>
<td>-0.01</td>
<td>0.33***</td>
<td>0.19***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Observations</td>
<td>2,122</td>
<td>1,927</td>
<td>1,927</td>
</tr>
</tbody>
</table>

*Note:* Cluster-robust SEs in parentheses; ***p<0.01, **p<0.05, *p<0.1.

*Source:* Authors’ analysis based on data described in the text.
shown in the subsection above), low-income students make negligible gains in computing skills (column 2) while losing in math skills (column 3).

As with our results for low-income students, attending vocational high school has negative impacts on low-ability students. Low-ability students who attend vocational versus academic high school are more likely to drop out compared to higher ability students (the dropout rate increases by 2.5 percentage points for every 1 SD decrease in baseline computer scores—table 6, column 1). In addition, by attending vocational (versus academic) high school, low-ability students are even less likely to gain computing skills compared to higher ability students (the endline computer scores decrease by 0.13 SDs for every 1 SD decrease in baseline computer scores—column 2). Finally, by attending vocational (versus academic) high school, low-ability students may see their math skills deteriorate more than higher ability students (by .06 SDs for every 1 SD decrease in baseline math scores—column 3).

### Table 5. Heterogeneous Impacts of Attending Vocational High School (versus Academic High School) on Low-income Student Outcomes (IV Analyses, 2011 HSEE Takers from 21 Counties)

<table>
<thead>
<tr>
<th></th>
<th>(1) Dropout</th>
<th>(2) Computer endline</th>
<th>(3) Math endline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Went to VET</td>
<td>0.00</td>
<td>0.12</td>
<td>−0.30**</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.08)</td>
<td>(0.12)</td>
</tr>
<tr>
<td>VET*Low-income</td>
<td>0.06*</td>
<td>−0.05</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.05)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Low-income</td>
<td>−0.02</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Male</td>
<td>0.01</td>
<td>0.06***</td>
<td>0.09**</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Age</td>
<td>0.01**</td>
<td>−0.01</td>
<td>−0.08***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Urban</td>
<td>0.00</td>
<td>−0.07**</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Student migrated</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Siblings</td>
<td>0.01</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Parents home</td>
<td>−0.03***</td>
<td>0.02</td>
<td>−0.04</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Parents no junior high school</td>
<td>0.00</td>
<td>−0.02</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Math baseline</td>
<td>0.00</td>
<td>0.02**</td>
<td>0.17***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Computer baseline</td>
<td>0.00</td>
<td>0.30***</td>
<td>0.13***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.04)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Centered HSEE score</td>
<td>−0.00**</td>
<td>0.00***</td>
<td>0.00***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Observations</td>
<td>3,600</td>
<td>3,303</td>
<td>3,303</td>
</tr>
</tbody>
</table>

*Note*: Cluster-robust SEs in parentheses; ***p<0.01, **p<0.05, *p<0.1.

*Source*: Authors’ analysis based on data described in the text.
Table 6. Heterogeneous Impacts of Attending Vocational High School (versus Academic High School) on Low-Ability Student Outcomes (IV Analyses, 2011 HSEE Takers from 21 Counties)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dropout</td>
<td>Computer Endline</td>
<td>Math endline</td>
</tr>
<tr>
<td>Went to VET</td>
<td>0.02</td>
<td>0.09</td>
<td>-0.31***</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.08)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>VET*computer_baseline</td>
<td>-0.03*</td>
<td>0.13***</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.04)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Male</td>
<td>0.01</td>
<td>0.06***</td>
<td>0.09**</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Age</td>
<td>0.01**</td>
<td>-0.01</td>
<td>-0.08***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Urban</td>
<td>0.00</td>
<td>-0.07**</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Student migrated</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Siblings</td>
<td>0.01</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Parents home</td>
<td>-0.03***</td>
<td>0.02</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Parents no junior high school</td>
<td>0.00</td>
<td>-0.02</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Low-income</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Math baseline</td>
<td>0.00</td>
<td>0.02**</td>
<td>0.17***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Computer baseline</td>
<td>0.02**</td>
<td>0.24***</td>
<td>0.10*</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.04)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Centered HSEE score</td>
<td>-0.00*</td>
<td>0.02***</td>
<td>0.03***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Observations</td>
<td>3,600</td>
<td>3,303</td>
<td>3,303</td>
</tr>
</tbody>
</table>

Note: Cluster-robust SEs in parentheses; ***p<0.01, **p<0.05, *p<0.1.
Source: Authors’ analysis based on data described in the text.

decrease in baseline computer scores, although the result is not statistically significant at the 10% level—column 3).

Taken together, the findings indicate that attending vocational high school may hurt disadvantaged (low-income and low-ability) students even more than their advantaged counterparts. Low-income and low-ability students who attend vocational (rather than academic) high school drop out more than the higher income and ability students. There is also some evidence to indicate that low-income and low-ability students are even less likely to gain computing skills than higher income and higher ability students and at least as likely to see a reduction in their math skills compared to higher income and higher ability students. These findings are true even though vocational schools are (by design) supposed to benefit such students. For this reason, according to our results, we conclude that low-income and low-ability students would have fared better in academic high schools.
3.3 IRT Gains in General and Specific Skills

The results in the previous subsections demonstrate that attending vocational high school has a negative impact on student outcomes. However, our analysis can go further. Because our standardized exams were vertically scaled using IRT, we are able to analyze the individual gains in general and specific skills for the sample vocational and academic high school students. This analysis will help us determine if vocational high school students are learning.

Surprisingly, the IRT-scaled gains show that vocational high school students are actually losing math skills (figure 2). The IRT-scaled math scores of students in vocational high school fell by 0.08 SDs from the beginning to the end of grade 10. By contrast, students in academic high schools gained 0.04 SDs in math over the same period. In other words, the results show that vocational high school students are not only falling behind academic high school students, they are actually losing skills they previously had.

There are somewhat more encouraging results in terms of specific skills. According the IRT-scaled test results, vocational high school students make modest gains in computer skills (figure 3). On average, the IRT-scaled computer scores of vocational high school students rose by 0.12 SDs. However, as would be expected (from the subsections above), vocational high school students made fewer gains in specific skills than academic high school students, even as they spend much less time in computer classes than their vocational counterparts. The computer scores of academic high school students (in nonelite academic high schools) rose by 0.23 SDs.

These results suggest that, in absolute terms, vocational high schools make only small contributions to or may even detract from human capital development. While it is true that vocational high school students make modest gains in their computing skills, as a whole their gains are less than those in academic high

14. In fact, this graph only examines the IRT-scaled math gains among the lowest scoring 50% of students at the baseline. We make this adjustment because our baseline results were right-censored (a ceiling effect). Including these students would have biased the estimate of gains upward, as students scoring full marks at the baseline actually could have scored higher. In spite of this adjustment and ceiling effect, our main analytic models which compare the impact of vocational versus academic high school are unaffected.

15. What explains the surprising result that academic high school students are learning so little math and vocational high school students are learning so little in computers? One reason to suspect that academic high school students learned so little mathematics is that our schools are nonelite academic high schools, where the quality of schooling is not always high. Another reason for the low mathematics gains for academic high school students is that they were entering a new school (their first year of high school). As is common in the United States, for example, students entering a new schooling environment may have muted learning gains as they adjust to their new environment (e.g., see Roderick and Camburn 1999).

16. Why did academic high school students appear to make higher gains in computers even compared to mathematics? Students have a minimum exposure to computers in junior high school. As such, their first real exposure to computers comes at the high school level. Hence, even though academic high schools only reserve one course for computers per week, the first systematic exposure to computers may have resulted in higher gains compared to mathematics. In addition, although it may appear peculiar that vocational high school students were learning fewer computer skills than academic high school students, the fact that the tests were constructed to match national standards suggests that vocational high schools were simply not teaching well (or students were not learning well).
More importantly, vocational high school students are actually losing math skills.

**IV. Conclusion**

Overall, VET at the high school level does appear to be meeting its mandate of equipping students with the human capital needed to succeed in China’s future economy. Specifically, attending vocational high school appears to cause students (in the computer major) to drop out of school if they are of low-income and of low-ability. Our results also show that attending vocational high school has no significant effect on specific skills and a substantial, negative impact on general skills relative to attending academic high school. This negative impact is pronounced among both low-income students and low-ability students. Finally, in absolute terms, vocational high school even detracts from students’ math skills over the course of the first year (from the start to the end of grade ten). If our results generalize to other provinces and majors, vocational high schools are failing to contribute to human capital development in China.17

17. It is conceivable that the students were performing poorly in vocational school because they were still adjusting to a new environment. While we do not have longer-term follow-up data of our sample students, we do have the dropout rate of the cohort of students in our sample over time. We started in the baseline with a total of 7,114 vocational high school students in our sample. Of these, 649 dropped out at the end of the first year (as noted in this paper). By the second year, an additional 751 students dropped out. That is a cumulative dropout rate of 20% over the first two years. (For comparison, the estimated two-year dropout rate in academic high schools is 5%—Loyalka et al., 2014). If students, in fact, felt like they were learning at the second year, the assumption is that it would have reduced the dropout rate. Since the dropout rate was even more serious in the second year, we conjecture that students were not learning more in their second year (assuming that other factors influencing dropout were also constant across the two years).
There are reasons to believe that these results are conservative. First, in our more robust models (the IV and matching estimates), we are actually comparing students around the HSEE cutoff. One implication of this method is that our results generalize to the “cream of the crop” in vocational high school. These are primarily students who scored high enough to be within reach of attending academic high school. If we were to use a counterfactual that allowed us to estimate the effect of attending vocational high school on all vocational high school students, the negative effects of vocational high school might be even larger.

Second, when selecting our sample, we chose schools with relatively large and stable enrollments in the computing major. If enrollments correlate with the quality of the school (as they do in academic schooling in China), our sample consists of higher-quality schools. If we estimated the effects of attending vocational high school among all vocational high schools, the negative impacts on dropout and skills would be even larger.

Overall, the lack of value-added in math or computing skills is likely to mute or decrease future labor market payoffs. Granted, it could be that vocational high schools produce value-added in noncognitive (or social and behavioral) skills, such that there is still a net labor market return even without value-added in general or specific (cognitive) skills. In fact, Kim (2013) finds positive labor market returns on vocational education in Korea. However, the fact that Chinese computer major students are losing their math skills and barely learning any

---

18. The reason is that students who scored at the bottom of the HSEE distribution have no opportunity to enter academic high school, thus making the assumption that they were randomly assigned to academic high school not credible. Likewise, students who score at the top of the HSEE distribution almost never attend vocational high school.
computing skills suggests that they also may not be acquiring social or behavioral skills either.19

Why does VET at the high school level appear to fail at generating human capital? While a full discussion of this question is beyond the scope of this study, one argument is that local governments (who are responsible for financing vocational high schools) are still failing to invest sufficient resources into vocational high schools. A related argument is that local governments favor academic over vocational high schools and deny appropriate resources like qualified teachers or finances to the latter (Yang 2012).

In fact, evidence suggests that this is not the case. We compare the vocational and academic high schools across a set of inputs—including the percentage of teachers with a college degree; the percentage of teachers with professional experience; computers per student; total school area (in square meters) per student; whether the school has laboratories; libraries or multimedia rooms; and expenditures per student (in RMB). We find that, with one exception, there are no substantial or statistically significant differences between vocational high schools and academic high schools (appendix table S7). The exception is that vocational high schools have more computers per student. As such, vocational high schools appear to be on equal (if not marginally better) footing with academic high schools in terms of basic inputs.

A second possibility is a lack of coordination and oversight to ensure the transformation of inputs (e.g., financial investments) into outputs (e.g., student skills). Multiple ministries/departments/bureaus oversee vocational education, thus reducing coordination and sharing of best practices between schools. Moreover, none of the ministries/departments/bureaus have developed protocols to systematically monitor vocational high school quality in a standard fashion. As such, there is almost no oversight over the quality of these schools.

A limitation of our study is that we focus only on students in the computer major and only test their math and computer skills. Ideally, our study would have included students in other majors and tested a wider set of skills. Therefore, strictly speaking, our results do not apply to other majors and other types of skills in vocational schooling in China. However, if our findings on quality are generally true and if the reasons for this poor quality are as we surmise, policymakers in China may wish to cease the large, almost indiscriminate investment into the vocational high school system. Instead they may wish to direct more resources toward the apparently more effective approach to human capital development: academic high school.

19. If the value-added of vocational high schools was so low, why did students still attend? While we cannot be sure, we conjecture that the failure of the government to provide clear, public (or open) information on the quality of schools has created information asymmetries. That is, students may not have actually known what the quality of the schools was before they decided to attend. As such, students believed that they would learn skills when they began vocational high schools (even if, on average, this was not true).
Furthermore, the results of this study should give pause to policymakers seeking to promote VET in other developing countries. Our results show that, at the margin, students in the most popular major in vocational school in two Chinese provinces lose general skills without any apparent gain in specific skills. While our results do not strictly apply to other developing countries, the fact that our findings are consistent with results from Indonesia (Chen 2009) and Romania (Malamud and Pop-Eleches 2008) suggests that other developing countries with substantial investments in vocational secondary education may also fail to enjoy significant returns to their investment. By diverting resources away from academic high school, developing countries may be reducing the number of students who can access a human-capital enhancing opportunity to attend academic high school. Together, such a policy could substantially hinder human capital production.

References


How does access to finance impact consumption volatility? Theory and evidence from advanced economies suggests that greater household access to finance smooths consumption. Evidence from emerging markets, where consumption is usually more volatile than income, indicates that financial reform further increases the volatility of consumption relative to output. This puzzle is addressed in the framework of an emerging economy model in which households face shocks to trend growth rate, and a fraction of them are financially constrained, with no access to financial services. Unconstrained households can respond to shocks to trend growth by raising current consumption more than the rise in current income. Financial reform increases the share of such households, leading to greater relative consumption volatility. Calibration of the model for pre- and post-financial reform in India provides support for the model's key predictions. JEL Codes: C50, E10, E21, E32

I N T R O D U C T I O N

Emerging economies have been seen to witness an increase in consumption volatility relative to output volatility after financial development. This behaviour appears puzzling since traditional models and evidence from advanced economies suggests that consumption should become smoother after financial constraints are reduced. This puzzle can be explained in a model featuring financial constraints and shocks to trend growth of productivity. The model predicts that...
relative consumption volatility rises when more consumers can access financial services.

The presence of financial constraints, such as credit constraints or lack of access to financial services in an economy, explains the excess volatility of consumption and its sensitivity to anticipated income fluctuations. A model featuring financially constrained consumers predicts that consumption cannot be smoothed fully. But in such a model, the volatility of consumption can be at least as high as income volatility or, at most, one. Further, if constraints are eased, the model predicts a reduction in relative consumption volatility.

Another feature of emerging economy models is the presence of shocks to trend growth of productivity. Large shocks to the permanent component of income originated from frequent policy regime shifts in emerging economies, relative to transitory income shocks, explain larger fluctuations in consumption relative to output fluctuations (Aguiar and Gopinath 2007). Unlike developed countries characterised by large transitory movements in income around the trend, shocks to trend growth are the primary source of fluctuations in emerging economies. When households anticipate a higher growth rate of income, which eventually leads to a rise in future income, they respond to this permanent income shock by increasing current consumption more than the rise in current income via borrowing against the future income or reducing current savings. As a result, consumption fluctuates more than income in emerging economies. This feature results in the relative volatility of consumption in emerging economies becoming greater than one.

A common feature of reform in emerging economies is financial sector reform. The increase in the access of households to finance resulting from reform allows households to smooth consumption over their lifetimes. But at the same time, emerging economies witness large shocks to the permanent component of income, relative to transitory income shocks. The combination of the response of households to permanent income shocks and the easing of financial constraints can yield an increase in the relative volatility of consumption.

The goal of this paper is to understand the joint impact of easing of financial constraints and permanent income shock on consumption volatility. This is analysed in a dynamic general equilibrium model with heterogeneous type agents. The model assumes that some households in the economy do not have access to finance. They can neither save nor borrow. These financially constrained households cannot smooth consumption over their lifetimes. The rest of the households in the economy are unconstrained and respond to a perceived income shock by smoothing consumption. Shocks to income that are perceived to be permanent lead to an increase in current period consumption higher than the increase in current period income. Only unconstrained households can increase consumption by more than the increase in income, either by borrowing against future income or reducing current savings. Constrained households can only increase consumption by the amount income has increased. Financial sector reform allows more households to access financial services. Now more households become unconstrained
and can respond to the income shock that they perceive to be permanent. The key prediction of this model is that financial development in an emerging economy leads to an increase in relative consumption volatility.

This prediction can be tested. The model is calibrated to Indian data for the pre- and post-reform years. All of the parameters, except for the share of financially constrained consumers, are kept unchanged. Financial inclusion is captured via a reduction in the fraction of constrained households in the post reform period. The results support the model’s key prediction.

This paper makes a contribution towards understanding the joint impact of financial development and permanent income shock on consumption volatility. It contributes to a growing literature that studies the effects of financial frictions on volatility. Earlier work mainly analyses the effect of domestic financial system development on output and consumption volatility through its effect on firms (Aghion et al. 2004, 2010). Some papers focus on the impact of financial globalisation on volatility (Aghion et al. 2004; Buch et al. 2005; Leblebicioglu 2009). The effect of domestic financial system development on output and consumption volatility is explored in a limited strand of literature. Iyigun and Owen (2004) propose a theory of income inequality in rich and poor countries as the cause of consumption volatility whose mechanics partly resemble those of the present model, once appropriately re-interpreted.

The model takes into account the broadly acknowledged fact that in emerging economies all consumers do not have access to finance (Honohan 2006). Financially constrained households are modelled as in Hayashi (1982) and Campbell and Mankiw (1991). The framework includes shocks to trend growth as in Aguiar and Gopinath (2007).

The rest of the paper is organised as follows: The Consumption Volatility and Financial Development section presents evidence on relative consumption volatility and financial development in emerging economies. The Consumption Volatility and Permanent versus Transitory Income Shocks section discusses the role of the relative magnitude of permanent and transitory income shocks for consumption volatility in developed vis-à-vis emerging economies. The Financial Frictions and Consumption Volatility: Theoretical Framework section presents the model and its predictions. The Case Study: Evidence for India section presents the calibration exercise and results. The Financial Development, Permanent Income Shock, and Relative Consumption Volatility in a Small Open Economy section presents the implications in a small open economy setup. The final section concludes.

**Consumption Volatility and Financial Development**

Recent empirical evidence on emerging economy business cycles shows an increase in the volatility of consumption relative to that of output after financial sector reform in Asia, Turkey, and India (Kim et al. 2003; Alp et al. 2012; Ghate et al. 2013). The relative volatility of consumption in the pre- and post-financial sector reform period for some developing countries are estimated (table 1). The
The choice of the date on which reform took place is based on Kim et al. (2003), Singh et al. (2005), Rodrik (2008), Alp et al. (2012), and Aslund (2012). The analysis is based on annual data for a set of emerging economies.1 The volatility of consumption relative to that of output in these countries, in the pre- and post-reform period, shows that many emerging economies exhibit similar behaviour in that relative consumption volatility increases after reform (table 1).

Financial development has been a major component of reform. A commonly used indicator of financial development, namely, total bank deposits to GDP ratio, for a set of emerging economies, on average, shows a rise in the indicator over time (figure 1). The rising trend in the ratio is also visible for individual countries (figure 1).

The indicators on financial depth, depicted by the density of commercial bank branches and depositors with commercial banks in emerging economies, in the

<table>
<thead>
<tr>
<th>Region &amp; reform date</th>
<th>Relative consumption volatility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-reform</td>
</tr>
<tr>
<td>Latin America: 1990</td>
<td></td>
</tr>
<tr>
<td>Chile</td>
<td>1.10</td>
</tr>
<tr>
<td>Colombia</td>
<td>0.97</td>
</tr>
<tr>
<td>Mexico</td>
<td>0.94</td>
</tr>
<tr>
<td>Peru</td>
<td>1.09</td>
</tr>
<tr>
<td>East Asia: 1996</td>
<td></td>
</tr>
<tr>
<td>Indonesia</td>
<td>2.45</td>
</tr>
<tr>
<td>Malaysia</td>
<td>1.36</td>
</tr>
<tr>
<td>Philippines</td>
<td>0.73</td>
</tr>
<tr>
<td>Korea</td>
<td>0.93</td>
</tr>
<tr>
<td>Taiwan</td>
<td>1.84</td>
</tr>
<tr>
<td>Thailand</td>
<td>0.88</td>
</tr>
<tr>
<td>East Europe: 1990</td>
<td></td>
</tr>
<tr>
<td>Turkey</td>
<td>1.07</td>
</tr>
<tr>
<td>Poland</td>
<td>0.92</td>
</tr>
<tr>
<td>Hungary</td>
<td>1.01</td>
</tr>
<tr>
<td>South Asia</td>
<td></td>
</tr>
<tr>
<td>India: 1992</td>
<td>0.83</td>
</tr>
<tr>
<td>Africa</td>
<td></td>
</tr>
<tr>
<td>South Africa: 1994</td>
<td>1.42</td>
</tr>
<tr>
<td>Mean</td>
<td>1.15</td>
</tr>
<tr>
<td>Std. dev.</td>
<td>0.44</td>
</tr>
</tbody>
</table>

Source: Datastream, author’s calculations.

This table shows the reform date and the volatility of consumption relative to that of output in the pre- and post-reform period for a set of emerging economies.

1. The span of the analysis varies across countries given the availability of the data. Table S1.1 in the Supplemental Appendix S1, available at http://wber.oxfordjournals.org/, lists period of analysis for each country. The reform date for each region, and the sources of the documentations indicating the reform dates are also reported in this table.
beginning and in the end of the last decade, indicate an increase in access of households to finance (table 2).

The above evidence suggests that the relative volatility of consumption rises after financial sector reform. This appears puzzling and cannot be explained by

**TABLE 2. Access to Finance**

<table>
<thead>
<tr>
<th>Country</th>
<th>Commercial bank branches per 100,000 adults</th>
<th>Depositors with commercial banks per 1,000 adults</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chile</td>
<td>13</td>
<td>18</td>
</tr>
<tr>
<td>Colombia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mexico</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>Peru</td>
<td>4</td>
<td>50</td>
</tr>
<tr>
<td>Indonesia</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Malaysia</td>
<td>13</td>
<td>..</td>
</tr>
<tr>
<td>Philippines</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Korea</td>
<td>17</td>
<td>19</td>
</tr>
<tr>
<td>Taiwan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>Turkey</td>
<td>13</td>
<td>..</td>
</tr>
<tr>
<td>Poland</td>
<td>37</td>
<td>46</td>
</tr>
<tr>
<td>Hungary</td>
<td>14</td>
<td>17</td>
</tr>
<tr>
<td>India</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>South Africa</td>
<td>5</td>
<td>10</td>
</tr>
</tbody>
</table>

*Source: Financial Inclusion, World Development Indicators.*

This table depicts the density of commercial bank branches and depositors with commercial banks in emerging economies in the beginning and in the end of the decade of 2000–10.
the existing literature. It supports the evidence in Kim et al. (2003), Alp et al. (2012), and Ghate et al. (2013), who allude to the increase in relative consumption volatility after financial sector reform.

**Consumption Volatility and Permanent Versus Transitory Income Shocks**

Empirical literature on business cycle stylised facts document business cycle properties in developed economies (Kydland and Prescott 1990; Backus and Kehoe 1992; Stock and Watson 1999; King and Rebelo 1999) and developing countries (Agenor et al. 2000; Rand and Tarp 2002; Male 2010). One of the key business cycle features that distinguishes emerging economies from advanced countries is the greater fluctuations in consumption relative to income fluctuations. Aguiar and Gopinath (2007) relate this difference in consumption behaviour in the two sets of countries, to the relative magnitude of permanent and transitory shocks to income.

The authors estimate a standard small open economy real business cycle model for Mexico, as a representative of the emerging economies, and Canada, representing advanced countries. The main finding is that large shocks to the growth rate of permanent components of productivity are the primary sources of fluctuations in emerging economies. In contrast, advanced economies are characterised by fluctuations around a stable trend, caused by large shocks to transitory component of productivity. The differences in technology shock processes cause households to respond differently to income shocks in developed and emerging economies. When households anticipate a higher growth rate of income which eventually leads to a rise in future income, they respond to this permanent income shock by increasing current consumption more than the rise in current income via borrowing against the future income or reducing current savings. As a result, consumption fluctuates more than income in emerging economies. This feature results in the relative volatility of consumption in emerging economies being greater than one.

*Positive Correlation between the Size of Trend Growth Shock and Relative Consumption Volatility: Evidence from Literature*

The positive correlation between the magnitude of shocks to trend growth and relative consumption volatility, found in the literature, is documented in table 3. The third and fifth columns of the table show technological shock processes for Mexico and Canada, along with output and consumption volatilities estimated from the model in Aguiar and Gopinath (2007). The second and fourth columns also document the empirical volatilities in output and consumption for these two countries. The table shows that Mexico, with consumption volatility relative to output volatility greater than one, is characterised by a larger shock to the growth rate of permanent component of technology $\sigma_g$ compared to the transitory shock $\sigma_a$. In contrast, Canada, with a relative consumption volatility less than
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mexico</td>
<td>Canada</td>
<td>Developed</td>
</tr>
<tr>
<td></td>
<td>Data</td>
<td>Model</td>
<td>Data</td>
</tr>
<tr>
<td>(\sigma_y)</td>
<td>2.40</td>
<td>2.13–2.40</td>
<td>1.55</td>
</tr>
<tr>
<td>(\sigma_c)</td>
<td>3.02</td>
<td>3.02–3.27</td>
<td>1.15</td>
</tr>
<tr>
<td>(\sigma_c/\sigma_y)</td>
<td>1.26</td>
<td>1.10–1.33</td>
<td>0.74</td>
</tr>
<tr>
<td>(\rho_g)</td>
<td>0.00–0.11</td>
<td>0.03–0.29</td>
<td></td>
</tr>
<tr>
<td>(\sigma_g)</td>
<td>2.13–3.06</td>
<td>0.47–1.20</td>
<td>2.89</td>
</tr>
<tr>
<td>(\rho_s)</td>
<td>0.95</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td>(\sigma_s)</td>
<td>0.17–0.54</td>
<td>0.63–0.78</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Source: Aguiar and Gopinath (2007), Naoussi and Tripier (2013), authors’ analysis outlined in the Consumption Volatility and Permanent versus Transitory Income Shocks section.

This table depicts cross country relative consumption volatility vis-à-vis the magnitude of shocks to trend growth documented from literature.
one, is characterised by larger transitory shocks compared to fluctuation in the permanent component of productivity.

Similarly, Naoussi and Tripier (2013) estimate a real business cycle model with transitory and trend shocks to productivity for eighty-two countries, including developed, emerging, and Sub-Saharan African (SSA) countries. They find that magnitudes of trend shocks are positively correlated with relative consumption volatilities. Columns 6 to 11 in table 3 summarise their findings. Relative consumption volatilities and shock to trend growth rate are found to be highest for SSA countries, followed by emerging and developed economies.

Finally, column 12 of table 3 shows the nature of technology shock processes for India. The estimation of the technology shock processes in India are outlined in the following section.

**Decomposition of Indian Total Factor Productivity (TFP) Series to Permanent and Transitory Components**

To have an account of transitory and trend growth shock in the Indian TFP series, the series is decomposed into permanent and transitory components using Kalman filter. First, the TFP series for India is estimated following an aggregate production function approach. The aggregate production function, representing the production sector in the model outlined in the next section, is defined following Aguiar and Gopinath (2007) as

\[
Y_t = e^{a_t}K_t^{1-\alpha}(G_t)^\alpha,
\]

where \(K_t\) is the aggregate stock of capital and \(\alpha \in (0, 1)\) denotes labour’s share of output. Households are assumed to supply unit labour inelastically. The parameters \(a_t\) and \(G_t\) represent productivity processes. The two productivity processes are characterised by different stochastic properties. The parameter \(a_t\) captures a transitory movement in productivity and is characterised by the following AR(1) process:

\[
a_t = \rho_a a_{t-1} + \varepsilon_a^t; \quad |\rho_a| < 1, \quad \varepsilon_a^t \sim N(0, \sigma_a^2).
\]

(2)

The parameter \(G_t\) represents the cumulative product of growth shocks as follows:

\[
\ln \left( \frac{g_t}{\mu_g} \right) = \rho_g \ln \left( \frac{g_{t-1}}{\mu_g} \right) + \varepsilon_g^t; \quad |\rho_g| < 1, \quad \varepsilon_g^t \sim N(0, \sigma_g^2),
\]

(3)

where \(\mu_g - 1\) is the long-run mean trend growth rate. The two different productivity processes are assumed to distinguish shock process in the level of productivity \(a_t\)
and the growth rate of productivity $g_t$. The growth shocks are incorporated in a labour-augmenting way to ensure the existence of a steady state where all variables grow at the rate $\mu_g$ and the tractability of analysis of cyclical properties of the model economy. In this analysis, the cyclical component of a variable $X_t$, that is, the deviation of the variable from its trend path is defined as $x_t = X_t/T_{t-1}$.

The Solow residual from the aggregate production function captures productivity processes that contains a transitory and a permanent component:

$$sr_t = a_t + \alpha \ln \Gamma_t = \ln Y_t - (1 - \alpha) \ln K_t.$$  \hspace{1cm} (4)

Since, the households supply unit labour inelastically and total mass of households is normalised to one, equation (4) measures the Solow residual in terms of per capita output and capital stock. In estimating the Solow residual for India, GDP at factor cost and net fixed capital stock, both in 2004-05 constant prices, proxy for output and capital stock, respectively. The data on GDP and net fixed capital stock are sourced from National Accounts Statistics. The labour force data are sourced from the World Bank. The value of labour share is set to 0.7 from Verma (2008). Given the availability of data on labour force and capital stock, the Solow residual series spans 1980-2009.

The transitory and permanent components in the Solow residual series for India are estimated using the Kalman filter. The underlying model is the following: the Solow residual series $sr_t$ is a sum of a trend component $T_t$ and a transitory or cyclical component $C_t$:

$$sr_t = T_t + C_t + V_t; \quad V_t \sim N(0, \sigma_v^2),$$

$$T_t = d + T_{t-1} + W_{1t}; \quad W_{1t} \sim N(0, \sigma_{W1}),$$

$$C_t = \rho_c C_{t-1} + W_{2t}; \quad |\rho_c| < 1; \quad W_{2t} \sim N(0, \sigma_{W2}).$$  \hspace{1cm} (5)

where $V_t$ represents measurement error. The trend component is assumed to follow a random walk process. This Trend-Cycle model in equation (5) can be represented in state-space form as:

$$sr_t = [1 \quad 1] \begin{pmatrix} T_t \\ C_t \end{pmatrix} + V_t,$$

$$\begin{pmatrix} T_t \\ C_t \end{pmatrix} = \begin{pmatrix} d & 0 \\ 0 & \rho_c \end{pmatrix} \begin{pmatrix} T_{t-1} \\ C_{t-1} \end{pmatrix} + \begin{pmatrix} W_{1t} \\ W_{2t} \end{pmatrix}. $$  \hspace{1cm} (6)

The first expression in equation (6) represents the observation equation in terms of the unobserved states. The second equation represents the transition dynamics of the state variables. Figure 2 depicts the Kalman-filtered trend growth rate and cyclical components of the Solow residual for India.
Decomposition of Indian TFP in permanent and transitory components shows that shocks to trend growth are a major source of fluctuations in Indian business cycle. The Kalman filtered estimate of $\sigma_{W2} = 0.32$ provides a measure of transitory shock $\sigma_a$, and the estimate of $\rho_c = 0.76$ gives the degree of persistence in transitory component of TFP. Next, an AR(1) model is fitted to the growth rate of the estimated permanent component of TFP. The persistence in the trend growth $\rho_g$ is found to be 0.27, while the estimate of $\sigma_g$ is 1.59. The value of $\sigma_g$ compared to $\sigma_a$ indicates that the shock to trend growth rate is substantially higher than the transitory shock. These estimates are shown in table 3 along with output and consumption volatilities during the period spanning the TFP series.

**Financial Frictions and Consumption Volatility: Theoretical Framework**

The theoretical literature on finance and macroeconomic volatility explores how financial integration and financial development affect output and consumption volatility through the channel of firms and households (Bernanke and Gertler 1989; Greenwald and Stiglitz 1993; Aghion et al. 2004; Iyigun and Owen 2004;

The early literature predicts that financial development reduces macroeconomic fluctuations (Bernanke and Gertler 1989; Greenwald and Stiglitz 1993). More recent literature suggests that the nature of relationship between financial development and macroeconomic volatility can be nonlinear (Aghion et al. 2004) and may depend on several factors, such as the composition of short-term and long-term investments in the economy (Aghion et al. 2010).

The Model

Consider a closed economy that is populated by a continuum of infinitely lived households and firms, both of measure unity. There exists a fraction $\lambda$ of households with no access to banking or other instruments to save. These consumers, who may be referred to as non-Ricardian households, are liquidity-constrained and unable to save or borrow to smooth consumption. They have no assets and spend all their current disposable labour income on consumption in each period.

Labour supply is inelastic as no labour-leisure choice is made by the representative household. Emerging economies are characterised by large size of informal employment where average hours of work are found to be higher than that in the formal sector employment (Blunch et al. 2001; International Labour Organization 2012). For instance, studies found that informal sector workers worked on average fifteen hours more than their counterparts in the formal sector (Blunch et al. 2001). Hence, in an emerging economy setup, it is reasonable to assume that households allocate their available labour-time to production as much as possible. The representative household is assumed to supply one unit of labour inelastically.

Both Ricardian and liquidity-constrained households have identical preferences defined over a single commodity,

\[ U(C_i) = \ln(C_i), \quad i = R, L, \]  

(7)

2. In India, more than 90% of the workforce and about 50% of the national product are accounted for by the informal economy (Report of the Committee on Unorganised Sector Statistics 2012). According to National Sample Survey Organisation (2004–05), of the total workers, 82% in the rural areas and 72% in the urban areas are engaged in informal sector. In terms of absolute numbers, out of the total 465 million people employed in the formal and informal sectors, only 28 million people (6% of the total employment) are employed in the formal sector, while 437 million workers (94% of the total employment) are in the informal sector (National Sample Survey Organisation 2009–10, (http://labour.gov.in/content/aboutus/about-ministry.php). Data on hours worked are not officially published in India. The officially published employment data captures the employment scenario in the formal sector, which constitutes only 6% of the total employment.
where $C_i$ denotes total consumption of the household of type $i$. Ricardian households are indexed as $R$ and liquidity-constrained households as $L$.

A Ricardian household maximises discounted stream of utility,

$$V_t = E_t \sum_{t=0}^{\infty} \beta^t \log(C_t^R),$$

subject to the following budget constraint,

$$C_t^R + I_t^R = R_t K_t^R + W_t,$$

where $\beta \in (0, 1)$ denotes the subjective discount factor. Here $C_t^R$ is total consumption of the Ricardian household in period $t$. The variables $I_t^R$ and $K_t^R$ denote investment and capital stock of the household, respectively. The economy-wide return to capital and wage rate are given by $R_t$ and $W_t$. In each period, the Ricardian household divides her disposable income, comprised of wage and rental income, into consumption and savings.

The stock of capital of the representative Ricardian household evolves via the following law of motion,

$$K_{t+1}^R = (1-\delta)K_t^R + I_t^R - \frac{\phi}{2} \left( \frac{K_{t+1}^R}{K_t^R} - \mu \right)^2 K_t^R.$$  

The investment is subject to quadratic capital adjustment cost as in Aguiar and Gopinath (2007).

Households who do not have access to financial services cannot save or borrow. Their behaviour is thus different from that of Ricardian consumers. Liquidity-constrained households maximise instantaneous utility $\log C_t^L$ subject to the following budget constraint in each period,

$$C_t^L = W_t,$$

where $C_t^L$ is total consumption of the liquidity-constrained household in period $t$. In each period, a liquidity-constrained household consumes its entire disposable income comprised of wage income.

The aggregate consumption is the weighted average of consumption by the liquidity-constrained households and the Ricardian households. The weights are the share of each type of households in the population.

$$C_t = \lambda C_t^L + (1-\lambda)C_t^R.$$
The aggregate capital stock and investment are, respectively, the following

\[ K_t = (1 - \lambda)K_t^R, \quad I_t = (1 - \lambda)I_t^R, \quad (13) \]

A representative firm produces a homogeneous good, by hiring one unit of labour from households and combining it with capital. The aggregate output is produced by Cobb Douglas technology that uses capital and unit labour as inputs:

\[ Y_t = e^{\alpha t}[(1 - \lambda)K_t^R]^{1-a}T_a^\alpha, \quad (14) \]

where \( \alpha \in (0, 1) \) represents labour’s share of output and \( e^{\alpha t} \) denotes the transitory component of total factor productivity. Here \( T_a \) is the permanent component of productivity. The two productivity processes are characterised by the following stochastic properties: total factor productivity evolves according to an AR(1) process as follows:

\[ a_t = \rho_a a_{t-1} + \epsilon_t^a, \quad (15) \]

with \( |\rho_a| < 1 \) and \( \epsilon_t^a \) represents iid draws from a normal distribution with zero mean and standard deviation \( \sigma_a \).

Following Aguiar and Gopinath (2007), the growth rate of labour productivity \( \Gamma_t \) is defined as

\[ \Gamma_t = g_t \Gamma_{t-1}. \quad (16) \]

The growth rate of labour productivity \( g_t \) follows an AR(1) process of the form:

\[ \ln \left( \frac{g_t}{\mu_g} \right) = \rho_g \ln \left( \frac{g_{t-1}}{\mu_g} \right) + \epsilon_t^g; \quad \epsilon_t^g \sim N(0, \sigma_g^2) \quad (17) \]

The resource constraint of the economy is given by

\[ C_t + I_t = Y_t \quad (18) \]

In a closed economy, total output is allocated between total consumption and investment as indicated by equation (18).

Since the realisation of \( g \) permanently influences \( \Gamma \), output is nonstationary with a stochastic trend. Output, consumption, investment, and capital stock are detrended by normalising these variables with respect to the trend productivity through period \( t-1 \). For any variable \( X \), its detrended counterpart is defined as \( x_t = X_t/\Gamma_{t-1} \).
With the initial capital stock $K_0$, the competitive equilibrium is defined as a set of prices and quantities $(R_t, W_t, y_t, c_t, cR_t, cL_t, i_t, k_t)$, given the sequence of shocks to TFP and labour productivity growth, that solves the maximisation problem of the household, optimisation by the firms, and satisfies the resource constraint of the economy.

**Predictions**

After normalisation of the variables by labour productivity in the previous period, the system of equations driving the dynamics of the model economy become

\[
\begin{align*}
1 &= \beta E_{t-1} \left[ \Omega_t \frac{c_{t-1}^{R}}{c_t^{R} g_t} \right], \\
\Omega_t &= (1 - \alpha)e^{at}(1 - \lambda)^{1-\alpha}(k_t^{R})^{-\alpha}g_t^{a} + (1 - \delta), \\
c_t^{R} &= \frac{(1-\alpha\lambda)}{1-\lambda}e^{at}[(1-\lambda)k_t^{R}]^{-\alpha}g_t^{a} + (1 - \delta)k_t^{R} \\
- g_t k_{t+1}^{R} - (\phi/2) \left( \frac{k_{t+1}^{R} g_t}{k_t^{R}} - \mu_g \right)^{2} k_t^{R}, \\
a_t &= \rho_a a_{t-1} + \varepsilon_t^a, \\
\ln \left( \frac{g_t}{\mu_g} \right) &= \rho_g \ln \left( \frac{g_{t-1}}{\mu_g} \right) + \varepsilon_t^g.
\end{align*}
\]

The first equation in the system of equations (19) describes intertemporal allocation of consumption by the Ricardian consumers where $\Omega_t$ is the gross return to capital. The third equation pertains to the resource constraint of the economy, after taking into account the consumption of liquidity-constrained households as in equation (11), total consumption in equation (12), dynamics of capital accumulation by the Ricardian households in equation (10), stock of capital and investment in the economy given in equation (13), and making use of the fact that $w_t = W_t/G_{t-1} = \alpha e^{at}[(1-\lambda)k_t^{R}]^{-\alpha}g_t^{a}$.

After log-linearising the system of equations (19) and given the total consumption of the economy as in equation (12), and making use of the equation (11) and the fact that $W_t = \alpha Y_t$ implying $c_t^{L} = \bar{y}_t$, one can arrive at the volatility of consumption relative to output as,

\[
\frac{\sigma_{\bar{y}}^2}{\sigma_{\bar{c}}^2} = \left( \frac{\bar{c}^R}{\bar{c}^L} \right)^2 (1 - \lambda)^2 \frac{\sigma_{\bar{x}}^2}{\sigma_{\bar{y}}^2} + \left( \frac{\bar{c}^L}{\bar{c}^L} \right)^2 \lambda^2.
\]

Here the fluctuations in a Ricardian household’s consumption and that in total
output are, respectively,

\[
\sigma^2_c = \left[ \frac{a_1^2 b_1^2}{1 - a_1^2} + b_2^2 \right] \sigma^2_a + \left[ \frac{a_2^2 d_1^2}{1 - a_1^2} + d_2^2 \right] \sigma^2_g,
\]

\[
\sigma^2_y = \left[ 1 + \frac{(1 - \alpha)^2 b_1^2}{1 - a_1^2} \right] \sigma^2_a + \left[ \alpha^2 + \frac{(1 - \alpha)^2 d_1^2}{1 - a_1^2} \right] \sigma^2_g.
\]

The Supplemental Appendix S2 describes the solution method in details.

The effects of transitory and permanent income shocks on the volatility of consumption relative to volatility of output in the economy can be summarised as follows.

**Proposition 1** With everything else remaining unchanged,

(i) Volatility of consumption of a liquidity-constrained household relative to output volatility is always unity, that is, \( \frac{\sigma_c}{\sigma_y} = 1 \), when \( \sigma_{e^*} > 0 \); \( \sigma_{e^*} > 0 \).

(ii) Due to a transitory shock in income, both volatility of consumption of a Ricardian household relative to output volatility and the volatility of total consumption relative to output volatility are lower than one, irrespective of the share of liquidity-constrained households in the population, that is, \( \frac{\sigma_c}{\sigma_y} < 1 \) and \( \frac{\sigma_c}{\sigma_y} < 1 \) for \( \lambda \in [0,1) \), when \( \sigma_{e^*} > 0 \); \( \sigma_{e^*} = 0 \).

(iii) Due to a shock to the trend growth rate, volatility of consumption of a Ricardian household relative to volatility of output always exceeds one, irrespective of the share of liquidity-constrained households in the economy, while the volatility of total consumption relative to output volatility depends on the share of liquidity-constrained households in the economy, that is, \( \frac{\sigma_c}{\sigma_y} > 1 \), and \( \frac{\sigma_c}{\sigma_y} \leq 1 \), for \( \lambda \in [0,1) \), when \( \sigma_{e^*} = 0 \); \( \sigma_{e^*} > 0 \).

(iv) In the presence of shock to the trend growth rate, both volatility of consumption of a Ricardian household relative to output volatility and the volatility of total consumption relative to volatility of output increases when the share of liquidity-constrained households in the economy decreases, that is, \( \frac{\partial (\sigma_c/\sigma_y)}{\partial \lambda} < 0 \), and \( \frac{\partial (\sigma_c/\sigma_y)}{\partial \lambda} < 0 \), for \( \lambda \in [0,1) \), when \( \sigma_{e^*} = 0 \); \( \sigma_{e^*} > 0 \).

The proof of Proposition 1 is presented in the Supplemental Appendix S2 in details.

Liquidity-constrained households who have no access to savings instruments can respond to any change in income by changing consumption by the amount of changed income. Hence volatility of consumption of a liquidity-constrained household relative to output volatility is always one irrespective of the nature of shock.
In response to a transitory income shock, a Ricardian household smooths consumption by re-allocating changed income between consumption and savings. Hence consumption fluctuates by a lesser amount compared to income fluctuation. Hence consumption volatility of a Ricardian household relative to output volatility, in response to a transitory income shock, is always less than one, irrespective of the level of financial development. In this scenario, the relative volatility of total consumption, when total consumption is a weighted average of the relative consumption volatility of a Ricardian household and that of a liquidity-constrained household, is also less than one in all states of financial development.3

Ricardian households perceive a rise in income in the future following a permanent income shock. They respond to it by raising current consumption more than the rise in current income by borrowing against future income or reducing current savings. Thus, relative volatility of consumption of a Ricardian household with respect to output volatility is greater than one. Relative volatility of total consumption, when total consumption is a weighted average of the relative consumption volatility of a Ricardian household and that of a liquidity-constrained household, may be smaller or higher than one depending on the size of $\lambda$.

Financial development reduces the share of liquidity-constrained households in the economy and hence allows more people to respond to the permanent income shock by raising current consumption more than the rise in current income. As a result, volatility of total consumption relative to output volatility increases with financial development.

Combining these observations, the main theoretical prediction of the model can be stated as follows:

**Main prediction:** Other things unchanged, under the occurrence of permanent income shock, financial development leads to a rise in the volatility of consumption in the economy relative to output volatility.4

The main prediction is tested by calibrating the model economy to Indian data. The hypothesis is tested for an emerging economy where relative consumption volatility shows an increase after witnessing of financial sector development.

### Case Study: Evidence for India

The model is calibrated for India, an emerging economy which has witnessed financial sector reform. Ang (2011) finds that financial liberalisation increases fluctuations in consumption in India during 1950–2005. Also, relative to income

3. The weights correspond to a combination of the share of consumption of the respective household type in total consumption and the share of such households in total population.

4. It follows from the implications of the main prediction of the model that in response to a negative permanent income shock, Ricardian households reduce current consumption by more than the decline in current income and raise investment in order to smooth consumption over the lifetime. Financial development will allow more people to respond to the negative income shock by reducing current consumption more than the fall in income. Volatility of total consumption relative to output volatility thus increases with financial development under negative trend growth shocks as well.
volatility, consumption volatility in India increased after reform (Ghate et al. 2013).

India has witnessed development of its domestic financial sector in the post-reform period, while remaining fairly closed in terms of capital account openness even after the reform. Thus India serves as an example of an emerging economy, with a low level of financial integration and a moderate expansion of domestic financial services. Financial development indicators show expansion of financial services in India from the pre- to post-reform periods (figure 3). Interestingly, the country witnessed a small decline in banking services before witnessing a sharp increase. This period is included in the post-reform sample to achieve reasonable sample size.

The model is simulated for the pre- and post-reform periods, keeping all deep parameters, except the share of non-Ricardian households the same for both periods. Expansion of the financial services is captured by a lower value of the share of liquidity-constrained households in the post-reform period. The purpose is to identify one of the key factors which may explain the differences in relative consumption volatility between pre- and post-financial reform periods. The model is simulated for two different values of the share of liquidity-constrained households.

**Figure 3. Financial Development in India**

This figure shows the behaviour of some financial development indicators in India. The upper two panels depict bank deposit to GDP ratio and the private credit to GDP ratio. The left lower panel shows number of bank branches per 100,000 people. The right lower panel shows number of bank accounts per 100,000 people. The density of bank accounts and that of bank branches, bank deposit to GDP ratio, and private credit to GDP are all seen to rise. The dashed lines show the mean values before and after financial reforms.

households and compares the simulated business cycle moments with business cycle stylised facts observed in pre- and post-reform India.

The key business cycle moments for per capita output, consumption, and investment at annual frequency are estimated. Output, consumption, and investment are measured by real GDP at factor cost, private consumption expenditure, and gross fixed capital formation for the period 1951–2010. To examine the transition in the business cycle stylised facts, the sample is divided into pre- (1951–91) and post-reform periods (1992–2010). Key business cycle moments are obtained from the hp-filtered cyclical components of per capita output, consumption, and investment.

The trend in one of the key variables of the present analysis, namely, relative consumption volatility, is depicted in figure 4. The mean of relative consumption volatility shows an increase in the post reform period (figure 4).

The change in business cycle facts for the Indian economy from 1951–2009 are depicted in table 4. Per capita Real GDP has become less volatile in the post-reform period in India. The level of volatility is still high and comparable to emerging economies. The absolute per capita consumption volatility, as well as the relative consumption volatility with respect to output, increased in the post-reform period. Per capita investment volatility show a small decline in the post-reform period, while volatility in investment relative to output volatility has increased following reform. Contemporaneous correlation of consumption and investment with output has increased in the post-reform period. No significant persistence in the output and consumption cycle is seen in the pre-reform period. In the post-reform period, output and consumption cycle are observed to have higher persistence. Persistence in the investment cycle rises in the post-reform period.

There has been a sharp increase in access to finance after reforms. The ratio of bank accounts to total population was merely 20% in 1980; it has jumped...
to above 70% in 2010, except for a period of decline in the trend during 1990–2005. Similarly, bank branches per 100,000 population in 2010 were more than double the value in 1970.

As seen in table 4, relative consumption volatility in India has risen from 0.83 during 1951–91 to 1.04 during 1992–2012. Thus, after improved access to savings instruments and credit, fluctuations in consumption relative to fluctuations in income has increased.

### Calibration

Table 5 summarises the benchmark parameter values used in the calibration exercise. The access of households to banking is captured by the number of bank accounts to population. Hence the proxy for \( \lambda \), that is, the share of liquidity-constrained households is derived from this ratio. The number of bank accounts to population ratios in 1980 and 2010 are used to calibrate the share of liquidity-constrained households in the pre- and post-reform periods. In 1980, 21.4% of the population had access to banking. Thus the share of households without access to finance, that is, \( \lambda \), is set to 0.786 in the pre-reform period. In 2010, 66.9% of the population had access to banking services. The value of \( \lambda \) is thus set to \( 1 - 0.669 = 0.331 \) in the post-reform period.

Some of the other parameter values are chosen based on the existing literature. A period is a year. The share of labour \( \alpha \) for India is 0.7 as in Verma (2008), while the rate of depreciation is 5% as in Virmani (2004).

Next, the annual discount rate is calibrated using annual data of real interest rates for India sourced from the World Bank. The real interest rate series reported in this database is the lending interest rate adjusted for inflation as measured by the GDP deflater. The trend real interest rate is estimated using the Hodrick-Prescott filter. The average value of the trend real interest rate during the sample period of 1980–2012 is \( \bar{R} = 6.16\% \). The Euler equation in steady state becomes \( \mu_g = \beta(1 + \bar{R}) \), where \( \mu_g - 1 \) is the average trend growth of productivity process.

### Table 4. Business Cycle Stylised Facts for the Indian Economy in the Pre- and Post-Reform Period

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Real GDP</td>
<td>2.25</td>
<td>1.00</td>
</tr>
<tr>
<td>Pvt. Cons.</td>
<td>1.86</td>
<td>0.83</td>
</tr>
<tr>
<td>Investment</td>
<td>5.26</td>
<td>2.34</td>
</tr>
</tbody>
</table>

*Source: National Accounts Statistics, Labour Bureau, authors’ estimates outlined in the Case Study section.*

This table reports the changes in business cycle facts for the Indian economy from the pre-reform to the post-reform periods. The span of the analysis is 1951–2009.
Table 5. Benchmark Parameter Values

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discount factor</td>
<td>$\beta$ 0.968</td>
</tr>
<tr>
<td>Rate of Depreciation</td>
<td>$\delta$ 5.000</td>
</tr>
<tr>
<td>Share of labour</td>
<td>$\alpha$ 0.700</td>
</tr>
<tr>
<td>Adjustment cost parameter</td>
<td>$\phi$ 2.820</td>
</tr>
<tr>
<td>Mean trend growth rate of labour productivity</td>
<td>$\mu_g - 1$ 2.790</td>
</tr>
<tr>
<td>Persistence in transitory component of technology</td>
<td>$\rho_t$ 0.760</td>
</tr>
<tr>
<td>Volatility in transitory component of technology</td>
<td>$\sigma_t$ 0.320</td>
</tr>
<tr>
<td>Persistence in growth of permanent component of technology</td>
<td>$\rho_g$ 0.266</td>
</tr>
<tr>
<td>Volatility of shock to permanent component of technology</td>
<td>$\sigma_g$ 1.590</td>
</tr>
</tbody>
</table>

Source: Virmani (2004), Verma (2008), Aguiar and Gopinath (2007), and authors’ estimates outlined in the Consumption Volatility and Permanent versus Transitory Income Shocks section and in the Case Study section.

This table summarises the parameter values used for the calibration exercise. Rate of depreciation, mean trend growth rate, and volatilities of trend growth rate and transitory component of TFP are in percentage (%).

and $\beta$ is the annual discount factor. The value of $\mu_g - 1$ is obtained from Kalman filtration of Solow residual series for India.5 The estimated value of $\mu_g - 1$ is 2.79%. It then follows from the Euler equation that the annual discount factor for India is $\beta = \mu_g/(1 + \bar{R}) = 1.0279/1.0616 = 0.968$.

The estimated shock processes in the transitory and the growth rate of permanent components of Solow residual for India are sourced from table 3. The parameter for capital adjustment cost $\phi$ is set to 2.82 from Aguiar and Gopinath (2007).

Effect of Financial Development on Relative Consumption Volatility

The model predicts that a decline in the share of liquidity-constrained households in the population would allow more people to respond to permanent income shocks. They can increase current consumption more than the rise in current income. This is predicted to result in a rise in the relative consumption volatility.

Main findings are the following. The relative consumption volatility shows a rise in the post-reform period (table 6). This result supports the key prediction of the model. Since financial development allows more people to access savings instruments, when households perceive a permanent income shock which raises both current and future income, more people can respond to the shock by reducing current savings and raising current consumption more than the rise in current income. As a result of financial development, the volatility of consumption relative to volatility of output rises.

This model also replicates the pattern of changes in absolute consumption volatility successfully. The model also captures a decline in the absolute output

5. The details of the estimation procedure and results are outlined in the Consumption Volatility and Permanent versus Transitory Income Shocks section.
volatility in the post-reform period as observed in the data. However, in terms of magnitude, the change in the output volatility is not substantial. With financial inclusion, more people can save, and, hence, investment volatility declines. The model shows a fall in the absolute volatility in investment in the post-reform period, as observed empirically. However, unlike the trend shown in the data, the simulated relative investment volatility declines in the post-reform period.

Next, the simulated correlation of consumption and investment cycles with the output cycle and their persistence with the empirical counterparts are compared in (table 7). The model shows a rise in the correlation of investment with output, as in the data. However, the magnitude of the rise is small compared to the trend shown by the data. The simulated correlation of consumption cycle with the output cycle shows a marginal decline after reform.

The pattern of model simulated persistence in output and consumption cycles matches broadly with the pattern observed in the data. However, the performance of the model is not satisfactory in terms of matching the persistence in the investment cycle. Finally, the model is found to replicate the cyclical pattern in output, consumption, and investment fairly well (figure 5).

### Sensitivity to the Measure of Financial Development

In the above analysis, the financial development is measured by the share of the population with bank accounts. As a robustness check, another measure of financial development, namely, the bank deposit to GDP ratio is used to obtain the fraction of liquidity-constrained households in the economy. By this measure, $\lambda$ is 0.687 in the pre-reform period. The value of $\lambda$ in the post-reform period is 0.305.

The key moments from the business cycle model for the pre- and post-reform periods based on this alternative measure of $\lambda$ are similar to those of the benchmark model (table 8 and 9).
Along with domestic financial deepening, opening up of the capital account, or financial liberalisation, has been a major component of the spectrum of reforms in emerging economies in the last two decades. This section explores the implications of financial deepening for the aggregate consumption fluctuations in an open economy framework.

It is assumed that financial transactions by Ricardian households take place through an internationally traded, one-period, risk-free bond as in Aguiar and Gopinath (2007). The budget constraint of the Ricardian households is modified for the open economy framework as

\[ CR_t + IR_t + BR_t / C_0 = RK_t KR_t + W_t. \]  

Here, the level of debt due in period \( t \) held by a Ricardian household is denoted by \( BR_t \) and \( R_t \) is the time \( t \) interest rate payable for the debt due in period \( t + 1 \). The economy-wide return to physical capital and wage rate are given by \( RK_t \) and \( W_t \), respectively. Access to international financial markets is assumed to be imperfect. The interest rate is subject to a premium associated to the riskiness of investing in emerging economies. This premium depends on the level of outstanding debt, taking the form used in Schmitt-Grohe and Uribe (2003),

\[ R_t = R^* + \psi \left( e^{\beta_{t+1} - b_t} - 1 \right). \]
Here the variable $R^*$ is the world interest rate exogenously given to the small open home country. The variable $\bar{b}$ denotes the steady state level of total debt, and $\psi$ ($\psi > 0$) is the elasticity of interest rate to changes in the indebtedness of the economy. The total debt of the economy $B_t$ is exogenously given to the representative agent who does not internalise the premium payable on the foreign interest rate determined by the indebtedness of the economy. However, in equilibrium, total foreign debt of the economy coincides with the amount of debt acquired by all the representative agents of the Ricardian type. Given the fraction of Ricardian households in the economy equal to $1 - \lambda$, the total debt in the economy amounts to $B_t = (1 - \lambda)B_t^R$, while the long run total debt is $\bar{b} = (1 - \lambda)\bar{b}^R$.

The resource constraint equation for the open economy is modified as follows:

$$C_t + I_t + TB_t = Y_t,$$

(23)

This figure compares cyclical movements in per capita GDP, consumption expenditure and investment with simulated output, and consumption and investment cycles for the pre- and post-reform periods. The left panel shows key macroeconomic cycles in the pre-reform period, whereas the right panel depicts post-reform cyclical fluctuations in the macroeconomic indicators.

Source: Authors’ estimates outlined in the Case Study section.
Table 8. Sensitivity Analysis with Respect to the Financial Development Parameter

<table>
<thead>
<tr>
<th></th>
<th>Std. dev.</th>
<th>Rel. std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Y</td>
<td>C</td>
</tr>
<tr>
<td>Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-reform</td>
<td>2.25</td>
<td>1.86</td>
</tr>
<tr>
<td>Post-reform</td>
<td>1.93</td>
<td>1.99</td>
</tr>
<tr>
<td>Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-reform</td>
<td>1.92</td>
<td>2.00</td>
</tr>
<tr>
<td>Post-reform</td>
<td>1.91</td>
<td>2.18</td>
</tr>
</tbody>
</table>

Source: Authors’ analysis outlined in the Case Study section.

This table presents business cycle moments from the simulated model for the pre- and post-reform period using an alternative measure of \( \lambda \). The measure used in this analysis is based on the deposit to GDP ratio. The absolute standard deviation numbers are in percentage (%). The relative standard deviations are in ratio. The patterns of transition of business cycle moments broadly resemble the benchmark analysis.

Table 9. Sensitivity Analysis with Respect to the Financial Development Parameter

<table>
<thead>
<tr>
<th></th>
<th>Correlation</th>
<th>Auto-correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C</td>
<td>I</td>
</tr>
<tr>
<td>Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-reform</td>
<td>0.70</td>
<td>0.19</td>
</tr>
<tr>
<td>Post-reform</td>
<td>0.92</td>
<td>0.76</td>
</tr>
<tr>
<td>Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-reform</td>
<td>0.99</td>
<td>0.23</td>
</tr>
<tr>
<td>Post-reform</td>
<td>0.96</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Source: Authors’ analysis outlined in the Case Study section.

This table shows that business cycle moments from the simulated model for the pre- and post-reform period using the alternative measure of \( \lambda \) based on deposit to GDP ratio. The patterns of transition of the moments broadly resemble the patterns from benchmark analysis.

where the trade balance \( TB_t \) is financed by the net flows of capital,

\[
TB_t = B_t - \frac{B_{t+1}}{1 + R_t}.
\] (24)

In an economy which is open on both trade and financial fronts, imports and total domestic output net of exports is allocated between total consumption and investment, where the difference between exports and imports are balanced by the financial flows as indicated by equations (23) and (24). The rest of the framework, such as the optimisation problem of the Ricardian and the liquidity-constrained households, firm’s profit maximisation behaviour, and the permanent and transitory
shock structures remain similar, as in the closed economy framework. By normalising the variables with respect to the permanent component of productivity at period $t-1$, the detrended system of equations are obtained. The Supplemental Appendix S3 contains the detrended system of equations pertaining to the open economy.

**Calibration to Indian Data**

In order to calibrate the open economy, value of the interest rate elasticity of indebtedness is set to 0.001, as in Aguiar and Gopinath (2007). The steady state level of debt to GDP ratios for the pre- and post-reform periods are set to the average values of the external debt to GDP ratios in 1971–91 and 1992–2012, respectively. The respective values are 16.30% and 21.39%.

The value of the risk-free world interest rate is set to satisfy the condition that $\beta(1+R^*) = \mu_g$, where $\mu_g - 1$ is the mean growth rate of the permanent component of TFP. The value of this parameter is set to 2.79% based on the estimated permanent component of TFP as outlined in the Consumption Volatility and Permanent versus Transitory Income Shocks section. The rest of the parameter values remain the same, as in the closed economy case.

Data show, in addition to business cycle stylised facts with respect to the key macroeconomic indicators in India (table 4), more than one-and-a-half times increase in the mean net exports to GDP ratio from pre- to the post-reform period in India (table 10). The business cycle volatilities, both absolute and relative, in trade balance to GDP ratio have also increased in the post-reform period. The trade balance to GDP ratio has become strongly counter cyclical after the reform, from being merely acyclical in the pre-reform period (table 10).

The empirical and simulated business cycle moments for the open economy in the pre- and post-reform periods are compared in tables 11 and 12. The open economy version of the model is able to replicate most of the patterns in the changes in stylised facts from the pre- to post-reform periods in India. As observed in the data, the model-simulated absolute volatilities in consumption and trade balance to GDP ratio have increased in the post-reform period, while that of investment has decreased. However, unlike in the data, the volatility of output in the model shows a rise in the post-reform period and the absolute volatility in the trade balance to GDP ratio exceeds output volatility.

So far as the relative volatilities are concerned, volatilities in consumption and trade balance to GDP ratio, relative to output volatility rise, reflecting trends observed in the data. However, unlike the pattern observed empirically, the relative volatility of investment falls. The relative volatility of investment resembles the pattern observed in the closed economy framework.

The model-simulated correlation of investment with output increases after the reform, although the model is not able to capture the sharp rise in the correlation

---

6. The annual series of external debt are sourced from WDI. The data spans from 1971–2012 and are in current US$. The GDP data, also in current US$, are sourced from WDI.
Table 10. Stylised Facts on Trade Balance to GDP Ratio in India in the Pre- and Post-Reform Period

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.99</td>
<td>3.48</td>
</tr>
<tr>
<td>Std. dev.</td>
<td>0.90</td>
<td>1.16</td>
</tr>
<tr>
<td>Rel. std. dev.</td>
<td>0.40</td>
<td>0.60</td>
</tr>
<tr>
<td>Cont. cor.</td>
<td>0.25</td>
<td>−0.69</td>
</tr>
<tr>
<td>First ord. auto. corr.</td>
<td>0.246</td>
<td>0.504</td>
</tr>
</tbody>
</table>


This table presents business cycle moments and the average value of the trade balance to GDP ratio for the pre- and post-reform periods.

Table 11. Simulated Business Cycle Volatilities from the Open Economy Model

<table>
<thead>
<tr>
<th></th>
<th>Std. dev.</th>
<th>Rel. std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Y</td>
<td>C</td>
</tr>
<tr>
<td>Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-reform</td>
<td>2.17</td>
<td>1.86</td>
</tr>
<tr>
<td>Post-reform</td>
<td>1.94</td>
<td>1.99</td>
</tr>
<tr>
<td>Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-reform</td>
<td>1.48</td>
<td>2.14</td>
</tr>
<tr>
<td>Post-reform</td>
<td>1.51</td>
<td>2.94</td>
</tr>
</tbody>
</table>

Source: Authors’ analysis outlined in the Financial Development, Permanent Income Shock, and Relative Consumption Volatility in a Small Open Economy section.

This table compares absolute and relative business cycle volatilities from the simulated model for the pre- and post-reform period with the pattern observed in the data. The volatilities are in percentage (%).

Table 12. Simulated Business Cycle Correlation and Persistence from the Open Economy Model

<table>
<thead>
<tr>
<th></th>
<th>Correlation</th>
<th>Auto-correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C</td>
<td>I</td>
</tr>
<tr>
<td>Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-reform</td>
<td>0.71</td>
<td>0.19</td>
</tr>
<tr>
<td>Post-reform</td>
<td>0.83</td>
<td>0.76</td>
</tr>
<tr>
<td>Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-reform</td>
<td>0.80</td>
<td>0.20</td>
</tr>
<tr>
<td>Post-reform</td>
<td>0.72</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Source: Authors’ analysis outlined in the Financial Development, Permanent Income Shock, and Relative Consumption Volatility in a Small Open Economy section.

This table compares business cycle correlation of various macroeconomic indicators with output cycle and persistence from the simulated model for the pre- and post-reform periods with the patterns observed in the data.
as observed in the data. The data shows that the correlation of trade balance to GDP ratio turns from acyclical to strongly counter-cyclical. Although the model shows that trade balance to GDP ratio has a negative correlation with output, and the magnitude of the correlation increases in the post-reform period, but it does not become strongly countercyclical after the reform. The correlation of consumption with output declines, whereas it increases in the data after the reform.

Discussion of the Results

The open economy framework, when calibrated to Indian data, supports the main prediction of rising relative consumption volatility with financial inclusion. Broadly, the model-simulated moments show similar patterns observed in the closed economic framework, except for a marginal rise in the output volatility in the post reform period.

One plausible reason for the open economy setup to show similar trends in the volatility and correlation of the key macroeconomic indicators, as in the closed economy scenario, is that financial deepening, in the present model, works through the household channel. Under strong permanent income shock, relative to transitory income fluctuations, Ricardian households behave in a similar manner in both closed and open economy setups. However, the extent of fluctuations is higher in an open economy. In response to permanent income shock, in an open economy, households can even raise current consumption more by using funds borrowed against future income. Hence fluctuation in consumption is even higher than the closed economy scenario. Financial inclusion, in this setup results in larger fluctuations in aggregate consumption. A sharp rise in consumption volatility with a relatively smaller decline in investment volatility causes a marginal rise in post-reform output fluctuations. Hence, the open and closed economy setups show qualitatively similar results.

In this open economy framework, consumers transact an internationally traded bond, which is the source of capital flows in the economy. A bulk of literature has explored macroeconomic effects of the interaction between financial openness and domestic financial development through firm borrowing channel (Aghion et al. 2004, 2010). Incorporating borrowing by firm in the model may provide an additional channel for the interaction between financial development and financial liberalisation to affect output and investment. However, in spite of the fact that India started liberalising capital account in 1991, the pace and the extent of easing restrictions on capital flows remained low compared to other emerging economies. The access to foreign capital by Indian households and firms are still limited due to a wide array of capital control measures existing in the country. The de jure measure of capital account openness based on the Chinn-Ito index shows that India is relatively closed compared to other large emerging economies (Patnaik and Shah 2012) (see figure 6). Households in India are not allowed to borrow abroad. There are a number of restrictions on foreign borrowing by firms, and both macro and firm level data indicate low exposure of
Indian firms to foreign capital. Given the low level of access to foreign capital by Indian households and firms, an open economy setup through the financial channel may not be appropriate to replicate the post-reform business cycle stylised facts in India.

India liberalised current account at a faster pace than capital account. Explicitly modelling the current account incorporating home and foreign goods in consumption and investment, as in Mendoza (1995) and Kose and Yi (2006) would provide an additional channel of trade liberalisation to affect macroeconomic volatility and cyclicality of various indicators with output.

---

7. Along with domestic financial deepening, opening up of the capital account, or financial liberalization, has been a major component of reforms in India since 1991. However, the access to foreign capital by Indian households and firms have remained limited. Households and banks in India are not allowed to borrow abroad. As far as borrowing by firms are concerned, Indian firms access foreign capital through two channels to leverage their operations. These are Foreign Direct Investment (FDI) and foreign borrowings. FDI in India (net inflows) has grown from USD 0.59 billion in 1993–94 to USD 30.76 billion in 2013–14 (Economic Outlook, Centre for Monitoring Indian Economy). However, the net FDI inflows in India accounts for only 1.78% of GDP in 2013–14. The share of net FDI inflows in India in total investment amounts to 5.24% in 2013–14. To compare with other emerging economies, for instance, net FDI inflows in Brazil in 2013 has been USD 80.84 billion, which is more than double the FDI inflows in India, while the net FDI inflows in China in 2013 has been USD 347.85, which is more than eleven times larger than the FDI flows in India (World Development Indicators). Looking deep into the firm-level database, only 623 firms are found to have foreign promoter (ownership) in a base of 26,725 companies at the end of 31st March, 2014 (Prowess, Centre for Monitoring Indian Economy). India holds stock under foreign borrowings of USD 53.92 billion in 2012–13 and 2013–14. The net inflow of foreign borrowings has accounted for only 0.63% of GDP in 2013–14. Again in a sample of 26,725 firms in the Prowess database, only a total of 642 companies are found to have had foreign borrowings over the years, while only 464 companies have executed for the financial year 2013–14.
Emerging economies have been seen to witness an increase in consumption volatility relative to output volatility after financial development. This behaviour appears puzzling since traditional models and evidence from advanced economies suggest that consumption should become smoother with increase in the access to financial services.

A distinguishing feature of developing economies is that a large share of the population does not have access to finance. In the last two decades, these economies have experienced reforms in the financial sector giving greater access to financial services for households and firms. Yet, these economies experienced an increase in consumption volatility relative to output volatility in the post-reform period. This paper addresses this empirical puzzle. This puzzle can be explained in a model featuring credit constraints and shocks to trend growth of productivity. The model predicts that relative consumption volatility will rise when more consumers can smooth consumption.

The model, when simulated for India before and after an increase in financial development, broadly replicates the rise in relative consumption volatility, as observed in the data. Most of the other empirical regularities observed in the data are also replicated by this model.

The benchmark model represents a closed economy, and the concept of financial development is limited to household’s access to financial services. The model assumes that the household sector is the sole channel for the financial development to work. This is one plausible reason for the model’s weak performance in replicating the business cycle patterns with respect to investment. By including credit-constrained firms in this framework, one can examine the role of financial development further. Extending the model with borrowings by firms will help in understanding how increase in households’ access to finance affects consumption-smoothing behaviour when production and demand for resources are subject to firm’s access to finance.

Finally, the open economy framework, following Aguiar and Gopinath (2007), assumes that consumers transact an internationally traded bond, which is the source of capital flows in the economy. A bulk of literature has explored macroeconomic effects of the interaction between financial openness and domestic financial development through the firm borrowing channel (Aghion et al. 2004, 2010). However, a wide array of capital control measures existing in India (Patnaik and Shah 2012) restricts access of Indian households and firms to foreign capital. Again, India liberalised current accounts at a faster pace than capital accounts. Hence an open economy framework, capturing trade liberalisation following Mendoza (1995) and Kose and Yi (2006), may help in improving the fit of the model in the open economy framework.

Further, differentiating between agricultural and nonagricultural goods in the consumption basket may help to capture the effects of structural shifts away from agriculture to nonagriculture on the post-reform stylised facts.
SUPPLEMENTARY MATERIAL

The supplemental appendices to this article are available at http://wber.oxfordjournals.org

CONFLICT OF INTEREST

None declared.

REFERENCES


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

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


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

Forthcoming papers in

- **Global Income Distribution: From the Fall of the Berlin Wall to the Great Recession**
  Christoph Lakner and Branko Milanovic

  Philip Keefer and Stuti Khemani

- **The Whole is Greater than the Sum of Its Parts: Complementary Reforms to Address Microeconomic Distortions**
  Raphael Bergoeing, Norman V. Loayza, and Facundo Piguillem

- **Risky Business: Political Instability and Sectoral Greenfield Foreign Direct Investment in the Arab World**
  Martijn Burger, Elena Ianchovichina, and Bob Rijkers

**MINI-SYMPOSIUM “DOING BUSINESS”**

- **The Impact of Business Environment Reforms on New Registrations of Limited Liability Companies**
  Leora Klapper and Inessa Love

- **Deals and Delays: Firm-level Evidence on Corruption and Policy Implementation Times**
  Caroline Freund, Mary Hallward-Driemeier, and Bob Rijkers

- **The Role of Regulation on Entry: Evidence from the Italian Provinces**
  Francesco Bripi
Learning Dynamics and Support for Economic Reforms: Why Good News Can Be Bad  
Sweder J. G. van Winbergen and Tim Willems
A Helping Hand or the Long Arm of the Law? Experimental Evidence on What Governments Can Do to Formalize Firms  
Gustavo Henrique de Andrade, Maram Bruhn, and David McKenzie
Economic Shocks and Subjective Well-Being: Evidence from a Quasi-Experiment  
Jacob Gerner Hariri, Christian Bjørnskov, and Mogens K. Justesen
Does Access to Foreign Markets Shape Internal Migration? Evidence from Brazil  
Laura Hering and Rodrigo Paillacar
The Decision to Invest in Child Quality over Quantity: Household Size and Household Investment in Education in Vietnam  
Hai-Anh H. Dang and F. Halsey Rogers
The Impact of Vocational Schooling on Human Capital Development in Developing Countries: Evidence from China  
Prashant Loyalka, Xiaoting Huang, Linxiu Zhang, Jianguo Wei, Hongmei Yi, Yingquan Song, Taojiang Shi, and James Chu
Financial Inclusion, Productivity Shocks, and Consumption Volatility in Emerging Economies  
Rudram Bhattacharya and Ila Patnaik