ISSN 0258-6770 (print) ISSN 1564-698X (online)

THE WORLD BANK ECONOMIC REVIEW

Volume 23 • 2009 • Number 3

A Symposium On Gender, Poverty and Demography

Gender, Poverty and Demography: An Overview
Mayra Buvinic, Monica Das Gupta, and Ursula Casabonne

Development, Modernization, and Childbearing: The Role of Family Sex Composition
Deon Filmer, Jed Friedman, and Norbert Schady

The Consequences of the "Missing Girls" of China
Avraham Y. Ebenstein and Ethan Jennings Sharygin

The Gender and Intergenerational Consequences of the Demographic Dividend: An Assessment of the Micro- and Macrolinkages between the Demographic Transition and Economic Development
T. Paul Schultz

Macroeconomic Stability and the Distribution of Growth Rates
Vatcharin Sirimaneetham and Jonathan R.W. Temple

The Effect of Male Migration on Employment Patterns of Women in Nepal
Michael Lokshin and Elena Glinskaya

Political Accountability and Regulatory Performance in Infrastructure Industries: An Empirical Analysis
Farid Gasmi, Paul Noumba Um, and Laura Recuero Virto

www.wber.oxfordjournals.org

OXFORD JOURNALS
OXFORD UNIVERSITY PRESS

Volume 23 • Number 3 • 2009

THE WORLD BANK ECONOMIC REVIEW

Pages 347–531

OXFORD

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

Annual Subscription Rate (Volume 23, 3 Issues, 2009): *Institutions*—Print edition and site-wide online access: £135/$202/€202, Print edition only: £128/$192/€192, Site-wide online access only: £128/$192/€192; *Corporate*—Print edition and site-wide online access: US$273/£177/€273, Print edition only: £192/$288/€288, Site-wide online access only: £192/$288/€288; *Personal*—Print edition and individual online access: £41/$61/€61. US$ rate applies to US & Canada, Euros€ applies to Europe, UK£ applies to UK and Rest of World. There may be other subscription rates available; for a complete listing, please visit www.wber.oxfordjournals.org/subscriptions. *Readers with mailing addresses in non-OECD countries and in socialist economies in transition are eligible to receive complimentary subscriptions on request by writing to the UK address below.*

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

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

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

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

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

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

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

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

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

Typeset by Techset Composition Limited, Chennai, India; Printed by Edwards Brothers Incorporated, USA.

Gender, Poverty and Demography: An Overview

Mayra Buvinic, Monica Das Gupta, and Ursula Casabonne

Much has been written on gender inequality and how it affects fertility and mortality outcomes as well as economic outcomes. What is not well understood is the role of gender inequality, embedded in the behavior of the family, the market, and society, in mediating the impact of demographic processes on economic outcomes. This article reviews the empirical evidence on the possible economic impacts of gender inequalities that work by exacerbating demographic stresses associated with different demographic scenarios and reducing the prospects of gains when demographic conditions improve. It defines four demographic scenarios and discusses which public policies are more effective in each scenario in reducing the constraints that gender inequality imposes on poverty reduction. JEL codes: J10, J13, J16, J18

There has been renewed interest in the links between demographic change and economic outcomes. This interest has focused primarily on the window of opportunity for accelerating economic growth presented by the increasing share of adults in the population relative to children and the elderly. But it is well known that a larger set of demographic processes can influence the prospects for poverty reduction and economic growth.

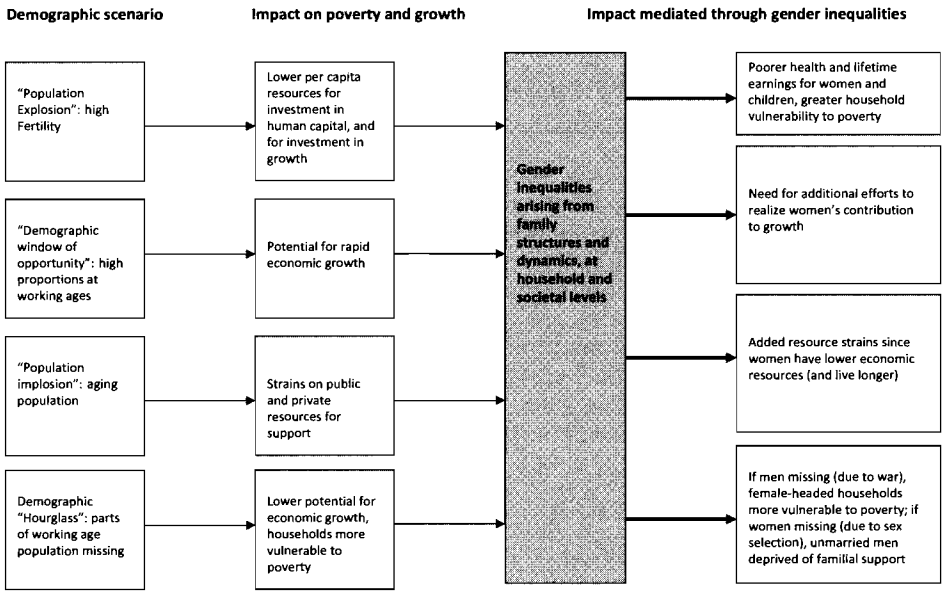
Much has also been written on how gender inequality affects fertility and mortality outcomes as well as economic outcomes. What is not well understood is the role of gender inequality in mediating the impact of demographic processes on economic outcomes.

This overview examines the impact of gender inequality on poverty through the prism of four dominant demographic conditions (figure 1). Gender equality does not necessarily mean equality of outcomes for men and women but rather equality of opportunity (and the ability to make choices) in the family, market, and society (World Bank 2007). Girls and women are typically more affected by inequalities in opportunities—which have serious (and often overlooked)

Mayra Buvinic (corresponding author) is the director of the Gender and Development Group at the World Bank; her email address is mbuvinic@worldbank.org. Monica Das Gupta is a senior social scientist in the Development Research Group at the World Bank; her email address is mdasgupta@worldbank.org. Ursula Casabonne is an operational analyst of the Gender and Development Group; her email address is ucasabonne@worldbank.org. The authors wish to acknowledge literature reviews by Sharon Ghuman and Sarah Hayford.

THE WORLD BANK ECONOMIC REVIEW, VOL. 23, NO. 3, pp. 347–369 doi:10.1093/wber/lhp013
Advance Access Publication November 9, 2009

FIGURE 1. Flowchart of Demography, Poverty, and Gender Relationships



Source: Authors' analysis.

implications for the perpetuation of poverty—but the impact of gender inequalities on men's well-being are also discussed. As there are few data on differences in opportunity, much of the discussion is necessarily based on studies of gender differentials in outcomes as a proxy measure of opportunity.

The evidence on these relationships varies in quantity and quality. An extensive literature establishes associations between variables but does not prove causality. This overview flags these analytical differences and focuses primarily on the relationship between gender inequality and poverty outcomes, which is backed by a growing body of micro studies using increasingly rigorous methods. Less attention is given to the relationship between gender inequality and economic growth, which is more difficult to measure and for which the evidence is much weaker (see Schultz 2009, in this issue).

The four demographic settings are as follows:

- "Demographic explosion"—countries with high fertility and declining child mortality, leading to high youth dependency ratios; lower per capita resources for investments in human capital, infrastructure, and economic growth; and greater difficulty for households to emerge from poverty.
- "Demographic window of opportunity"—countries where declining fertility has led to a high proportion of working-age adults in the population relative to children and the elderly, offering the prospect of

increasing savings and economic growth if appropriate policies are put in place to use the expanded labor force productively.

- "Demographic implosion"—countries where the population is aging rapidly as a result of continuing low fertility and declining adult mortality, straining public and private resources for supporting the elderly, a disproportionate number of whom are women since women typically outlive men.
- "Demographic hourglass"—countries where the prime working-age population is declining because of premature adult mortality (from armed conflict or disease, such as HIV/AIDS), raising dependency ratios, increasing households' vulnerability to poverty, and reducing the potential for economic growth. This is the reverse of the demographic window of opportunity. While countries progress over time from the first to the third scenario, this hourglass effect can occur under any of the other scenarios if there is an unusual mortality spike.

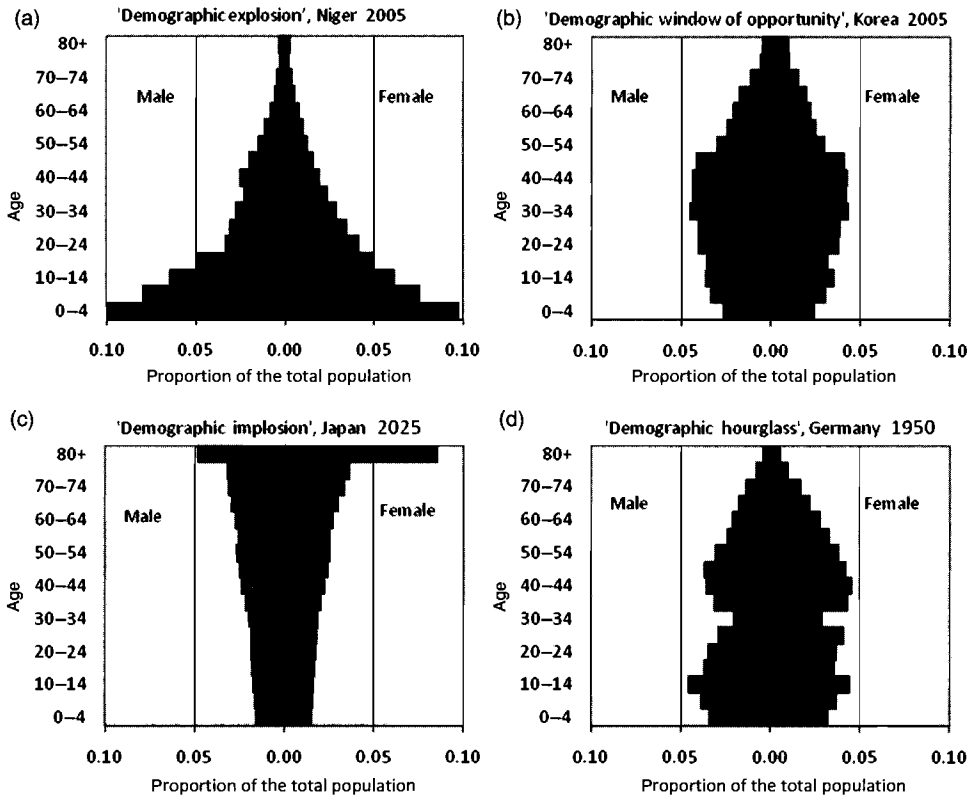
These four demographic settings are illustrated with examples in figure 2.

Gender inequalities exacerbate demographic stresses and limit potential gains when demographic conditions improve.[1] In high-fertility settings, gender inequalities slow fertility decline, negatively affecting women's health and lifetime earnings and reducing prospects for income growth for current and future generations. In the second scenario, gender inequalities restrict women's participation in productive employment and thereby lower the potential economic growth dividend. In rapidly aging populations, the strains of supporting the elderly are exacerbated by gender inequality in access to productive assets and employment. But where gender inequalities are so grave as to result in significant numbers of "missing" women due to sex selection, men may be deprived of familial support in their old age. And under the hourglass scenario, gender inequalities increase the vulnerability of children and the elderly to poverty if men die, because women are less well placed to support their dependents single handedly—though this can be partly offset by an increase in potential economic niches left by the men who die.

Map 1 identifies countries by their dominant demographic scenario. The most demographically stressed countries are those that still have high fertility but are also subject to the hourglass scenario as a result of high HIV prevalence rates or armed conflict. This combination of demographic stresses severely constrains these countries' efforts to reduce poverty and increase growth.

1. In turn, changing demographic dynamics can affect gender inequality. For example, it can affect spousal dynamics. Since men typically marry women younger than themselves, the age gap can increase with high fertility and the resultant growth in size of successive births cohorts. This gives husbands a potential edge over their wives, which diminishes when falling fertility reduces the scope for finding a wife much younger than oneself. Little is known about how this affects other outcomes, such as investment in human capital.

FIGURE 2. Estimated Population Distributions by Age and Sex, Illustrating Different Demographic Settings



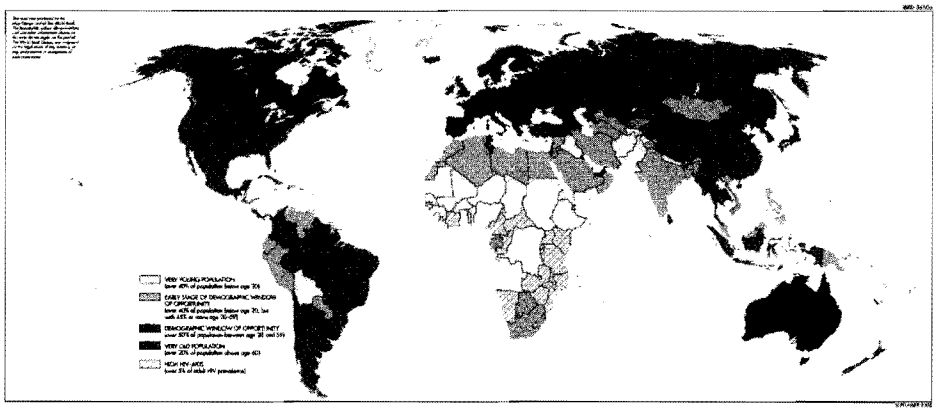
Note: Data refer to the medium variant assumption of fertility trends.
Source: Authors' analysis based on data from United Nations Population Division (2006).

The article is organized as follows. The first four sections synthesize the main findings to date for the four demographic settings. Section V presents some suggestions for research and policy.

I. "DEMOGRAPHIC EXPLOSION": HIGH FERTILITY AND RAPID POPULATION GROWTH

It is an arithmetic truth that rapid population growth strains resources at both societal and household levels. Rapid population growth (see figure 2a) reduces available resources per capita for public investment in services such as health and schooling, as well as for investment in growth through expansion of infrastructure and employment opportunities. These reductions in investment per capita take a toll on prospects for economic growth and poverty reduction, though in the very long sweep of history higher population density can be

MAP 1. Map of Demographic Settings, 2005



Note: Data refer to the medium variant assumption of fertility trends.
Source: Authors' analysis based on data from United Nations Population Division (2006).

associated with more rapid technological change (Kremer 1993). At the household level, having many children puts pressure on the household budget and can be associated with poorer human capital outcomes through many channels, to the detriment of children, women, and the household economy. The evidence is discussed below.

Effects on Child Health and Well-being

Parental investment in children may be diluted as the number of children increases. Poorer families, in particular, may have trouble feeding and schooling their children, and the difficulty may increase with family size. Several studies have examined this quantity–quality tradeoff, controlling for the potential endogeneity between family size and child outcomes.

Evidence of this tradeoff is found in the developing world, for example in Indonesia, where weight for age is lower in later-born children (Henderson and others 2008). In a review of studies from around the world, Behrman, Alderman, and Hoddinott (2004) find that low birth weight is associated with stunting, poorer cognitive development, and lower adult productivity. In the developed world, where parental incomes are relatively high, less dramatic effects are found in some studies. In the United States, children's likelihood of attending private school falls with increasing number of siblings (Conley and Glauber 2006). In Norway, later born children have been found to have lower IQs (Black, Devereux, and Slavanes 2007). Data from the United Kingdom are also suggestive of the disadvantages faced by low-birth-weight children—children born prematurely were shorter as adults, with lower earnings, and were less likely to have professional or managerial jobs (Strauss 2000). Other studies

in developed countries find no effect of greater numbers of children, for example, in Israel (Angrist, Lavy, and Schlosser 2005).

Gender adds a further twist to this story of resource dilution, since there is a strong preference for sons in many developing economy settings. In South and East Asia, parental son preference is strong and results in significant "culling" of daughters to keep family size down (Chung and Das Gupta 2007), thereby reducing the resource dilution caused by continuing to bear children until the desired number of sons is achieved. This pattern of excess female child mortality is not found in Sub-Saharan Africa (Garenne 2003). Filmer, Friedman, and Schady's (2009) article in this issue shows that parents are more likely to stop bearing children if they have a son, so girls tend to have more siblings. This means that son preference, combined with incomplete culling of girls, leads to larger family size. This effect is strongest in South Asia, followed by Central Asia and the Middle East and North Africa.

Effects on Women's Health, Labor Force Participation, and Earnings

Women pay a high price for high fertility in maternal mortality, a major cause of death for young women in high-fertility settings (WHO 2007). Moreover, women's mortality risk remains elevated long after childbirth: a study in Bangladesh found that the risk is nearly twice as high as normal for up to two years after childbirth (Menken, Duffy, and Kuhn 2003). The risk is even higher for poor women, who have less access to quality care during pregnancy and childbirth (Bloom, Wypij, and Das Gupta 2001).

Childbearing can also take a toll on women's labor force participation, productivity, and earnings. In Bangladesh, Joshi and Schultz (2007) find that lower fertility was associated with a rise in women's earnings, better maternal and child health, and higher schooling for the sons in the family. Studies in settings as diverse as Brazil and the Philippines indicate that childbearing is associated with lower women's labor force participation (Connelly and others 2006; Adair and others 2002). However, other studies find only a weak correlation between declining fertility and women's labor supply. This may be due to the intervening role of family structures and dynamics, including fosterage, that affect the allocation of labor in the household and the compatibility between paid work and child care. In addition, cultural practices may constrain women's access to jobs and productive resources, as well as female labor demand.

Women's lifetime earnings are negatively associated with the number of children, and this is especially the case for women who are less educated and those who begin childbearing early. In the United States, this association is strongest among poor and less educated women (Angrist and Evans 1996). Similarly, in the United Kingdom, a woman with no job qualifications and two children has half the total lifetime earnings of her childless counterpart, and a mother of four has less than a fifth (Matheson and Summerfield 2001). In Chile, Barbados, Guatemala, and Mexico, Buvinic (1998) finds that adolescent

childbearing was associated with lower monthly earnings for mothers and lower child nutritional status only among poor people. In agrarian settings, women's participation in the labor force and lifetime economic productivity may be less disrupted by childbearing, but countering this hypothesis, tabulations of time-use data from countries in Africa, Asia, and Latin America, including several still largely agrarian societies, indicate that women spend twice as much time on unpaid care work as do men (Budlender 2008).

Effects of Girls' Education and Urbanization on Fertility

Women's education is strongly negatively correlated with fertility (Cochrane 1979). More recent studies that have controlled for potential endogeneity between the variables have established this link more conclusively in a range of settings, including Guatemala, Indonesia, and Nigeria (Behrman and others 2006; Breierova and Duflo 2004; Osili and Long 2004). However, further evidence is still needed to establish causality unambiguously and clarify the principal pathways by which women's schooling affects fertility. In many settings, including China and several African countries, urbanization has also been found to be correlated with lower fertility, partly due to higher opportunity costs (education and employment) and greater access to family planning services (Brockerhoff 1998; Goldstein, White, and Goldstein 1997).

Women's disadvantage in schooling is narrowing in most regions of the developing world. Grant and Behrman (2008) find that the gender gap in school enrollment at ages 10–12 (but not at older ages) fell between 1990–99 and 2000–05 in South Asia, Middle East and North Africa, and West and Central Africa. In Latin America, Southeast Asia, and Southeast Africa, they find that girls ages 10–12 have somewhat higher levels of primary school enrollment than do boys and that their secondary school attainment (conditional on enrollment) is higher than that of boys'.

II. The "Demographic Window of Opportunity"

The demographic window of opportunity is the period following fertility decline, when the share of working-age people in the population rises and dependency ratios are low (see figure 2b). This increases per capita income and per capita availability of public resources to invest in human capital and the infrastructure for economic growth. Aggregate savings can be raised, and the expanded labor force can be used to increase the pace of economic growth. This in turn helps speed fertility decline, in a virtuous cycle of high growth and low fertility. Eventually the population starts aging, and this window of opportunity closes.

Realizing the Demographic Dividend

The extent to which this window of opportunity can be converted into a demographic dividend of increased economic growth depends on the effectiveness of

state policies. Realizing the dividend requires early investments in schooling and health, so that the working population is educated and healthy. Robust evidence shows that while expanding the quantity of education is important, its quality is critical for economic growth (Hanushek and Woessmann 2008).[2] Policies, therefore, should encourage investments in school quality and in the expansion of employment opportunities, especially in industries that can absorb the semi-skilled labor that predominates in developing economies. Promoting international trade and a favorable investment climate and reducing labor market rigidities increase the demand for labor.

East Asian countries implemented such policies, and a third of the nearly tripling of real per capita income over 1965–90 has been attributed to their ability to harness the demographic shifts to advantage (Bloom and Williamson 1998). By contrast, Latin America's underperformance compared with East Asia during this period has been linked to protectionist trade policies that impeded realization of this dividend (IDB 2000). However, establishing causality between demographic and economic change is methodologically complex, given simultaneity between demographic and economic variables. Shultz (2009) finds, for instance, that results vary enormously depending on the level of analysis. He finds large estimates at the aggregate level (even when cross-country regressions use instruments to control for simultaneity) but only weak, insignificant estimates with more rigorous household survey data for the association between a rising share of households with working-age individuals and increasing aggregate demand for savings (following the life-cycle savings model).

Gender Inequality and the Demographic Dividend

Gender inequality can mediate the effect of the demographic bulge on economic growth in a number of ways. First, it can slow the speed of fertility decline and therefore the timing and size of the window of opportunity. This can result in a shallow but prolonged window of opportunity, not the sharp surge of working-age population seen in East Asia.

Second, gender inequality in schooling can limit the potential for economic growth by restricting the pool of talent and reducing average labor force quality (World Bank 2001). Galor and Weil (1996) conclude that gender inequalities in schooling and employment can seriously diminish countries' prospects for growth and poverty reduction. Studies in a range of settings indicate that expanding women's education and control over household resources is associated with better child health and education (King and Hill 1993; World Bank 2007), boosting the potential for future productivity and economic growth.

2. Improving health outcomes has been widely found to be associated with better cognitive outcomes as well as greater labor productivity and higher income at the micro level, but the impact of health improvements on economic growth at the macro level is difficult to measure and more mixed (Glewwe and Miguel 2008; Jack and Lewis 2007).

Third, increased female labor force participation (and greater gender equality in labor markets) contributes to the demographic dividend. It has been argued that women's entrance into the labor force was one of the most important features in East Asia's demographic dividend (Mason 2006). Shultz (2009) argues that the gains from increases in female labor supply will be larger than those from changing age structures.

Cultural restrictions can constrain women's economic contributions. In the Middle East, such restrictions have prevented women from taking advantage of the opportunities created by economic opening (Schultz 2009). And because of cultural restrictions, it is also common for women to be less likely to own or have secure access to productive assets, such as land. Deininger (2008) shows that in Ethiopia increasing women's tenure security (by issuing land titles in the names of both husbands and wives) increases farm productivity and empowers women in the household.

In societies with strongly patrilineal kinship systems, parents have lower incentives to invest in schooling their daughters since they cannot receive support from married daughters. Filmer (2000) finds a large gender gap in female enrollment in selected countries in South Asia and the Middle East, especially in poorer households.

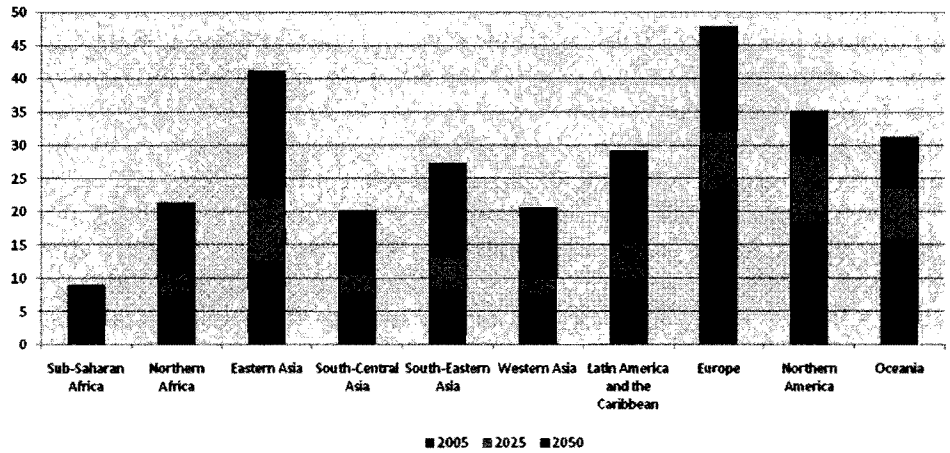
III. "Demographic Implosion": Rapidly Aging Populations

The rapid fertility decline in much of the developing world is graying the population much faster than in the developed world and typically at lower levels of income, straining sources of old age support. Although the share of older people in industrialized nations remains higher, by mid-century 80 percent of the world's elderly will be living in developing economies. The most rapid growth of the elderly population will be in East Asia, which experienced the most rapid fertility declines (figure 3). South Asia and the Middle East will have more time to put formal old age support systems in place, and for Sub Saharan Africa the problem is far in the future.

Systems of Old Age Support

In the developed world, formal systems support the elderly and protect against poverty in old age, although an increasing number of countries face problems with the continued financing of these systems as the proportions of the elderly rise (Williamson and Smeeding 2004). Some middle-income countries such as Brazil and South Africa show that modest cash transfers directly targeted to the elderly can significantly reduce poverty and extreme poverty (Case 2004; de Carvalho 2008). Interestingly, when women received the transfers they also used them to improve their grandchildren's health and schooling, an effect that was not measurable when men received the transfers (Carvalho 2008; Duflo 2003). These measures may be fiscally unaffordable in many low-income countries, however, where until recently the formal sector has been small and citizens have

FIGURE 3. Increases in the Old Age Dependency Ratio, 2005–50 (Ratio of Projected Population Ages 65+ to ages 15–64)



Note: Data refer to the medium variant assumption of fertility trends.
Source: Authors' analysis based on data from United Nations Population Division (2006).

not contributed to pension funds. These countries may also lack the administrative capacity to deliver such targeted support. In most developing economies, non-contributory pensions provide negligible support (Casabonne 2007).

Traditional systems of familial support are still largely in place in the developing world, with children the main source of material and physical support for their elderly parents. The proportions of elderly living with their children are highest in Asia, followed by Africa and Latin America (United Nations 2005). Coresidence with children is associated with a lower likelihood of being poor in old age, although causality has not been established. Smeeding and others (2008) find in middle-income countries (China; Taiwan, China; and Mexico) that people in multigenerational households have net disposable income poverty rates that are not much different from those in rich countries. Non-coresident children may also help support their parents financially and otherwise.

With urbanization and modernization, the proportion of elderly living in multigenerational households shrinks: between the early 1980s and 2000, it fell from 78 to 66 percent in Mexico, and from 95 to 58 percent in Taiwan, China (Smeeding and others 2008). However, familial support can show remarkable resilience even after decades of social change. In Taiwan, China, only 29 percent of elderly respondents said that pensions or retirement benefits constituted their main source of income (Chan 2005). Persistence in patterns of familial support offers a window of time for countries to build up the resources to establish more formal old age support policies. These policies need to be carefully designed, since government transfers to the needy can diminish the flow of private transfers from relatives (Jimenez and Cox 1992).

Gender Inequalities Heighten the Problems of Old Age Support

Women are more vulnerable than men to poverty in old age, for several reasons. First, they have lower lifetime earnings as a basis for earnings-linked support systems and for personal savings, because they participate less in the formal labor force, are paid less for their participation, and childbearing lowers their lifetime earnings curve. And they own fewer assets than men (World Bank 2007). In sum, women make large non-monetary contributions to their families, but are in turn more dependent on them for support.

Second, women live longer than men, on average, largely because of women's biological advantages and men's greater tendency to engage in risky behavior.[3] Gender differences in opportunity also contribute, for example, through men's greater exposure to hazardous occupations. Greater longevity can be desirable, but it also means that women are likely to be widowed and perhaps lose access to all or part of their husband's income or pension. Moreover, women are exposed to a longer period of old age, with potentially poor health, when the need for support is higher. On the other hand, studies indicate that women are better integrated than men into family and social networks, giving them sources of resilience (Knodel and Ofstedal 2003).

The elderly are more vulnerable where kinship systems prescribe that only sons can support their aged parents and less vulnerable when children of either sex can provide parental support. Fertility decline further increases the likelihood of elderly parents having no son able to support them. This suggests that the need for state support may be higher in some societies than in others and that policies for state support may need to be tailored to prevailing cultural patterns.

An especially striking manifestation of how old age support can be affected by gender inequalities is currently unfolding in China, where significant proportions of girls are "missing" as a result of strong son preference. While this problem has been widely discussed, Ebenstein and Sherigin's (2009) article in this issue explores its dynamics and implications. They show that significant proportions of men in China will remain single and will face an old age without the physical and financial support of a spouse and children. The regions of China with the highest levels of culling of girls are also the prosperous regions of the country, able to attract marriageable women from poorer areas. Thus, Ebenstein and Sherigin show that the unmarried men will be concentrated among those who are poorer and less educated, living in regions with lower employment opportunities and resources for providing public support to

3. Genetic and physiological factors predispose women to greater longevity (Waldron 2005)—while men have higher levels of engagement in life-threatening behaviors such as smoking and drinking and violence as well as higher exposure to hazardous occupations, resulting in higher mortality rates from causes such as lung cancer, accidents, suicide, and homicide (McKee and Shkolnikov 2001; Waldron 2005).

their citizens. Ironically, a cultural bias against girls will increase male poverty and vulnerability, especially in old age.

IV. "Demographic Hourglass": Working-age Adults Missing

If premature mortality shrinks the prime working-age population, there are economic and social consequences at both macro and micro levels that are opposite those of the demographic window of opportunity. At the macro level, the labor shortfall and the rise in dependency ratios diminish the potential for economic growth and poverty reduction. At the micro level, households suffer from the loss of income from the missing adult, and the remaining members of the household are put under greater stress to provide and care for household members. Familial support for the elderly and for children becomes stressed. On the other hand, the labor shortages can benefit women and others who might normally have greater difficulty finding a job.

Demographic Effects of Premature Adult Mortality and Skewed Age–Sex Distributions

One major factor contributing to hourglass populations is armed conflict, which typically boosts mortality among men more than among other population groups. The resulting shortage of working-age men can be dramatic: in 1950, after World War II, there were 40 percent more women than men ages 25–39 in Germany, and 57 percent more in what is now the Russian Federation (see figure 2d).[4] In Africa in 2000, there were an estimated 51 male deaths per 100,000 compared with 15 female deaths per 100,000—for low- and middle-income countries the average was 6.2 male deaths per 100,000 (WHO 2002)—suggesting excess male mortality due to violence. Studies show the devastating effects of conflict at the macro and micro levels and have also analyzed the economic causes of conflict (Collier and Sambanis 2005; Stewart and Fitzgerald 2001). However, studies of gender differentials in the effects of conflict are limited by the difficulties of data collection in the aftermath of war and of the attribution of causality, so they are not reviewed here.[5]

Another possible factor producing hourglass populations is the reverse skewing of the population age–sex distribution as a result of strong societal son preference, which increases girls' mortality. Current sex ratios at birth in China, as indicated by the 2005 One Percent Population Sample Survey, imply that these birth cohorts will have nearly 20 percent fewer female than male adults. The fallout for men has been discussed above. However, son preference may not shrink the total labor force, since it implies a choice not to reduce family size, but to alter its sex composition. Indeed, if parents are inefficient at

4. United Nations Population Division. http://esa.un.org/unpp.

5. The effect of large-scale labor outmigration is also not discussed, since the literature shows that labor migrants are not typically lost to their place of origin unless they take their families with them.

culling daughters, son preference may increase the size of birth cohorts (Filmer, Friedman, and Schady 2009).

HIV/AIDS is another important factor contributing to high levels of premature adult mortality and shrinking the labor force. By 2020, some of the worst affected countries (Botswana, Lesotho, Swaziland, and Zimbabwe) will have lost an estimated 35 percent or more of their working-age populations (ILO 2006). A high prevalence of HIV/AIDS is estimated to have a large negative impact on GDP growth (Corrigan, Glomm, and Mendez 2005). The mechanisms through which this happens include loss of labor productivity because of illness and death, increased health care expenditures that lead to dissaving, and lower capital accumulation and expenditures on schooling.

HIV Prevalence and Gender Inequality

Both men and women are affected by HIV, but in Africa many more women than men are affected and at younger ages than men. Recent surveys in eight African countries show HIV prevalence rates of 1.2–1.8 times higher among women than among men in six of these countries, particularly in those below ages 35–40 (Mishra and others 2005). The Joint United Nations Programme on HIV/AIDS estimates that women make up three-quarters of Africans ages 15–24 who are HIV positive (UNAIDS 2008).

Impact on Orphans. Orphanhood can be associated with poorer schooling and health outcomes for children, especially in poorer households. In a sample of 39 Demographic and Health Surveys, poor households in about a third of countries had a significant orphan disadvantage, but there was not a larger gender gap for orphans than for non-orphans in most countries (Ainsworth and Filmer 2002).

Studies in Tanzania, where longitudinal data have been carefully collected and analyzed, find that children who lose their mother fare worse than those who lose only their father. Their schooling suffered more if they lost their mother, especially in poorer households (Ainsworth, Beegle, and Koda 2005). They were also disadvantaged in height, especially if they were younger when orphaned (Beegle, De Weerdt, and Dercon 2008). Orphaned girls face an additional disadvantage as a consequence of their gender: they become exposed to HIV earlier than other children, especially if they come from poorer households (Beegle and Krutikova 2007).

Impact on Spouses and Children. Having a working-age adult ill with HIV takes a heavy toll on other household members. Children are diverted to activities other than schooling. In Tanzania, longitudinal data show that children sharply reduced their hours spent in school before the death but returned to school after the death (Ainsworth, Beegle, and Koda 2005). Girls are sometimes disproportionately affected (Yamano and Jayne 2004).

An evaluation of the impact of antiretroviral therapy in Kenya found that access to antiretroviral drugs helps keep HIV-infected adults in the labor force. This is associated with lower participation of women in cash-generating work, in turn freeing children (boys more than girls) from additional household work to return to school (Thirumurthy, Zivin, and Goldstein 2005).

IMPACT ON AGING PARENTS. Studies from a range of settings find that aging parents face heavy burdens if they have a child with HIV. First, the toll of caring for an adult child with HIV can deplete their resources, especially for the poorer elderly. Studies in Asia (Cambodia and Thailand) and Southern Africa find that to meet these costs, parents sell assets, use their savings, take loans, and work extra hours (Gregson 2008). Several of these studies are based on surveys of self-reported status after the child becomes ill or dies, so the results of Adhvaryu and Beegle (2008), which are based on longitudinal data, are of special value. They find that in Tanzania parents who lose adult children deplete their assets to meet the additional expenses and that thereafter older women increase their working hours, working longer into old age.

Second, aging parents support their orphaned grandchildren. In both Asia (Cambodia and Thailand) and Southern Africa, surveys indicate that grandparents are the primary caregivers for about a third of orphaned children (Deininger, Garcia, and Subbarao 2003; Gregson 2008) and partially support others.

Third, losing adult children increases old-age vulnerability to poverty. For about half of poorer parents in Cambodia and Thailand who had lost an adult child, that child was the parents' main source of support (Knodel 2006). Bereaved parents can potentially turn to other children for support, but their need for support is intensified by their depleted assets. And high levels of mortality can offset the potential risk diversification of having several children. It is estimated that by 2010 in South Africa, nearly one in five people older than 60 will have no surviving children (Merli and Palloni 2006).

V. DISCUSSION AND POLICY IMPLICATIONS

Viewed through the lens of the dominant demographic conditions in a country, the development implications of gender inequality become clearer. These implications relate in particular to the value of girls in the family, the quality of family planning services, gender equity in schooling and employment, and old age support. Cultural factors leading to a preference for sons over daughters are typically viewed as a problem for girls and women, since they may result in the culling of unwanted daughters and reduce parental willingness to invest in girls' health and schooling.

Filmer, Friedman, and Schady's (2009) article in this issue shows that son preference can also increase the likelihood that girls will be exposed to parental resource dilution: parents continue childbearing until they reach their desired

number of sons, so girls tend to belong to larger families. However, as Ebenstein and Sherigin's (2009) article in this issue shows, from a demographic scenario perspective, high male to female sex ratios at birth or a culture in which only sons are able to provide old age support affects not only young girls, but also older men's vulnerability to poverty and the country's ability to support its elderly population. Culture, fortunately, is malleable, and studies show that son preference can diminish in the face of modernization, especially if the media and other sources are used to reshape attitudes toward daughters (Chung and Das Gupta 2007).[6]

In addition to the influence of the media, the surest way to increase the value of girls and women in the family and society is to invest in reducing gender inequalities in schooling and employment. These are smart investments across the different demographic scenarios. However, family planning programs should head the list in settings where high fertility is a major threat to countries' prospects for economic growth and household poverty reduction.[7] Such programs are particularly cost-effective in high-fertility settings and in settings with high HIV prevalence and should be a priority investment for the 63 countries that currently experience high fertility (see map 1), especially the 15 countries in Sub-Saharan Africa that show both high fertility and hourglass mortality due to HIV/AIDS. The direct health benefits of family planning for women and children from increasing birth intervals, reducing teen pregnancies, and preventing mother-to-child HIV transmission are well known (Levine and others 2006). In addition, Schultz (2009) shows that in a social experiment in Matlab, Bangladesh, family planning increased women's wages and their investments in child quality (sons' schooling and daughters' health). These health and economic payoffs to family planning should accelerate fertility decline in high-fertility, high–hourglass mortality countries.[8]

To fully reap the benefits of the demographic dividend as countries transition from the high-fertility scenario to the early stages of the demographic window of opportunity, investments in schooling that increase education quality for all and reduce gender gaps in school enrollment and completion are

6. Development has two countervailing effects on the culling of girls. On the one hand, improved financial and physical access to better sex-selective technology tends to increase culling. On the other hand, modernization changes attitudes in favor of greater gender equality. Studies indicate that the second effect comes to outweigh the first.

7. Pritchett (1994) argues that desired family size rather than family planning programs is a major factor contributing to fertility reduction. However, his analysis ignores the fact that successful family planning programs focus intensively on reducing desired family size. For example, in India, the media are used heavily through direct advertising, soap operas, and the like to disseminate the idea that smaller families are happier families. Jensen and Oster's (2008) study of cable television in India shows that media exposure reduces fertility.

8. These benefits of family planning in high-fertility Sub-Saharan African countries, where women record high unmet need for contraception (19.4 percent of women, according to Levine and others 2006), should influence desired fertility and counter Pritchett's (1974) argument that desired fertility and not family planning is the predominant factor in explaining fertility decline.

especially important. Reducing gender gaps in school attainment without increasing school quality and the cognitive skills of female and male students will not have desired effects on individual earnings and economic growth (Hanushek and Woessmann 2008). These efforts should pay special attention to increasing the access of girls from socially excluded groups, who constitute the largest proportion of girls not in primary school, and narrowing gender gaps in secondary schooling, where gender disparities are the widest and the returns to girls' schooling the greatest (Tembon 2008). Investments also need to be balanced in the growing number of cases where boys' schooling attainment is lagging behind or reversing.

There is a rich literature on policy options to improve student performance and reduce gender disparities in schooling. Schooling reforms that seek to improve student performance can take decades (20–30 years), but improved student performance is critically linked to economic growth, and evidence indicates that it benefits girls more than boys (Tembon 2008). Reducing gender disparities in schooling includes supply-side interventions, such as broadening school options for girls and, notably, incentives so that parents send children to school. Demand-side interventions such as conditional cash transfer programs have been found to be effective in expanding child schooling, sometimes more for girls than for boys (Fiszbein and Schady 2009). In settings such as Bangladesh, where cultural factors preclude parents' receiving support from married daughters, disincentives for investing in schooling for daughters can be offset through scholarships or stipends for girls (Khandker, Pitt, and Fuwa 2003).

Schooling opportunities for all, but especially for girls, can be severely compromised in hourglass settings because of HIV/AIDS and conflict. Policy alternatives include accelerated learning programs and targeted programs to improve the school-to-work transition and skill deficits of poor adolescent girls and young women—a critical target group or entry point for interventions that seek to break the intergenerational transmission of poverty. Rigorous evaluations of a generation of demand-driven youth training programs in Latin America have found that they successfully ease young women's transition into jobs by promoting equal access to women, especially through training in nontraditional skills, stipends for childcare, and strong links between training and private sector labor demand (Attanasio, Kugler, Meghir 2008; Ñopo, Robles, and Saavedra 2007). The challenge is to roll out these designs in high-fertility, high HIV/AIDS, and post-conflict hourglass settings with substantially less institutional capacity.

Schultz's (2009) article in this issue surveys the literature on the macro and micro links between the demographic window of opportunity and economic development, showing how labor market and microcredit policies could be better designed to increase women's income-earning potential. Policies that expand labor demand and create economic opportunities for women are fundamental for reducing the intergenerational transmission of poverty in

high-fertility and high young-adult-mortality scenarios, to help realize the demographic dividend and to increase private savings for old age. These policies, in turn, will have positive impacts on girls' schooling by narrowing the gap between the returns to girls' and boys' schooling and helping to change parents' perception that girls are less valuable than boys. The time to implement these policies is now—in the past decade and a half girls' schooling has improved noticeably but women's labor force participation rates have barely budged (World Bank 2007).

Trade policies can expand women's employment by increasing employment in sectors that favor women, such as the garment industry, electronics, and high-value agricultural exports (cut flowers and fruits)—while also reducing the gender gap in employment in other industries. But there is also evidence that trade works to narrow the occupational gender wage gap in richer but not in poorer countries (Oostendorp 2009). It may be that women need a certain skill level to take advantage of trade-related opportunities.

This argues for skills training and labor-intermediation programs targeted to youth and women to facilitate their entry into the employment opportunities generated by the opening of markets to international trade. More generally, women tend to be more vulnerable to labor market conditions than do men and to experience higher unemployment rates and may therefore benefit disproportionately from an overall expansion in job opportunities. For example, Kolev and Sirven (2007) find a positive relationship between male and female employment ratios in 21 African countries, and they find the lowest gender gaps in employment in countries with the largest male employment ratios.

Experience with employment generation (public works) programs designed to cushion the impact of economic crisis on poor people shows that the programs can ensure high female participation by including incentives to attract female labor, such as numerical targets for women's participation, use of community-based intermediary agencies, work sites close to home, home-based production, and availability of child care (Buvinic 2008).

In addition to wage employment, expanding women's access to entrepreneurship and self-employment is critical to boost women's economic opportunities and contributions and is especially timely in demographic explosion and hourglass scenarios with undeveloped labor markets. Doing this requires increasing women's access to credit and other inputs. Substantial experience with microfinance agencies shows that their programs effectively reach women and can have larger benefits for female borrowers than for male borrowers (Armendariz de Aghion and Morduch 2005; Pitt and Khander 1988). Less is known about successful models to increase women's access to formal financial services, while recent studies in Italy, Eastern Europe, and Central Asia find that women are disadvantaged in obtaining finance from commercial banks (Alesina, Lotti, and Mistrulli 2008; Sabarwal and Terrell 2008).

Policies also need to be carefully designed to help women remain in the labor force when they have children. Access to child care is a common feature of public

works programs that reach women, and public investments in pre-primary schooling increase maternal labor supply (Berlinski and Galiani 2007). However, labor market regulations that seek to protect women by requiring employers to pay for maternity leave and similar fringe benefits can reduce employers' willingness to hire women. Public funding for such fringe benefits may be more effective. Further, most mandated fringe benefits are withheld from wages, leaving employers' cost of labor relatively unaffected (Shultz 2009).

As populations age, governments increasingly need to consider safety net and pension options for vulnerable elders, typically older women. Modest targeted non-contributory pensions have been found effective in reducing old age poverty in Brazil and South Africa and might be used to help poor parents elsewhere cope with the costs of caring for HIV-affected children and their orphans.

Substantial research gaps remain on the relationships among demographic factors, gender inequalities, and economic outcomes. Foremost, they include looking beyond associations between variables to testing for causality between, for instance, female schooling or female wages and lower fertility. In addition, evidence is needed on the relationship (and pathways) between growth outcomes and increased gender equality in schooling and the labor market. Also, evaluation research needs to better understand how specific policies affect women's labor supply, wages, and savings. How is women's labor force participation affected by gender inequalities in wages and discrimination against women in labor markets? And, in turn, how does women's participation in the labor force affect aggregate savings and investments? While difficult, another major research need is isolating the impact of demographic changes brought about by conflict on women's labor force participation, gender inequality, and family well-being.

Progress in establishing stronger links between gender equality and poverty and growth and in moving from establishing associations to asserting causality will increasingly be possible with growing investments in gender-informed panel studies with large samples and natural or scientific experiments. Filling these research gaps should help in identifying effective policy interventions for creating a virtuous cycle of increased gender equality, poverty reduction, and economic growth.

Funding

The World Bank's research initiative on gender, poverty, and demography was partly funded by a generous contribution from the government of Norway.

References

Adair, Linda, David Guilkey, Eilene Bisgrove, and Socorro Gultiano. 2002. "Effect of Childbearing on Filipino Women's Work Hours and Earnings." *Journal of Population Economics* 15(4):625–45.

Adhvaryu, Achyuta, and Kathleen Beegle. 2008. "The Impact of Adult Deaths on the Elderly in Tanzania." Background Paper for the Gender, Poverty, and Demography Initiative. Gender and Development Unit, World Bank, Washington, DC.

Ainsworth, Martha, and Deon Filmer. 2002. "Poverty, AIDS, and Children's Schooling: A Targeting Dilemma." Policy Research Working Paper 2885. World Bank, Washington, DC.

Ainsworth, Martha, Katheleen Beegle, and Godlike Koda. 2005. "The Impact of Adult Mortality and Parental Deaths on Primary Schooling in North-western Tanzania." *Journal of Development Studies* 41(3):412–39.

Alesina, Alberto F., Francesca Lotti, and Paolo E. Mistrulli. 2008. "Do Women Pay More for Credit? Evidence from Italy." Harvard Institute of Economic Research Discussion Paper 2159. Mass: Cambridge.

Angrist, Joshua D., and William N. Evans. 1996. *Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size.* NBER Working Paper 5778. Cambridge Mass.: National Bureau of Economic Research.

Angrist, Joshua D., Victor Lavy, and Analia Schlosser. 2005. *New Evidence on the Causal Link between the Quantity and Quality of Children.* NBER Working Paper 11835. Cambridge Mass.: National Bureau of Economic Research.

Armendariz de Aghion, Beatriz, and Jonathan Morduch. 2005. *The Economics of Microfinance.* Cambridge, Mass.: MIT Press.

Attanasio, Orazio, Adriana Kugler, and Costas Meghir. 2008. *Training Disadvantaged Youth in Latin America: Evidence from a Randomized Trial.* NBER Working Paper 13931. Cambridge, Mass.: National Bureau of Economic Research.

Beegle, Kathleen, and Sofya Krutikova. 2007. "Adult Mortality and Children's Transition into Marriage." Policy Research Working Paper 4139. World Bank, Washington, DC.

Beegle, Kathleen, Joachim De Weerdt, and Stefan Dercon. 2008. "Adult Mortality and Consumption Growth in the Age of HIV/AIDS." *Economic Development and Cultural Change* 56(2):299–326.

Behrman, Jere R., Harold Alderman, and John Hoddinott. 2004. "Hunger and Malnutrition." In Bjorn Lomborg ed., *Global Crises, Global Solutions.* Cambridge U.K.: Cambridge University Press.

Behrman, Jere R., Alexis Murphy, Agnes Quisumbing, Usha Ramakrishna, and Kathryn Young. 2006. "What Is the Real Impact of Education on Age of First Parenthood and Family Formation?" Background Paper for *World Development Report 2007.* World Bank, Washington, DC.

Berlinski, Samuel, and Sebastian Galiani 2007. "The Effects of a Large Expansion of Pre-primary School Facilities on Preschool Attendance and Maternal Employment." *Labor Economics* 14(3):665–80.

Black, Sandra, Paul J. Devereux, and Kjell Salvanes. 2007. "Older and Wiser? Birth Order and IQ of Young Men." IZA Discussion Paper 3007. Institute for the Study of Labor, Bonn, Germany.

Bloom, David, and Jeffrey G. Williamson. 1998. "Demographic Transitions and Economic Miracles in Emerging Asia." *World Bank Economic Review* 12(3):419–55.

Bloom, Shelah S., David Wypij, and Monica Das Gupta. 2001. "Dimensions of Women's Autonomy and the Influence on Maternal Health Care Utilization in a North Indian City." *Demography* 38(1):67–78.

Breierova, Lucia, and Esther Duflo. 2004. *The Impact of Education on Fertility and Child Mortality: Do Fathers Really Matter Less Than Mothers?* NBER Working Paper 10513. Cambridge, Mass.: National Bureau of Economic Research.

Brockerhoff, Martin., 1998. "Migration and the Fertility Transition in African Cities." In Richard E. Bilsborrow ed., *Migration, Urbanization, and Development: New Directions and Issues.* Norwell, Mass.: Kluwer Academic Publishers.

Budlender, Debbie. 2008. "The Statistical Evidence on Care and Non-Care Work across Six Countries." Gender and Development Programme Paper 4. Geneva: United Nations Research Institute for Social Development.

Buvinic, Mayra. 1998. "The Costs of Adolescent Childbearing: Evidence from Chile, Barbados, Guatemala and Mexico." *Studies in Family Planning* 29(2):201–09.

———. 2008. "The Global Financial Crisis: Assessing Vulnerability for Women and Children, Identifying Policy Responses." Paper prepared for the 53rd Session of the UN Commission on the Status of Women, 25 February–7 March, New York.

Casabonne, Ursula. 2007. "Population Aging, Gender and Non Contributory Pensions in Developing Countries." Background Paper for the 'Gender, Poverty and Demography Initiative. World Bank, Gender and Development Unit, Washington, DC.

Case, Anne. 2004. "Does Money Protect Health Status? Evidence from South African Pensions." In David Wise ed., *Perspectives on the Economics of Aging*. Chicago: University of Chicago Press.

Chan, Angelique. 2005. "Aging in Southeast and East Asia: Issues and Policy Directions." *Journal of Cross-Cultural Gerontology* 20(4):269–84.

Chung, Woojin, and Monica Das Gupta. 2007. "The Decline of Son Preference in South Korea: The Roles of Development and Public Policy." *Population and Development Review* 33(4):757–83.

Cochrane, Susan Hill. 1979. *Fertility and Education*. Baltimore, Md.: Johns Hopkins University Press.

Collier, Paul, and Nicholas Sambanis eds. 2005. *Understanding Civil War*. Vols 1 and 2. World Bank, Washington, DC.

Conley, Dalton, and Rebecca Glauber. 2006. "Parental Educational Investment and Children's Academic Risk: Estimates of the Impact of Sibship Size and Birth Order from Exogenous Variation in Fertility." *Journal of Human Resources* 41(4):722–37.

Connelly, Rachel, Deborah S. DeGraff, Deborah Levison, and Brian P. McCall. 2006. "Tackling the Endogeneity of Fertility in the Study of Women's Employment in Developing Countries: Alternative Estimation Strategies using Data from Urban Brazil." *Feminist Economics* 12(4):561–97.

Corrigan, Paul, Gerard Glomm, and Fabio Mendez. 2005. "AIDS, Human Capital, and Growth." *Journal of Development Economics* 77(1):107–24.

de Carvalho, Filho, and Irineu Evangelista. 2008. "Household Income as a Determinant of Child Labor and School Enrollment in Brazil: Evidence from a Social Security Reform." IMF Working Paper 08/241. Washington, DC: International Monetary Fund.

Deininger, Klaus. 2008. "Rural Land Certification in Ethiopia Empowers Women." *World Bank Gender Action Plan Newsletter*. March, Washington, DC.

Deininger, Klaus, Marito Garcia, and K. Subbarao. 2003. "AIDS-Induced Orphanhood as a Systemic Shock: Magnitude, Impact, and Program Interventions in Africa." *World Development* 31(7):1201–20.

Duflo, Esther. 2003. "Grandmothers and Granddaughters: Old-Age Pensions and Intrahousehold Allocation in South Africa." *World Bank Economic Review* 17(1):1–25.

Ebenstein, Avraham, and Ethan Sherigin. 2009. "The Consequences of the "Missing Girls" of China." *World Bank Economic Review* 23(3).

Filmer, Deon. 2000. "The Structures of Social Disparities in Education: Gender and Wealth." Policy Research Working Paper 2268. World Bank, Washington, DC.

Filmer, Deon, Jed Friedman, and Norbert Schady. 2009. "Development, Modernization, and Son Preference: The Role of Family Sex Composition." *World Bank Economic Review* 23(3).

Fiszbein, Ariel, and Norbert Schady. 2009. *Conditional Cash Transfers: Reducing Present and Future Poverty*. A World Bank Policy Research Report. World Bank, Washington, DC.

Galor, Oded, and David N. Weil. 1996. "The Gender Gap, Fertility, and Growth." *American Economic Review* 86(3):374–87.

Garenne, Michel. 2003. "Sex Differences in Health Indicators among Children in African DHS Surveys." *Journal of Biosocial Science* 35(4):601–14.

Glewwe, Paul, and Edward Miguel. 2008. "The Impact of Child Health and Nutrition on Education in Less Developed Countries." In T. Schultz, and John Strauss eds., *Handbook of Development Economics*. Vol. 4. Amsterdam: North-Holland.

Goldstein, Sidney, Michael White, and Alice Goldstein. 1997. "Migration, Fertility, and State Policy in Hubei Province, China." *Demography* 34(4):481–91.

Grant, Monica J., and Jere R. Behrman. 2008. "Gender Gaps in Educational Attainment in Less Developed Countries." *Paper presented at the Population Association of America Meeting*, April 17–19, New Orlean, La.

Gregson, Simon. 2008. "What Becomes of the Left Behind? Survivors in Aids-afflicted Populations in Sub-Saharan Africa and Asia." *Paper presented at Meeting on Gender, Poverty and Demography*, March 25, World Bank, Washington, DC.

Hanushek, Eric A., and Ludger Woessmann. 2008. "The Role of Cognitive Skills in Economic Development." *Journal of Economic Literature* 46(3):607–68.

Henderson, Daniel J., Daniel L. Millimet, Christopher F. Parmeter, and Le Wang. 2008. "Fertility and the Health of Children: A Non-parametric Investigation." In Daniel Millimet, Jeffrey Smith, and Edward Vytlacil eds., *Modelling and Evaluating Treatment Effects in Econometrics*. Vol. 21. Bingley, U.K.: Emerald Group Publishing Ltd.

IDB (Inter-American Development Bank). 2000. *Development beyond Economics: Economic and Social Progress in Latin America*. Washington, DC: Inter-American Development Bank.

ILO (International Labour Organization). 2006. *HIV/AIDS and Work: Global Estimates, Impacts on Children and Youth, and Response*. Geneva: International Labour Office.

Jack, William, and Maureen Lewis. 2007. "Health Investments and Economic Growth: An Overview," *Paper presented at the Commission on Growth and Development Workshop on Health and Growth*, October 16, Washington, DC.

Jensen, Robert, and Emily Oster. 2008. "The Power of TV: Cable Television and Women's Status in India." *Quarterly Journal of Economics* 124(3):1057–94.

Jimenez, Emmanuel, and Donald Cox. 1992. "Private Transfers and Public Transfers in Developing Countries: A Case of Peru." *World Bank Economic Review* 6(1):155–69.

Joshi, Shareen, and T. Paul Schultz, 2007. "Family Planning as an Investment in Development: Evaluation of a Program's Consequences in Matlab, Bangladesh." Working Paper 951. Yale University Economic Growth Center, Hew Haven, Conn.

Khandker, Shahidur R., Mark Pitt, and Nobuhiko Fuwa. 2003. "Subsidy to Promote Girls' Secondary Education: The Female Stipend Program in Bangladesh." World Bank, Washington, DC.

King, Elizabeth M., and M. Anne Hill. 1993. *Women's Education in Developing Countries: Barriers, Benefits, and Policies*. World Bank, Washington, DC.

Knodel, John. 2006. *Poverty and the Impact of AIDS on Older Persons: Evidence from Cambodia and Thailand*. University of Michigan Population Studies Center Research Report. Ann Arbor, Mich.: University of Michigan.

Knodel, John, and Mary Beth Ofstedal. 2003. "Gender and Aging in the Developing World: Where Are the Men?" *Population and Development Review* 29(4):677–98.

Kremer, Michael. 1993. "Population Growth and Technological Change: One Million B.C. to 1990." *Quarterly Journal of Economics* 108(3):681–716.

Kolev, Alexandre, and Nicholas Sirven. 2007. "Gender Disparities in Africa's Labor Market: A Cross-country Comparison using Standardized Survey Data." French Development Agency, Paris.

Levine, Ruth, Ana Langer, Nancy Birdsall, Gaverick Matheny, Merrick Wright, and Angela Bayer. 2006. "Contraception." In D.T Jamison, J.G. Breman, A.R. Measham, G. Alleyne, M. Claeson, D.B. Evans, P. Jha, A. Mills, and P. Musgrove eds., *Disease Control Priorities in Developing Countries*. World Bank, Washington, DC.

Mason, Andrew. 2006. "Capitalizing on the Demographic Dividend." In *Population and Poverty: Population and Development Strategies*. New York: United Nations Population Fund.

Matheson, Jil, and Carol Summerfield. 2001. *Social Trends 31*. London: U.K. Office for National Statistics.

McKee, Martin, and Vladimir Shkolnikov. 2001. "Understanding the Toll of Premature Death among Men in Eastern Europe." *British Medical Journal* 323(3):1051–55.

Menken, Jane, Linda Duffy, and Randall Kuhn. 2003. "Childbearing and Women's Survival: New Evidence from Rural Bangladesh." *Population and Development Review* 29(3):405–26.

Merli, M. Giovanna, and Alberto Palloni. 2006. "The HIV/AIDS Epidemic, Kin Relations, Living Arrangements, and the African Elderly in South Africa." In B. Cohen, and J. Menken eds., *Aging in Sub-Saharan Africa: Recommendations for Furthering Research.* Washington, DC: National Academies Press.

Mishra, Vinod, Arnold Fred, Fredrick Otieno, Anne Cross, and Hong Rathavuth. 2005. "Education and Nutritional Status of Orphans and Children of HIV-infected Parents in Kenya." Demographic and Health Surveys Working Paper 24. ORC Macro, Calverton, MD.

Ñopo, Hugo, Miguel Robles, and Jaime Saavedra. 2007. "Occupational Training to Reduce Gender Segregation: The Impacts of Pro Joven." Inter-American Development Bank Research Department Working Paper 623. Inter-American Development Bank, Washington, DC.

Oostendorp, Remco H.. 2009. "Globalization and the Gender Wage Gap." *World Bank Economic Review* 23(1):141–61.

Osili, Una Okonkwo, and Bridget Terry Long. 2004. "Does Female Schooling Reduce Fertility? Evidence from Nigeria." Indiana University–Purdue University, Indianapolis, Indiana.

Pitt, Mark M., and Shahidur R. Khandker. 1998. "The Impact of Group-Based Credit Programmes on Poor Households in Bangladesh: Does Gender of the Participant Matter?" *Journal of Political Economy* 106(5):958–96.

Pritchett, Lant H. 1994. "Desired Fertility and the Impact of Population Policies." *Population and Development Review* 20(1):1–56.

Sabarwal, Shwetlena, and Katherine Terrell. 2008. "Does Gender Matter for Firm Performance? Evidence from the East European and Central Asia Region." Policy Research Working Paper 4705. World Bank, Washington, DC.

Schultz, T. Paul. 2009. "The Gender Consequences of the Demographic Dividend: An Assessment of the Micro and Macro Linkages between the Demographic Transition and Economic Development." *World Bank Economic Review* 23(3).

Smeeding, Timothy, Qin Gao, Peter Saunders, and Coady Wing. 2008. "Elder Poverty in an Aging World: Conditions of Social Vulnerability and Low Income for Women in Rich and Middle-Income Nations." Paper prepared for the Gender and Development Unit, World Bank, Washington, DC.

Stewart, Frances, and Valpy Fitzgerald eds. 2001. *War and Underdevelopment.* Vols 1 and 2. Oxford, U.K.: Oxford University Press.

Strauss, Richard S. 2000. "Adult Functional Outcomes of Those Born Small for Gestational Age: Twenty-six-Year Follow-up of the 1970 British Birth Cohort?" *Journal of the American Medical Association* 283(5):625–32.

Tembon, Mercy. 2008. "Overview." In M. Tembon, and L. Fort eds., *Girls' Education in the 21st Century: Gender Equality, Empowerment and Economic Growth.* World Bank, Washington, DC.

Thirumurthy, Harsha, Joshua G. Zivin, and Markus Goldstein. 2005. *The Economic Impact of AIDS Treatment: Labor Supply in Western Kenya.* NBER Working Paper 11871. Cambridge, Mass.: National Bureau of Economic Research.

UNAIDS (Joint United Nations Programme on HIV/AIDS). 2008. *Addressing the Vulnerability of Young Women and Girls to Stop the HIV Epidemic in Southern Africa.* Geneva: Joint United Nations Programme on HIV/AIDS.

United Nations. 2005. *Living Arrangements of Older Persons around the World.* New York: Department of Economic and Social Affairs.

United Nations Population Division. 2006. *World Population Prospects: The 2006 Revision.* New York: Department of Economic and Social Affairs.

Waldron, Ingrid. 2005. "Gender Differences in Mortality—Causes and Variation in Different Societies." In Peter Conrad, and Rochelle Kern eds., *The Sociology of Health and Illness: Critical Perspectives.* New York: Worth-St. Martin's Press.

Williamson, James, and Timothy M. Smeeding. 2004. "Sliding into Poverty? Cross-National Patterns of Income Change and Benefit Decay in Old Age." Center for Retirement Research at Boston College. Working Paper 2004-25. Boston: Boston College.

World Bank. 2001. "Engendering Development: Through Gender Equality in Rights, Resources, and Voice." A World Bank Policy Research Report. New York: Oxford University Press.

———. 2007. *Global Monitoring Report 2007: Confronting the Challenges of Gender Equality and Fragile States.* World Bank, Washington, DC.

WHO (World Health Organization). 2002. *World Report on Violence and Health*, Geneva: World Health Organization.

———. 2007. *Maternal Mortality in 2005.* Geneva: World Health Organization.

Yamano, Takashi, and Thomas S. Jayne. 2004. "Working-age Adult Mortality and Primary School Attendance in Rural Kenya." International Development Collaborative Policy Briefs KE-TEGEMEO-PB-05. Michigan State University Department of Agricultural Economics, East Lansing, Mich.

Development, Modernization, and Childbearing: The Role of Family Sex Composition

Deon Filmer, Jed Friedman, and Norbert Schady

Does the sex composition of existing children in a family affect fertility behavior? An unusually large data set, covering 64 countries and some 5 million births, is used to show that fertility behavior responds to the presence—or absence—of sons in many regions of the developing world. The response to the absence of sons is particularly large in Central Asia and South Asia. Modernization does not appear to reduce this differential response. For example, in South Asia the fertility response to the absence of sons is larger for women with more education and has been increasing over time. The explanation appears to be that a latent demand for sons is more likely to manifest itself when fertility levels are low. As a result of this differential fertility behavior, girls tend to grow up with significantly more siblings than do boys, with potential implications for their well-being when quantity–quality tradeoffs result in fewer material and emotional resources allocated to children in larger families. JEL codes: J16, J13, O15

A family preference for sons over daughters may manifest itself in various ways. An especially stark dimension is the excess mortality among girls documented in several Asian countries (see, for example, Zeng and others 1993 for China; Muhiri and Preston 1991 for Bangladesh; and Das Gupta 1987 for India). A similar phenomenon has been documented in the Middle East (Yount 2001). Son preference can also manifest itself through lower investments in the human capital of girls. Pande (2003) documents lower nutrition and immunization rates among girls in India. School enrollment and attainment among girls

Deon Filmer (corresponding author) is a lead economist in the Development Economics Research Department at the World Bank; his email address is dfilmer@worldbank.org. Jed Friedman is a senior economist in the Development Economics Research Department at the World Bank; his email address is jfriedman@worldbank.org. Norbert Schady is a senior economist in the Development Economics Research Department at the World Bank; his email address is nschady@worldbank.org. The authors thank Monica Das Gupta, Peter Lanjouw, Cynthia Lloyd, T. Paul Schultz, and three anonymous referees for valuable comments and suggestions, and Ryan Booth and Nicholas Ingwersen for outstanding research assistance. They are grateful for financial support from the Hewlett Foundation's Trust Fund on Fertility, Reproductive Health, and Socioeconomic Outcomes and the Government of Norway.

THE WORLD BANK ECONOMIC REVIEW, VOL. 23, NO. 3, pp. 371–398 doi:10.1093/wber/lhp009
Advance Access Publication October 23, 2009

lags behind that of boys in many South Asian, Middle Eastern, and North African countries (Filmer 2005).[1]

This study focuses on one manifestation of a "preference" for sons—a greater propensity for continued childbearing given an all-female rather than an all-male composition of existing children in the family. Such behavior could be the result of taste-based sex discrimination or of economic concerns, such as higher costs of investing in girls than in boys or lower pecuniary returns to investments in girls than in boys. Therefore, while differential fertility-stopping behavior is related to preferences, it is the result of a larger set of factors.

There are numerous possible reasons for observing differential fertility-stopping behavior in the developing world. Typically, they derive from conditions found in many traditional rural societies, such as inheritance systems that pass assets to sons, intergenerational insurance systems in which sons care for parents in old age, or production systems with low pecuniary returns to women's work (and to investments in women's human capital). General development processes and modernization, including urbanization, the dissolution of traditional rural communities, and increasing female education and labor force participation, are expected to work against these pressures for differential fertility-stopping behavior in settings where it exists (see, for example, Chung and Das Gupta 2007). This article explores the extent of son-preferred differential fertility-stopping behavior in the developing world; how it varies across countries and regions; whether it is associated with measures of modernization, such as urbanization, women's education, and wealth; and its potential consequences for household demographic composition and the investment in girls' human capital.

A handful of empirical studies have investigated differential fertility-stopping behavior at various levels of economic development. Hank and Kohler (2000) focus on European countries. Using Fertility and Family Surveys for 17 countries, they find substantial heterogeneity across countries, with a tendency toward a mild preference for a mixed-sex composition of children in a family. Their data suggest a preference for girls in the Czech Republic, Lithuania, and Portugal. Andersson and others (2006) use historical data from Denmark, Finland, Norway, and Sweden to show no effect of sex on fertility for second births, a desire for sex balance at third births, and heterogeneity across countries at fourth births (son preference in Finland and daughter preference in the other three countries).

For developing countries, most of the literature has focused on individual Asian countries with a prevalence of discrimination against women.[2] An

1. See World Bank (2001) for a more general discussion of differences between boys and girls in inputs and outcomes.

2. For example, Park (1983), Arnold (1985), Bairagi (1987), and Larsen, Chung, and Das Gupta (1998) show the strong impact of son preference on future fertility in the Republic of Korea; Arnold, Choe, and Roy (1998), Drèze and Murthi (2001), and Jensen (2007) find evidence that son preference affects fertility behavior in India; Haughton and Haughton (1995) show a similar pattern in Vietnam; while Pong (1994) and Leung (1998) document the pattern among ethnic Chinese in Malaysia. One study addresses the issue in Egypt, with a similar finding of son preference affecting fertility behavior (Yount, Langsten, and Hill 2000).

important exception to these country-specific studies is Arnold (1992, 1997), who considers the impact of sex ratios on subsequent fertility behavior across many developing countries. Arnold (1992) shows that the most typical pattern in the 26 countries he studied is of a preference for at least one son and one daughter. He finds some weak evidence for son-preferred differential fertility-stopping behavior in North Africa and Sri Lanka. Arnold (1997) analyzes data for 44 countries but focuses largely on the effect of sex ratios on *stated* fertility preferences and on some fertility behaviors, such as current pregnancy status and average birth spacing. He finds regional variation in the extent of an association between sex ratios and the outcomes he analyzes, with the strongest results suggesting son-preferred differential fertility-stopping behavior for the Asian and North African countries.

This article uses information on 5 million births by 1.3 million mothers in 64 countries to analyze how the sex mix of children in a family affects fertility decisions in the developing world. The article extends the literature in important ways. The analysis includes a large number of developing countries from disparate regions. The article documents not only regional patterns in son-preferred differential fertility-stopping behavior, but also within-region differences by location (urban or rural), education (women who have completed primary school and those with less schooling), wealth levels (above and below the median of a composite measure of assets), and over time (different birth cohorts of mothers). The article analyzes the extent to which observed patterns in son-preferred differential fertility-stopping behavior strengthen or weaken as the total number of children decreases. Moreover, finally, the results are linked to the wider literature on sex composition and resource dissolution in larger families.

I. Methods and Data

This section describes the methodology, starting with a model for estimating the impact of the sex balance of children in a family on the probability of subsequent births. It then details the data used for the analysis.

Estimating the Impact of Sex Balance on Fertility Behavior

The basic model estimates:

$$B_{wn+1} = a + b_{mn} \cdot M_{wn} + b_{fn} \cdot F_{wn} + u_{wn} \quad \text{for } n \geq 2 \tag{1}$$

where B_{wn+1} is a zero or one outcome variable indicating a birth for woman w with a preexisting number of children n; M_{wn} is a variable equal to one if woman w had no sons at family size n; F_{wn} is a variable equal to one if woman w had no daughters at family size n; and the term u_{wn} is a random error. This regression is run separately for each existing family size.

The omitted category in the regression is women who have at least one son and one daughter. The coefficients b_{mn} and b_{fn} can therefore be understood as probabilities of additional childbearing for women who have children of only one sex, relative to those who have children of both sexes. Positive coefficients are evidence of preferences for a sex mix of children over children of one sex only. A significantly positive difference between the two coefficients ($b_{mn}-b_{fn} > 0$) indicates that a woman is more likely to have another birth if she has no sons than if she has no daughters. As in much of the literature (see Keyfitz 1968 and Repetto 1972 for early examples), this is referred to as son-preferred differential fertility-stopping behavior. Though sometimes referred to here as "son preference," the meaning refers exclusively to fertility decisions, as described above, rather than to other possible manifestations of differential behavior toward sons and daughters after birth, as might be evident in differences in mortality, nutritional status, or school enrollment by sex. A negative difference ($b_{mn}-b_{fn} < 0$) indicates daughter preference in childbearing.

Because calculating separate estimates for each pre-existing family size produces a large number of coefficients for b_{mn} and b_{fn}, for most results the focus is on averages across different family sizes—for individual countries or regions and for specific groups (by education, location, wealth, and birth cohort). For this purpose, the means b_m and b_f are defined as follows:

$$b_g = \sum_{n=2}^{\infty} w_{gn} \cdot b_{gn} \quad \text{for } g = m, f \tag{2a}$$

where w_{gn} is the relative weight for family size n (and the weights sum to one). With independence assumed across parities, the corresponding standard error of b_g can also be calculated as follows:

$$s_g = \sqrt{\sum_{n=2}^{\infty} w_{gn}^2 \cdot v_{bgn}} \quad \text{for } g = m, f \tag{2b}$$

where v_{bgn} is the square of the estimated standard error of b_{gn}.[3]

One concern is that including in this analysis women who have not yet completed fertility may bias the results if women who enter childbearing at later ages have different preferences from those who begin childbearing earlier or if birth spacing is partly a function of the sex mix of existing children. To

3. A related alternative approach is to pool all observations at different parities and estimate a model that relates the probability of an additional birth as a function of the share of sons among existing children. Since women appear more than once if they progress beyond three children—for example, a woman with four children would appear twice, once for the transition from two to three children and again from three to four—the model would also include additional controls for the existing family size at each observation. This model can be supplemented with other observable information, such as the location and education of the mother. Analysis of this model serves as a robustness check for the main results and is discussed later.

overcome this problem, the sample is generally limited to women ages 40–49, on the assumption that these women have completed their lifetime fertility (the data do not include women older than 49). To highlight the largely consistent estimates obtained with the two approaches, results based on the entire sample are occasionally compared with those for women ages 40–49.

An important part of the analysis is the exploration of heterogeneity. In addition to heterogeneity by family size, the article explores differences based on location, education, and wealth. In the case of rural or urban location, the following regression is run:

$$B_{wn+1} = a + R_w + b_{mn} \cdot R_w \cdot M_{wn} + b_{fn} \cdot R_w \cdot F_{wn} + c_{mn} \cdot (1 - R_w) \cdot M_{wn} + c_{fn} \cdot (1 - R_w) \cdot F_{wn} + u_{wn} \quad \text{for} n \geq 2 \tag{3}$$

where the R_w is an indicator variable equal to one for women in rural areas; $R_w \cdot M_{wn}$ and $R_w \cdot F_{wn}$ equal one for women in rural areas who have had no sons or no daughters; and $(1 - R_w) \cdot M_{wn}$ and $(1 - R_w) \cdot F_{wn}$ are equal to one for women in urban areas who have had no sons or no daughters. The aggregated coefficients b_m, b_f, c_m, and c_f are reported, along with tests for significant differences between them (based on the formulas in (2a) and (2b)). This arrangement enables testing whether any observed son (or daughter) preference differs in rural and in urban areas by testing whether $(b_m - b_f) = (c_m - c_f)$, a test of difference-in-differences. A similar logic applies to differences by education levels and wealth.

A woman's reported current residential location defines the indicator variable used to test for differences between women in urban and rural areas. To test for differences by education, the indicator variable used splits the sample into those who have completed fewer than six years of schooling and those who have completed six or more. (Six years of schooling corresponds to completing primary school in most countries in the sample.[4]) The analysis by household wealth is based on a composite measure of household durable goods—an approach popularized by Filmer and Pritchett (2001).[5] For each country, the indicator variable divides the sample according to whether the household falls above or below the median household wealth scale.

To investigate whether son-preferred differential fertility-stopping behavior increases or decreases over time across birth cohorts of women, differential

4. A different approach was also used, calculating the median years of education for women in each country and dividing the sample into those above and those below the median. These results were very similar to those reported here.

5. One drawback with this measure is that it reflects household wealth only at the time of the interview, whereas this study considers the full fertility history of each mother—a history that can stretch back 20 years or more. Thus, the wealth index is not an entirely accurate measure of resources available to mothers at the time of decisions about fertility continuation, although there is a positive correlation between current and previous levels of wealth. Considering these interpretive difficulties, this article does not stress the results based on wealth. Early applications of this asset index approach include Pollitt and others (1993) and Rivera and others (1995).

fertility-stopping behavior is calculated within each country for every one-year birth cohort—for example, women in India born in 1945—and then the corresponding regional averages in each year are calculated—for example, for women in South Asia in 1945. A first step is to graph these regional averages. As a more formal test of changes in differential fertility-stopping behavior, separate regressions are run on a set of five-year birth cohort dummy variables by region, to test for differences in these dummy variables. One concern with these estimates is that any observed changes in differential fertility-stopping behavior across birth cohorts could be driven by changes in the countries that make up the regional averages—some countries have surveys only in earlier years and therefore enter only into calculations of regional averages for early birth cohorts, while other countries have surveys only in later years and enter only into regional calculations for later cohorts. Thus, estimates are also presented that keep fixed the countries in each regional sample and the weight given to each in calculating the regional average.

As a final step in the analysis, a multivariate framework is applied based on location–education–cohort cells. This is done primarily because, as shown, prevailing fertility rates have a significant effect on estimated differential fertility-stopping behavior and are correlated with other observable factors. The basic regression is then:

$$(b_m - b_f)_{rht} = \beta_r D_r + \beta_h D_h + \beta_t D_t + \beta_F F_{rht} + u_{rht} \qquad (4)$$

where $(b_m - b_f)_{rht}$ is the measure of differential fertility-stopping behavior, as before, for a given location–education–birth cohort cell; D_r and D_h are dummy variables for women in rural areas and high-education women; D_t is a measure of a woman's birth cohort (in practice, birth cohorts in this part of the analysis are aggregated over three years, to keep the sample sizes reasonable); and F_{rht} is the average number of children born to women in a given location–education–birth cohort cell.[6] The resulting sample includes 3,456 observations for 64 countries. Each country-year contributes four observations corresponding to the four location–education groups for women born in that year. In estimating equation (4), observations are weighted by N, the number of women in each cell. By giving greater weight to cells with larger sample sizes, this method more precisely estimates values of differential fertility-stopping behavior.

Data

Data are from 158 Demographic and Health Surveys (DHS) for the 64 countries listed in the appendix. The data contain the complete retrospective fertility histories of 1.3 million women in the 64 countries, as well as socioeconomic

6. Household wealth is not included in this analysis because of the limitations discussed earlier; however, results are largely unchanged when wealth is included.

information such as educational attainment, ownership of durable goods, and household location.[7]

For comparisons across developing country regions, countries are assigned to geographic regions following World Bank definitions: East Asia and Pacific, Europe and Central Asia, Latin America and the Caribbean, Middle East and North Africa, South Asia, and Sub-Saharan Africa (see the appendix). Note that the countries observed in the East Asia and Pacific region include only countries in Southeast Asia and that those in the Europe and Central Asia region include only countries in Central Asia, and hence these regions are referred to here as Southeast Asia and Central Asia.

In general, observations in each survey are weighted by their expansion factors, which reflect differences in the probability that households are sampled in the DHS.[8] When regional averages are constructed, observations are reweighted so that each country contributes its relative population share to the regional sample; population estimates for 2000 are used.[9] A series of robustness tests show that the findings are largely similar regardless of whether weighted or unweighted regional averages are used.

II. Effects of the Sex-mix Composition of Existing Children on Fertility Behavior

This section presents results for the effects of the sex-mix composition of existing children on fertility behavior by region, mothers' characteristics, mothers' birth cohort, and implications for gender differences in the number of siblings.

Differential Stopping Behavior by Global Region

Table 1 presents the results by region. For each region, the 2+ family size row presents the averages across all family sizes. Although the averages include the results for all family sizes, size-specific coefficients are reported only for family sizes of 2–5 children because the results for higher numbers of children are very noisy and represent less than 5 percent of the total number of births.

7. Supplemental appendix table S1 presents further descriptive statistics for the study populations including total fertility for women ages 40 and older, the mean son–daughter ratio, the percentage of households without a son, the percentage of households without a daughter, and the ratio of reported "ideal" number of sons to "ideal" number of daughters.

8. When a country has more than one survey, all surveys are pooled and the sampling weights are adjusted so that each survey is equally weighted. For example, surveys were administered in Cambodia in 2000 and 2005. To derive the Cambodia database, data from the two surveys were pooled and the survey weights were adjusted so that each survey contributed half the weighted observations to the analysis. Pooling data across surveys enables increasing the number of observations for each country and therefore increases the precision of the estimates.

9. In other words, if one country has twice the population of another in the same region, it will contribute twice the weighted observations to the analysis.

TABLE 1. Differential Fertility-stopping Behavior among Women Ages 40–49 at the Time of the Survey, by Region (Probability of an additional birth as a function of sex-mix composition of existing children)

Region and family size[a]	Probability of additional childbearing after zero sons (b_m; b_{mn})	Probability of additional childbearing after zero daughters (b_f; b_{fn})	Differential fertility-stopping behavior ($b_m - b_f$; $b_{mn} - b_{fn}$)	Significance of difference (p-value)	Mean number of children	Mothers' ideal ratio of sons to daughters[b]
Latin America and Caribbean						
2 +	0.030***	0.019	0.011	0.541	5.08	0.97
2	0.026***	0.016	0.009	0.457		
3	0.020***	0.011	0.009	0.211		
4	0.041***	0.048***	−0.007	0.724		
5	−0.013**	0.048***	−0.061	0.003***		
Middle East and North Africa						
2 +	0.074***	0.016**	0.058	0.000***	6.04	1.13
2	0.018**	0.014***	0.004	0.520		
3	0.037***	0.013	0.024	0.033**		
4	0.037***	0.009	0.028	0.065		
5	0.056**	0.030*	0.026	0.225		
Central Asia						
2 +	0.118***	0.022	0.096	0.000***	4.14	1.02
2	0.089***	0.032***	0.057	0.039**		
3	0.122***	0.011***	0.110	0.001***		
4	0.166***	0.060***	0.106	0.004***		
5	0.168***	0.032	0.136	0.002***		
South Asia						
2 +	0.107***	0.029***	0.078	0.000***	4.94	1.37
2	0.054***	−0.007**	0.060	0.010***		
3	0.107***	0.012	0.095	0.062		
4	0.137***	0.020***	0.116	0.034**		
5	0.142***	0.047***	0.095	0.010**		

Southeast Asia						
2 +	0.052***	0.015	0.037	0.040**	4.74	1.01
2	0.035**	0.016***	0.019	0.354		
3	0.031	0.042***	−0.011	0.785		
4	0.068	0.020**	0.048	0.341		
5	0.099**	0.047***	0.053	0.317		
Sub-Saharan Africa						
2 +	0.024***	0.024***	0.000	0.982	6.63	1.08
2	0.005**	0.002	0.003	0.543		
3	0.012	−0.005	0.017	0.005***		
4	0.021***	0.010	0.011	0.276		
5	0.004	0.010	−0.006	0.740		

**Significant at the 5 percent level; ***significant at the 1 percent level.

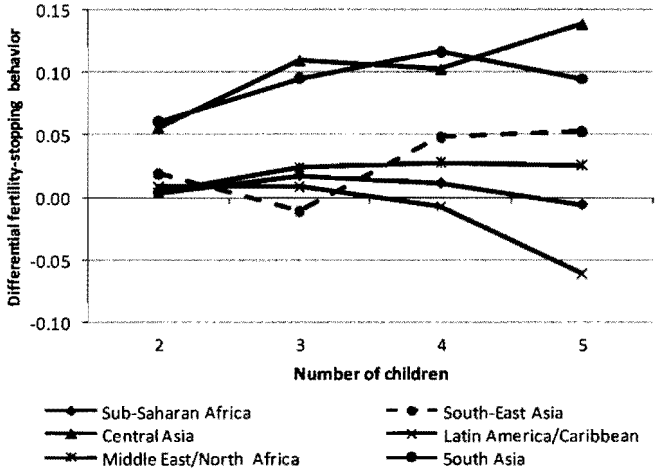
Note: Table reports the estimated probability of an additional birth as a function of having no boys and no girls. Models are estimated at the region level and include country dummy variables. The sample is limited to women ages 40–49, who are most likely to have completed their fertility.

a. Family size 2+ estimates are weighted averages for family sizes of two or more children (see text for details).

b. As reported by mothers to survey enumerators, who routinely ask mothers for their "ideal" number of children, separately for boys and girls. The ratio is the mean desired number of boys divided by the mean desired number of girls.

Source: Authors' analysis of DHS data shown in the appendix.

FIGURE 1. Differential Fertility-stopping Behavior by Region and Parity (Five-year Moving Averages)



Source: Authors' analysis of DHS data shown in the appendix.

The results show clear evidence that many families in all regions in the developing world prefer a mixed-sex composition of children. All the regional averages of b_m and b_f are positive, and many are significant: relative to families with both boys and girls, who are the omitted category in the regressions, families with only boys or only girls are more likely to have another birth.

In addition, the results shows a son-preferred differential fertility-stopping behavior in many regions in the developing world (see table 1, columns 3 and 4). The largest effects are found for Central Asia, where families are 9.6 percentage points more likely to have an additional child if they have had no sons than if they have had no daughters, and South Asia, where the corresponding difference is 7.8 percentage points. Significant, but smaller degrees of son-preferred differential fertility-stopping behavior are apparent in the Middle East and North Africa (5.8 percentage points) and in Southeast Asia (3.7 percentage points). There is no clear evidence of a son-preferred differential in fertility-stopping behavior for either Sub-Saharan Africa or Latin America and the Caribbean.[10]

Because it is difficult to take in all of the coefficients at a glance, the parity-specific results shown in table 1 are summarized in figure 1. Son-preferred differential fertility-stopping behavior appears to grow with the number of children in the two regions where it is most pronounced, Central Asia and South Asia. For example, families in South Asia who have already had four or five

10. Country-specific analyses were also conducted. In the two regions with the clearest evidence of son-preferred differential fertility-stopping behavior (Central Asia and South Asia), these results hold equally for almost all countries in the regions (see supplemental appendix table S2). For the other regions, there is more variability in the country-level results.

children are approximately 14 percentage points more likely to have an additional child if all of their children have been girls rather than boys.

This increase in differential fertility-stopping behavior by number of children is perhaps not surprising: the mean number of children is 4.1 in Central Asia and 4.9 in South Asia. Since the average family expects to have a reasonably large number of children, the sex of children in families with fewer children does not matter as much in determining future fertility because parents expect to have more children, regardless of the sex of their children at the time. In families with more children, however, parents are closer to achieving their total desired number of children, and hence the sex-mix composition of children already born becomes an important determinant of future childbearing. Such patterns are less apparent in the Middle East and North Africa, Southeast Asia, and Latin America, in line with either the smaller degree of son-preferred differential fertility-stopping behavior or the absence of such preference in these regions.[11]

In addition to identifying differences across cohorts in these basic patterns, table 1 is informative about the extent to which the "ideal" balance between the number of boys and girls reported by mothers is a good indication of fertility behavior. This can be seen by comparing columns 3 and 6 of table 1. A clear subjective preference for sons is apparent in South Asia and Middle East and North Africa, as is a clear behavioral preference for sons with regard to the decision to continue child bearing. However, another region that exhibits a significant pattern of son-preferred differential fertility-stopping behavior, Central Asia, reports a subjective preference for a near equality of sons and daughters. In contrast, mothers in Sub-Saharan Africa report a subjective preference for sons, but families do not exhibit son preference in actual fertility behavior.[12] In Latin America and the Caribbean, mothers express a slight preference for daughters,

11. Given the preferred parameterization—binary controls for "no sons" and "no daughters"—aggregating results for family sizes of one child with those of family sizes of two or more children would create an inconsistency. With a family size of one child, the model can include only one dummy variable (either "no sons" or "no daughters"). The two models would need to be estimated separately, and the coefficients on the two variables would merely be transformations of one another. The excluded category in these models would be a family with one son or one daughter. This is unlike the main estimations, where families with children of at least one of *each* sex serve as the excluded group. The interpretation is therefore slightly different, and so families with only one child are not included in the analysis. A related model was estimated, however, that investigates the probability of an additional birth, controlling for the sex of the first child. Supplemental appendix table S3 reports these results, which also show son-preferred differential fertility-stopping behavior in South Asia even for decisions after the first child. However, the analysis shows that families in Latin America are significantly more likely to stop child bearing after the first birth if that birth is a daughter rather than a son.

12. The lack of observed differential fertility-stopping behavior in Sub-Saharan Africa could be due to several factors, but one important factor is surely the high level of fertility. Completed fertility in Sub-Saharan Africa is by far the highest and the proportion of households with children of only one sex the lowest across all regions. However, supplemental appendix table S1 also suggests that there is wide variation within Sub-Saharan Africa in the ratio of "ideal" number of sons to "ideal" number of daughters. Therefore, to the extent that reported "ideal" ratio reflects latent sex preference in family composition, Sub-Saharan Africa is not a uniformly son-preferring region, unlike, say, South Asia.

but actual fertility behavior exhibits no distinct pattern. Clearly, subjectively stated preferences over the sex-mix composition of children more accurately predict actual fertility behavior in some regions than in others.[13]

Table 2 presents a series of robustness tests to these basic findings, focusing on the aggregate effects averaged across all family sizes (number of children). The first panel uses the number of women ages 40–49 as the weight for aggregating across countries within regions rather than the total population of a country. These weights are generated using data on the share of women ages 40–49 and applying these estimates to estimates of the total female population.[14] The stability of the results to this alternative approach to weighting is apparent. The only major difference between this first panel and table 1 is that the slight son-preferred differential fertility-stopping behavior found in East Asia is no longer statistically significant.

The results are similar if instead of giving greater weight to countries with larger populations, only the expansion factors in the surveys are used (see table 2, second panel). The only difference is that now son-preferred differential fertility-stopping behavior is slightly muted in South Asia—a difference between b_m and b_f of 4.6 percentage points compared with 7.8 percentage points in table 1. The results are still similar if even these survey weights are disregarded, so that each sample observation in each region is given the same weight (third panel). If anything, these results suggest an even greater degree of son-preferred differential fertility-stopping behavior in Central Asia and South Asia than do the results in table 1. Moreover, finally, son-preferred differential fertility-stopping behavior continues to be apparent in the three regions where it is most pronounced in table 1—Middle East and North Africa, Central Asia, and South Asia—when all women ages 15–49 at the time of the survey are included, not just women who are most likely to have completed their fertility (fourth panel).[15]

Differential Fertility-stopping Behavior by Mothers' Characteristics

This section investigates how the strong son-preferred differential fertility-stopping behavior exhibited in some regions varies across common

13. Supplemental appendix table S4 reports the alternative specification mentioned earlier that pools the parity-specific data and estimates differential fertility-stopping behavior as a function of the ratio of sons to total number of children, controlling for family size. Similar to table 1 in this article, this analysis finds significant son-preferred differential fertility-stopping behavior in the Middle East and North Africa, Central Asia, and South Asia, suggesting that the article's main findings are robust to this alternative measure of differential fertility-stopping behavior. The son-preferred differential fertility-stopping behavior estimates in these three regions actually grow in magnitude when select mothers' observables such as location, education, and age are also controlled for. These results with covariates are presented in the second panel of Supplemental appendix table S4.

14. Both statistical constructs are from a World Bank database accessed at: http://go.worldbank.org/N2N84RDV00.

15. Of course, since this panel includes all women, not just those who have completed their fertility, the total number of children is lower in all regions.

TABLE 2. Differential Fertility-stopping Behavior among Women at the Time of the Survey, with Different Weights, by Region (Probability of an additional birth as a function of sex-mix composition of existing children)

Region	Probability of additional childbearing after zero sons (b_m)	Probability of additional childbearing after zero daughters (b_f)	Differential fertility-stopping behavior ($b_m - b_f$)	Significance of difference (p-value)	Mean number of children	Mothers' ideal ratio of sons to daughters[a]
Women ages 40–49, population of women ages 40–49 adjusted weights						
Latin America and Caribbean	0.030***	0.020	0.011	0.545	5.01	0.97
Middle East and North Africa	0.076***	0.016**	0.061	0.000***	5.99	1.13
Central Asia	0.120***	0.023	0.097	0.000***	4.07	1.02
South Asia	0.109***	0.028***	0.081	0.000***	4.89	1.37
Southeast Asia	0.051***	0.021	0.030	0.115	4.74	1.01
Sub-Saharan Africa	0.023**	0.024***	−0.001	0.925	6.52	1.08
Women ages 40–49, population-unadjusted weights						
Latin America and Caribbean	0.018	0.018	0.000	0.984	5.31	0.93
Middle East and North Africa	0.072***	0.016**	0.057	0.000***	6.46	1.10
Central Asia	0.133***	0.049***	0.084	0.001***	3.77	1.03
South Asia	0.080***	0.034***	0.046	0.001***	5.45	1.41
Southeast Asia	0.055***	0.017	0.038	0.048**	4.84	0.99
Sub-Saharan Africa	0.032***	0.017**	0.015	0.165	6.62	1.04
Women ages 40–49, no weights						
Latin America and Caribbean	0.031***	0.031***	0.000	0.977	5.17	0.92
Middle East and North Africa	0.075***	0.013***	0.061	0.000***	5.82	1.15

(Continued)

TABLE 2. Continued

Region	Probability of additional childbearing after zero sons (b_m)	Probability of additional childbearing after zero daughters (b_f)	Differential fertility-stopping behavior ($b_m - b_f$)	Significance of difference (p-value)	Mean number of children	Mothers' ideal ratio of sons to daughters[a]
Central Asia	0.150***	0.017	0.133	0.000***	3.77	1.05
South Asia	0.119***	0.025***	0.094	0.000***	4.67	1.34
Southeast Asia	0.044***	0.020***	0.024	0.020**	4.95	0.99
Sub-Saharan Africa	0.025***	0.019***	0.006	0.482	6.73	1.06
Full sample of women, population-adjusted weights						
Latin America and Caribbean	0.042***	0.026***	0.016	0.134	5.08	0.95
Middle East and North Africa	0.063***	0.020***	0.043	0.000***	6.04	1.12
Central Asia	0.124***	0.037***	0.087	0.000***	4.14	1.03
South Asia	0.102***	0.013***	0.089	0.000***	4.94	1.35
Southeast Asia	0.046***	0.023***	0.023	0.100	4.74	1.01
Sub-Saharan Africa	0.018***	0.021***	−0.003	0.609	6.63	1.09

**Significant at the 5 percent level; ***significant at the 1 percent level.

Note: Table reports the estimated probability of an additional birth as a function of having no boys and no girls. Models are estimated at the region level and include country dummy variables. Estimates are for families with three or more children (see text for details).

a. As reported by mothers to survey enumerators, who routinely ask mothers for their "ideal" number of children, separately for boys and girls. The ratio is the mean desired number of boys divided by the mean desired number of girls.

Source: Authors' analysis of DHS data shown in the appendix.

measures of "modernization"—rural–urban location, education, and wealth. Although results are reported for all regions, the discussion focuses on Central Asia and South Asia, where the aggregate results show the greatest son-preferred differential fertility-stopping behavior.

The patterns are somewhat different in the two regions. In both South Asia and Central Asia, there is son-preferred differential fertility-stopping behavior in both urban and rural regions, among more and less educated women, and among both households with more and those with less wealth (table 3, columns 3 and 7). However, the difference-in-difference results suggest that in South Asia son-preferred differential fertility-stopping behavior is higher in urban than in rural areas (although not significantly so), among women with more education levels than those with less, and in households with more wealth than in those with less. Some of the differences are quite large: For example, women with six or more years of schooling are 19 percentage points more likely to have an additional child if they do not have boys than if they do not have girls (column 3), while women with less than six years of schooling are only 7 percentage points more likely to do so (column 7).[16] In Central Asia, the picture is more mixed: Son-preferred differential fertility-stopping behavior is also higher in urban than in rural areas, but higher among women with low levels of education than among those who have completed at least primary school. Further, there is no significant difference among households in Central Asia at different wealth levels.

Many express the belief that as societies and economies develop, the traditional social practices that may enforce or perpetuate a preference for sons weaken. This could happen, for example, if women gain greater autonomy and control a greater share of the household's economic resources (see, for example, the discussions in Haddad, Hoddinot, and Alderman 1997). Under this assumption, greater son-preferred differential fertility-stopping behavior might be expected in rural than in urban areas, among women with less education, and among poorer women. The results here do not support that, however, either overall or for regions in which son preference is most pronounced (see table 3). This is consistent with earlier findings of greater male preference in Indian households with more educated household heads (Behrman 1988).

Differential Fertility-stopping Behavior over Time

To examine changes across birth cohorts, differential fertility-stopping behavior is calculated for each regional cohort cell, as described above. The results

16. Women who are educated or live in urban areas potentially have greater access to technologies that allow them to select the sex of a child. This might affect a small number of the women in the sample (those in the latest cohorts in some countries). However, the effect on estimated differential fertility-stopping behavior is not clear since differential fertility-stopping behavior is by definition a behavior conditional on the existing sex mix of children, regardless of whether that mix arose through natural means or with the assistance of sex-selective technology.

TABLE 3. Differential Fertility-stopping Behavior by Select Mother or Household Characteristics for Women Ages 40–49, by Region (Probability of an additional birth as a function of sex-mix composition of existing children)

Region	Probability of additional childbearing after zero sons (b_m)	Probability of additional childbearing after zero daughters (b_f)	Differential fertility-stopping behavior (b_m-b_f)	Mean number of children	Probability of additional childbearing after zero sons (b_m)	Probability of additional childbearing after zero daughters (b_f)	Differential fertility-stopping behavior (b_m-b_f)	Mean number of children	Difference-in-difference (column 3–column 7)
	Urban				*Rural*				*Difference*
Latin America and Caribbean	0.041***	0.049***	−0.009	4.46	0.044**	−0.011	0.055	6.05	−0.064
Middle East and North Africa	0.048***	0.009	0.039***	5.08	0.076***	0.019	0.057***	6.94	−0.018
Central Asia	0.125***	0.033**	0.091***	3.55	0.098***	0.036**	0.063***	5.07	0.028
South Asia	0.137***	0.032***	0.105***	4.27	0.098***	0.026***	0.072***	5.22	0.033
Southeast Asia	0.077***	0.023**	0.054***	4.29	0.042**	0.013	0.029	4.94	0.025
Sub-Saharan Africa	0.041***	0.030**	0.012	5.55	0.019**	0.023**	−0.004	7.05	0.016
		Six or more years of schooling				*Less than six years of schooling*			*Difference*
Latin America and Caribbean	−0.003	0.063***	−0.066***	3.46	0.031***	0.006	0.025	5.91	−0.090**
Middle East and North Africa	0.109***	0.044***	0.064***	3.78	0.074***	0.011	0.062***	6.57	0.002
Central Asia	0.107***	0.046***	0.061***	3.64	0.136***	−0.001	0.137***	4.65	−0.076**
South Asia	0.198***	0.004	0.193***	3.32	0.094***	0.029***	0.066***	5.35	0.128**
Southeast Asia	0.062***	0.020	0.042**	4.20	0.049***	0.023	0.026	5.19	0.017
Sub-Saharan Africa	0.047***	−0.007	0.054**	5.10	0.019	0.027***	−0.008	7.05	0.062**
		Above-median-wealth households[a]				*Below-median-wealth households*[a]			*Difference*
Latin America and Caribbean	0.020	0.043**	−0.023	3.55	0.056***	0.053***	0.003	5.07	−0.026
Middle East and North Africa	0.042***	0.037***	0.005	5.17	0.040**	0.008	0.032	6.55	−0.027
Central Asia	0.119***	0.028	0.091***	3.66	0.116***	0.027	0.089***	4.67	0.002
South Asia	0.144***	0.028***	0.116***	4.43	0.086***	0.026**	0.060***	5.54	0.056**
Southeast Asia	0.079***	0.036***	0.043	4.23	0.042**	−0.003	0.045**	4.98	−0.002
Sub-Saharan Africa	0.033***	0.008	0.025	6.31	0.026**	0.019	0.007	6.62	0.018

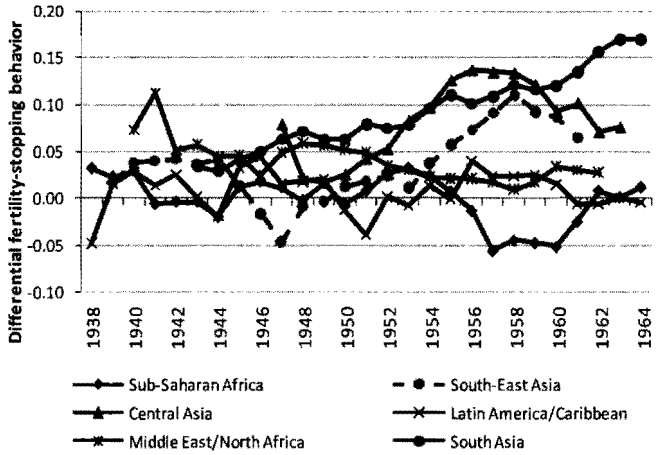
**Significant at the 5 percent level; ***significant at the 1 percent level.

Note: Table reports the estimated probability of an additional birth as a function of having no boys and no girls. Models are estimated at the region level and include country dummy variables. Estimates are for families with two or more children (see text for details).

a. The analysis by household wealth is based on a composite measure of household durable goods, with households categorized as above or below the median of a composite measure of assets.

Source: Authors' analysis of DHS data shown in the appendix.

FIGURE 2. Differential Fertility-stopping Behavior by Region and Mother's Year of Birth (Five-year Moving Averages)



Source: Authors' analysis of DHS data shown in the appendix.

are summarized in figure 2, which shows the five-year moving average of differential fertility-stopping behavior by region. In most regions, there is no systematic pattern. In South Asia, however, son-preferred differential fertility-stopping behavior increases across birth cohorts and is almost 15 percentage points higher for the latest birth cohorts than for the earliest ones. The other region with a high degree of son preference, Central Asia, shows an initial increase in son-preferred differential fertility-stopping behavior, followed by a decrease, although the absolute levels remain high throughout.

To test whether these changes across birth cohorts are significant, differential fertility-stopping behavior is first regressed on a linear cohort trend, separately by region. Each observation is weighted by the number of women in that cohort-year cell, which gives greater weight to the more precisely calculated cell averages. The coefficient on the cohort trend in this regression for South Asia is highly significant (0.007, with a standard error of 0.002), which suggests that son-preferred differential fertility-stopping behavior has been increasing by about 0.7 percentage points with each successive cohort. The corresponding coefficient for Southeast Asia is also significant (0.005, with a standard error of 0.002). None of the other coefficients is close to standard levels of significance.

There are two potential problems with figure 2 and the corresponding regression analysis. The first is that a linear cohort trend may not do justice to the data; this is particularly apparent for Central Asia, with its inverted U-shaped pattern. To address this concern, differential fertility-stopping behavior is regressed on five-year birth cohort dummy variables, again separately by region. The results—the regression analog of the pattern observed in figure 2—again show the clearest pattern for South Asia, where son-preferred differential

fertility-stopping behavior rises monotonically across five-year birth cohorts (table 4). The increase is 10-fold, from 0.017 for the cohort born in 1941–45, to 0.170 for the cohort born in 1961–65.

The second, more difficult problem is that the regional averages for different birth cohorts may be driven by different countries, depending on the years in which they conducted the DHS. For example, the data from Sri Lanka, where the only DHS was carried out in 1987, enters the average for South Asia for the early birth cohorts but not for the later ones, while the data for Nepal, where DHS were carried out in 1996, 2001, and 2006, enters the regional averages for the later birth cohorts, but not the earlier ones. To address this concern, the sample was limited to countries with a DHS both in 1995 or earlier and in 2000 or later. This greatly reduces the number of countries, from 65 to 27. However, cohort-specific measures of son-preferred differential fertility-stopping behavior can be calculated for these countries for women born in every year between 1945 and 1960, and thus regional averages can be calculated that keep the weights fixed for each country across birth cohorts. (The sample is limited to women ages 40 and older, as before.)

When both the sample of countries and the weight of each country in the regional average are kept fixed, son-preferred differential fertility-stopping behavior still increases across birth cohorts in South Asia, although the pattern is less dramatic and the difference across cohorts is no longer significant (see table 4, bottom panel). In other regions, the patterns are less clear and are generally not significant. What is clear is that there is no decline in son-preferred differential fertility-stopping behavior in any region where it exists for yet another standard measure of modernization—the passage of time.

A Simple Multivariate Framework

The sociodemographic characteristics explored in table 3—mother's education, urban location, and household wealth—are likely correlated with each other. Thus, it is possible that the association between son-preferred differential fertility-stopping behavior and each of these characteristics is really driven by one main social indicator. Furthermore, prevailing fertility levels may have an effect on differential fertility-stopping behavior since in a high-fertility environment fewer families face differential stopping decisions because of the greater likelihood of mixed-sex composition at larger family sizes. This section thus uses the aggregated location–education–cohort cell data described earlier to estimate the multivariate framework given by equation (4).

In bivariate regressions, urban residence and higher educational attainment are both associated with higher differential fertility-stopping behavior, although not significantly so (table 5, columns 1 and 2). These results are consistent with those in table 3. In addition, however, there is a significant negative association between the average number of children and differential fertility-stopping behavior (column 3)—the point estimate implies that

TABLE 4. Differential Fertility-stopping Behavior Regressed on Region Interacted with Five-year Cohorts of Mother Birth Year, for Women Ages 40–49, by Region

Region	Mothers' birth year cohort	Region-cohort interaction[a]	F-test[b] All interactions equal	F-test[b] First and last equal
All countries for cohorts 1941–65				
Latin America and Caribbean	1941–45	−0.004	0.784	0.904
	1946–50	0.013		
	1951–55	−0.009		
	1956–60	0.025		
	1961–65	0.001		
Middle East and North Africa	1941–45	0.062	0.851	0.733
	1946–50	0.055		
	1951–55	0.031		
	1956–60	0.010		
	1961–64	0.040		
Central Asia	1946–50	0.017	0.412	0.403
	1951–55	0.085**		
	1956–60	0.141***		
	1961–65	0.094		
South Asia	1941–45	0.017	0.001***	0.000***
	1946–50	0.067***		
	1951–55	0.078***		
	1956–60	0.120***		
	1961–65	0.170***		
Southeast Asia	1941–45	0.024	0.027**	0.874
	1946–50	0.002		
	1951–55	0.013		
	1956–60	0.108***		
	1961–63	0.033		
Sub-Saharan Africa	1941–45	−0.001	0.025**	0.895
	1946–50	0.000		
	1951–55	0.034		
	1956–60	−0.047***		
	1961–65	−0.006		
Countries with differential fertility-stopping behavior for cohorts 1946–60[c]				
Latin America and Caribbean	1946–50	0.020	0.410	0.491
	1951–55	−0.020		
	1956–60	0.000		
Middle East and North Africa	1946–50	0.050	0.593	0.311
	1951–55	0.024		
	1956–60	0.010		
Central Asia	1946–50	0.084	0.710	0.456
	1951–55	0.147***		

(Continued)

TABLE 4. Continued

Region	Mothers' birth year cohort	Region-cohort interaction[a]	F-test[b] All interactions equal	F-test[b] First and last equal
	1956–60	0.148***		
South Asia	1946–50	0.093***	0.219	0.275
	1951–55	0.080***		
	1956–60	0.120***		
Southeast Asia	1946–50	0.007	0.124	0.615
	1951–55	−0.038		
	1956–60	0.024		
Sub-Saharan Africa	1946–50	0.018	0.042**	0.037**
	1951–55	0.016		
	1956–60	−0.035**		

**Significant at the 5 percent level; ***significant at the 1 percent level.

a. The results in this column are the coefficients of the interaction terms.

b. The *F*-tests are region specific. The results are the *p*-values for the *F*-tests. Data are weighted by sample size.

c. Countries include Bangladesh, Bolivia, Burkina Faso, Cameroon, Colombia, Côte d'Ivoire, Dominican Republic, Egypt, Ghana, Haiti, India, Indonesia, Kenya, Madagascar, Malawi, Mali, Morocco, Namibia, Niger, Nigeria, Peru, Philippines, Rwanda, Senegal, Tanzania, Turkey, Uganda, Zambia, and Zimbabwe.

Source: Authors' analysis of DHS data shown in the appendix.

a decrease in average family size of one child more than offsets a switch from rural to urban location and almost offsets a switch from low to high schooling levels.

The key results include the measures of location, education, and the mean number of children for each country, year, location, and education cell (see table 5, columns 4 and 5). Once the average number of children is included in the model, the association between son-preferred differential fertility-stopping behavior and urban residence and between differential fertility-stopping behavior and education becomes negative (column 4). This reverses the bivariate findings and suggests that the higher son-preferred differential fertility-stopping behavior in urban areas and among more educated mothers can be "explained" by differences in overall fertility levels.[17] Including global dummy variables for each birth year, as a way of flexibly controlling for any secular changes, barely affects the results for these three indicators (column 5).

In sum, the cell-level results suggest that the number of children women expect to have over their lifetimes is an important determinant of son-preferred differential fertility-stopping behavior. When fertility levels are high, the

17. This finding is in character with Das Gupta and Mari Bhat (1997), who argue that fertility decline may lead to an intensification of discrimination against girls if the total number of children that couples desire falls more rapidly than the total number of desired sons.

TABLE 5. Multivariate Correlates of Differential Fertility-stopping Behavior

	Regression				
Variable	(1)	(2)	(3)	(4)	(5)
Urban	0.014			−0.023**	−0.021**
	(0.010)			(0.010)	(0.010)
Six or more years of schooling		0.027		−0.026***	−0.022***
		(0.020)		(0.009)	(0.009)
Mean number of children			−0.021*	−0.029**	−0.027**
			(0.011)	(0.013)	(0.012)
Birth year dummy variables	No	No	No	No	Yes
Number of observations	3,456	3,456	3,456	3,456	3,456
R-squared	0.00	0.01	0.04	0.05	0.06

*Significant at the 10 percent level; **significant at the 5 percent level; ***significant at the 1 percent level.

Note: Numbers in parentheses are robust standard errors. Each observation is a country, urban–rural, high–low education, year of birth cell. Data are weighted by sample size and country population in 2000.

Source: Authors' analysis of DHS data shown in the appendix.

absence of boys in earlier births is not an important driver of childbearing decisions—at all but the largest family size, most couples expect to have more children, no matter what the sex-mix composition of earlier births. However, as family size decreases, a higher fraction of couples find themselves having to choose whether to have an additional child at a point when they are already close to their expected family size and all their children are of the same sex. At this point, the sex-mix composition of their children—in particular, whether there is at least one boy—appears to play an important role in their decision.

Sex Differences in Number of Siblings

If families are more likely to have an additional child when they have no sons than when they have no daughters, girls may grow up in households with more siblings than do boys. Of course, the number of siblings that boys or girls have will also be determined by mortality—which may vary with family size and by a child's sex.

The mean number of siblings for girls and boys ages 0–15 years is higher for girls than for boys in regions where there is son-preferred differential fertility-stopping behavior (table 6). For example, in South Asia girls have about 0.13 more siblings than boys, on average; in Central Asia, the comparable number is 0.10. In contrast, in Sub-Saharan Africa, boys and girls have the same number of siblings on average. Moreover, if girls are discriminated against relative to boys after birth in regions where there is son-preferred differential fertility-stopping behavior, like South Asia and Central Asia, and

TABLE 6. Mean Number of Siblings of Children ages 0–15

	Children of women ages 40 and older			All children		
Region	Sons	Daughters	Sons–daughters	Sons	Daughters	Sons–daughters
Latin America and Caribbean	4.99	5.06	−0.07***	3.08	3.14	−0.06***
Middle East and North Africa	5.27	5.29	−0.02	3.67	3.73	−0.06***
Central Asia	4.27	4.37	−0.10**	2.63	2.77	−0.14***
South Asia	4.59	4.72	−0.13***	2.81	2.96	−0.15***
Southeast Asia	4.46	4.52	−0.07***	2.82	2.86	−0.04***
Sub-Saharan Africa	5.49	5.49	0.01	3.55	3.56	−0.01**

**Significant at the 5 percent level; ***significant at the 1 percent level.

Source: Authors' analysis of DHS data shown in the appendix.

therefore suffer excess mortality,[18] these results would generally underestimate the differences in sibship size by sex that result from son-preferred differential fertility-stopping behavior.

An extensive literature documents associations between larger family size and poorer outcomes for children in developed and developing countries (see, for example, Behrman and Wolfe 1986; Horton 1986; Conley and Glauber 2006, and the references therein). Having more siblings dilutes household and parental resources and may result in quantity–quality tradeoffs. Estimating the causal effect of the number of siblings on child outcomes is difficult, however, because of the likelihood of omitted family characteristics that may bias results. Nevertheless, insofar as some of the association between the number of children and poor outcomes is causal, it suggests that son preference, as manifested in sex-specific differential fertility-stopping behavior, may have adverse implications on the outcomes for girls, who will tend to grow up in larger families. Moreover, the differences in family size by children's sex are largest in regions where girls are more likely to suffer discrimination in other ways, in particular in South Asia (see table 6).

III. CONCLUSION

This article has investigated the fertility response to the sex-mix composition of children in a family using data from 158 DHS carried out in 64 countries. Sex composition of earlier births is a significant determinant of subsequent fertility in many developing countries. Fertility behavior is consistent with son preference in many regions of the developing world, with the clearest patterns apparent in South Asia and Central Asia. Specifically, the absence of sons increases

18. On India, see, for example, Das Gupta (1987), Behrman and Deolalikar (1990), and Rose (1999).

the probability of an additional birth by significantly more than the absence of daughters. This phenomenon is referred to as son-preferred differential fertility-stopping behavior.

Exploration of heterogeneity shows that widely used measures of "modernization," including urbanization, higher education levels, and household wealth, are associated with an increase in son-preference, as captured in differential fertility-stopping behavior. The presumption that this manifestation of son preference will dissipate over time is also not supported by the data. The results from regressions using a simple multivariate framework suggest that this may be a result of reductions in family size with increased modernization. While it is possible that greater urbanization, female education, and household wealth all reduce a latent son preference, the reductions in fertility that accompany modernization also make it more likely that a latent son preference can be detected in behavior. For this reason, social policies that aim to limit fertility may, as an unintended consequence, bring son-preferred differential fertility-stopping behavior to the fore.

Finally, one implication of son-preferred differential fertility-stopping behavior is that girls tend to have more siblings than boys. This is an important finding in itself, as it likely has consequences for the development of boys and girls in infancy, childhood, and adolescence. Moreover, insofar as there are quantity–quality tradeoffs that result in fewer material and emotional resources allocated to children in larger families, son preference in fertility decisions can have important indirect implications for investments and for the well-being of girls relative to boys.

Supplementary Material

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

Appendix: Sample Countries, Surveys, and Number of Mothers and Births

Country	Region	Year of survey	Number of mothers observed	Number of births observed
Armenia	Central Asia[a]	2000, 2005	8,648	21,583
Bangladesh	South Asia	1993–94, 1996–97, 1999–2000, 2004	36,169	127,486
Benin	Sub-Saharan Africa	1996, 2001, 2006	22,688	95,989

(Continued)

Continued

Country	Region	Year of survey	Number of mothers observed	Number of births observed
Bolivia	Latin America and Caribbean	1989, 1993–94, 1998, 2003–04	31,431	121,101
Brazil	Latin America and Caribbean	1986, 1991–92, 1996	12,050	37,871
Burkina Faso	Sub-Saharan Africa	1992–93, 1998–99, 2003	19,168	84,320
Burundi	Sub-Saharan Africa	1987	2,777	11,886
Cambodia	Southeast Asia[b]	2000, 2005	20,721	81,447
Cameroon	Sub-Saharan Africa	1991, 1998, 2004	14,243	56,254
Central African Republic	Sub-Saharan Africa	1994–95	4,388	16,936
Chad	Sub-Saharan Africa	1996–97, 2004	10,508	47,187
Colombia	Latin America and Caribbean	1986, 1990, 1995, 2000, 2005	50,573	141,967
Comoros	Sub-Saharan Africa	1996	1,695	7,913
Congo, Rep. of	Sub-Saharan Africa	2005	5,152	16,687
Côte d'Ivoire	Sub-Saharan Africa	1994, 1998–99, 2005	11,895	45,803
Dominican Republic	Latin America and Caribbean	1986, 1991, 1996, 1999, 2002	33,677	113,636
Ecuador	Latin America and Caribbean	1987	3,117	11,835
Egypt	Middle East and North Africa	1988, 1992–93, 1995–96, 2000, 2003, 2005	70,394	276,509
Ethiopia	Sub-Saharan Africa	2000, 2005	19,482	84,055
Gabon	Sub-Saharan Africa	2000–2001	4,499	16,878
Ghana	Sub-Saharan Africa	1988, 1993–94, 1998–99, 2003	14,449	55,788
Guatemala	Latin America and Caribbean	1987, 1995, 1998–99	16,804	72,032
Guinea	Sub-Saharan Africa	1999, 2005	11,672	50,058
Haiti	Latin America and Caribbean	1994–95, 2000, 2005	16,294	63,814
Honduras	Latin America and Caribbean	2005	13,991	50,093
India	South Asia	1992–93, 1998–2000, 2005–06	244,831	800,833

(Continued)

Continued

Country	Region	Year of survey	Number of mothers observed	Number of births observed
Indonesia	Southeast Asia[b]	1987, 1991, 1994, 1997, 2002–03	111,864	370,441
Kazakhstan	Central Asia[a]	1995, 1999	6,013	14,972
Kenya	Sub-Saharan Africa	1988–89, 1993, 1998, 2003	22,504	94,497
Kyrgyzstan	Central Asia[a]	1997	2,776	8,781
Lesotho	Sub-Saharan Africa	2004	4,832	14,708
Liberia	Sub-Saharan Africa	1986	4,231	17,264
Madagascar	Sub-Saharan Africa	1992, 1997, 2003–04	15,447	61,383
Malawi	Sub-Saharan Africa	1992, 2000, 2004	23,353	92,634
Mali	Sub-Saharan Africa	1987, 1995–96, 2001	21,004	98,580
Mexico	Latin America and Caribbean	1987	5,776	22,676
Morocco	Middle East and North Africa	1987, 1992, 2003–04	18,970	80,669
Mozambique	Sub-Saharan Africa	1997, 2003	16,530	63,195
Namibia	Sub-Saharan Africa	1992, 2000	8,490	28,318
Nepal	South Asia	1996, 2001, 2006	23,042	84,505
Nicaragua	Latin America and Caribbean	1997–98, 2001	18,971	70,977
Nigeria	Sub-Saharan Africa	1990, 1999, 2003	17,209	74,438
Niger	Sub-Saharan Africa	1992, 1998, 2006	18,194	87,107
Pakistan	South Asia	1990–91	5,905	27,369
Paraguay	Latin America and Caribbean	1990	3,970	153,46
Peru	Latin America and Caribbean	1986, 1991–92, 1996, 2000, 2004	60,700	217,275
Philippines	Southeast Asia[b]	1993, 1998, 2003	26,609	98,932
Rwanda	Sub-Saharan Africa	1992, 2000, 2005	17,876	771,14
Senegal	Sub-Saharan Africa	1986, 1992–93, 1997, 2005	23,525	102,547
South Africa	Sub-Saharan Africa	1998	8,223	22,934
Sri Lanka	South Asia	1987	5,388	17,701
Sudan	Sub-Saharan Africa	1989–90	5,277	25,805
Tanzania	Sub-Saharan Africa	1991–92, 1996, 1999, 2004	23,504	96,542

(Continued)

Continued

Country	Region	Year of survey	Number of mothers observed	Number of births observed
Thailand	Southeast Asia[b]	1987	6,025	17,803
Togo	Sub-Saharan Africa	1988, 1998	8,825	37,051
Trinidad and Tobago	Latin America and Caribbean	1987	2,440	7,837
Tunisia	Middle East and North Africa	1988	3,856	16,463
Turkey	Central Asia[a]	1993, 1998, 2003	18,861	59,996
Uganda	Sub-Saharan Africa	1988–89, 1995, 2000–2001, 2006	20,946	92,326
Uzbekistan	Central Asia[b]	1996	3,018	96,50
Vietnam	Southeast Asia[b]	1997, 2002	10,742	29,900
Yemen	Middle East and North Africa	1991–92	5,378	29,803
Zambia	Sub-Saharan Africa	1992, 1996–97, 2001–02	17,013	70,726
Zimbabwe	Sub-Saharan Africa	1988–89, 1994, 1999, 2005–06	17,881	62,855
64 countries	6 regions	158 surveys	1,336,484	4,931,081

a. None of the countries observed in this region is in the part of the region traditionally referred to as Eastern Europe, and so this region is referred to in the analysis as Central Asia only.

b. None of the countries observed in this region is in the part of the region traditionally referred to as the Pacific or in the Northeastern region of Asia, and so this region is referred to in the analysis as Southeast Asia only.

References

Andersson, G., H. Karsten, M. Rønson, and A. Vikat. 2006. "Gendering Family Composition: Sex Preferences for Children and Childbearing Behavior in the Nordic Countries." *Demography* 42(2):255–67.

Arnold, F. 1985. "Measuring the Effect of Sex Preference on Fertility: The Case of Korea." *Demography* 22(2):280–88.

———. 1992. "Sex Preference and Its Demographic and Health Implications." *International Family Planning Perspectives* 18(3):93–101.

———. 1997. *Gender Preferences for Children.* DHS Comparative Studies 23. Calverton, Md.: Macro International.

Arnold, F., M. K. Choe, and T. K. Roy. 1998. "Son Preference, the Family-building Process, and Child Mortality in India." *Population Studies* 52(3):301–15.

Bairagi, R. 1987. "A Comment on Fred Arnold's "Measuring the Effect of Sex Preference on Fertility." *Demography* 24(1):137–38.

Behrman, J. R. 1988. "Intrahousehold Allocation of Nutrients in India: Are Boys Favored? Do Parents Exhibit Inequality Aversion?" *Oxford Economic Papers* 40(1):32–54.

Behrman, J. R., and A. B. Deolalikar. 1990. "The Intrahousehold Demand for Nutrients in Rural South India: Individual Estimates, Fixed Effects, and Permanent Income." *The Journal of Human Resources* 25(4):665–96.

Behrman, J. R., and B. L. Wolfe. 1986. "Child Quantity and Quality in a Developing Country: Family Background, Endogenous Tastes, and Biological Supply Factors." *Economic Development and Cultural Change* 34(4):703–20.

Chung, W., and M. Das Gupta. 2007. "The Decline of Son Preference in South Korea: The Roles of Development and Public Policy." *Population and Development Review* 33(4):757–83.

Conley, D., and R. Glauber. 2006. "Parental Educational Investment and Children's Academic Risk: Estimates of the Impact of Sibship Size and Birth Order from Exogenous Variation in Fertility." *Journal of Human Resources* 41(4):722–37.

Das Gupta, M. 1987. "Selective Discrimination against Female Children in Rural Punjab, India." *Population and Development Review* 13(1):77–100.

Das Gupta, M., and P. N. Mari Bhat. 1997. "Fertility Decline and Increased Manifestation of Sex Bias in India." *Population Studies* 51(4):307–15.

Drèze, J., and M. Murthi. 2001. "Fertility, Education, and Development: Evidence from India." *Population and Development Review* 27(1):33–63.

Filmer, D. 2005. "Gender and Wealth Disparities in Schooling: Evidence from 44 Countries." *International Journal of Education Research* 43(6):351–69.

Filmer, D., and L. Pritchett. 2001. "Estimating Wealth Effects without Income or Expenditure Data—or Tears: Educational Enrollment in India." *Demography* 38(1):115–32.

Haddad, L., J. Hoddinot, and H. Alderman. 1997. *Intrahousehold Resource Allocation in Developing Countries: Models, Methods, and Policy.* Baltimore: Johns Hopkins University Press.

Hank, K., and H.-P. Kohler. 2000. "Gender Preferences for Children in Europe: Empirical Results from 17 FFS Countries." *Demographic Research* 2 (Article1). www.demographic-research.org/Volumes/Vol2/1/2-1.pdf.

Haughton, J., and D. Haughton. 1995. "Son Preference in Vietnam." *Studies in Family Planning* 26(6):325–37.

Horton, S. 1986. "Child Nutrition and Family Size in the Philippines." *Journal of Development Economics* 23(1):161–76.

Jensen, R. 2007. "Equal Treatment, Unequal Outcomes? Generating Sex Inequality through Fertility Behavior." Harvard University, John F. Kennedy School of Government, Cambridge, Mass. http://www.watsoninstitute.org/pub_detail.cfm?id=811.

Keyfitz, N. 1968. *Introduction to the Mathematics of Population.* Reading, Mass.: Addison-Wesley.

Larsen, U., W. Chung, and M. Das Gupta. 1998 "Fertility and Son Preference in Korea" *Population Studies* 52(3):317–25.

Leung, S. 1998. "On Tests for Sex Preferences." *Journal of Population Economics* 1(2):95–114.

Muhiri, P. K., and S. H. Preston. 1991. "Effects of Family Composition on Mortality Differentials by Sex among Children in Matlab, Bangladesh." *Population and Development Review* 17(3):415–34.

Pande, R. 2003. "Selective Gender Differences in Childhood Nutrition and Immunization in Rural India: The Role of Siblings." *Demography* 40(3):395–418.

Park, C. B. 1983. "Preference for Sons, Family Size, and Sex Ratio: An Empirical Study in Korea." *Demography* 20(3):333–52.

Pollit, E., K. S. Gorman, P. Engell, R. Martorell, and J. A. Rivera. 1993. "Early Supplementary Feeding and Cognition: Effects over Two Decades." *Monographs of the Society for Research in Child Development*, Serial No. 235, 58(7):1–99.

Pong, S. 1994. "Sex Preference and Fertility in Peninsular Malaysia." *Studies in Family Planning* 25(3):137–48.

Repetto, R. 1972. "Son Preference and Fertility Behavior in Developing Countries." *Studies in Family Planning* 3(4):70–76.

Rose, E. 1999. "Consumption Smoothing and Excess Female Mortality in Rural India." *Review of Economics and Statistics* 81(1): 41–49.

Rivera, J. A., R. Martorell, M. T. Ruel, J. P. Habicht, and J. D. Haas. 1995. "Nutritional Supplementation during Pre-school Years Influences Body Size and Composition of Guatemalan Adolescents." *Journal of Nutrition* 25(4S):1068S–77S.

World Bank. 2001. *Engendering Development through Gender Equality in Rights, Resources, and Voice*. Washington, D.C.: World Bank and New York: Oxford University Press.

Yount, K. M. 2001. "Excess Mortality of Girls in the Middle East in the 1970s and 1980s: Patterns, Correlates, and Gaps in Research." *Population Studies* 55(3):291–308.

Yount, K. M., R. Langsten, and K. Hill. 2000. "The Effect of Gender Preference on Contraceptive Use and Fertility in Rural Egypt." *Studies in Family Planning* 31(4):290–300.

Zeng Yi, T. P., G. Baochang, X. Yi, L. Bohua, and L. Yongpiing. 1993. "Causes and Implications of the Recent Increase in the Reported Sex Ratio at Birth in China." *Population and Development Review* 19(2):283–302.

The Consequences of the "Missing Girls" of China

Avraham Y. Ebenstein and Ethan Jennings Sharygin

In the wake of the one-child policy of 1979, China experienced an unprecedented rise in the sex ratio at birth (ratio of male to female births). In cohorts born between 1980 and 2000, there were 22 million more men than women. Some 10.4 percent of these additional men will fail to marry, based on simulations presented here that assess how different scenarios for the sex ratio at birth affect the probability of failure to marry in 21st century China. Three consequences of the high sex ratio and large numbers of unmarried men are discussed: the prevalence of prostitution and sexually transmitted infections, the economic and physical well-being of men who fail to marry, and China's ability to care for its elderly, with a particular focus on elderly males who fail to marry. Several policy options are suggested that could mitigate the negative consequences of the demographic squeeze. JEL codes: I18, J11, J12, J13, J26, N35

In an attempt to halt explosive population growth in China, the framers of the one-child policy of 1979 projected that if every woman of childbearing age had an average of 1.5 children, China would reach a peak population of approximately 1.2 billion in 2030, slowly declining thereafter to an ideal level of 700 million by the late 21st century (Yu 1980, projection 4). While these projections were remarkably accurate considering the available information, officials did not fully anticipate the impact of the fertility controls on the sex ratio at birth (the ratio of male to female births) and the social consequences of high sex ratios.[1]

Avraham Ebenstein is a Robert Wood Johnson Scholar in Health Policy at Harvard University; his email address is aebenste@rwj.harvard.edu. Ethan Sharygin né Jennings (corresponding author) is a Ph.D. student at the Population Studies Center at the University of Pennsylvania; his email address is garba@pop.upenn.edu. The authors thank Steven Leung for excellent research assistance, Sharona Shuster and Claudia Sitgraves for their careful editing, and Monica Das Gupta and Bill Lavely for their helpful comments and suggestions. An additional debt of gratitude is owed for the careful attention of four reviewers—the journal editor and three anonymous referees.

1. Song Jian, a leading scientist and politician credited with innovations in science and mathematics, was charged with developing policies to put China's population trajectory on the optimal path (Scharping 2003). This second-best scenario (after the ideal of one child per couple) was projected to result in a total population of 1.17 billion in 2025, declining to 777 million by 2080. While Song's projections did not incorporate the dramatic change in the sex ratio of births following introduction of the one-child policy, they did account for the already higher sex ratio of births in China.

THE WORLD BANK ECONOMIC REVIEW, VOL. 23, NO. 3, pp. 399–425 doi:10.1093/wber/lhp012
Advance Access Publication November 5, 2009

Government controls on marriage and childbirth instituted in the 1970s were intended to reduce population growth through delayed marriages, longer gaps between births, and lower lifetime fertility, a set of policies known as *wan xi shao* (later, longer and fewer). In 1979, a countrywide one child per couple policy was introduced. As the policy was codified and policy enforcement diffused throughout the country over the 1980s, parents unhappy with the prospect of never having a son became an increasingly common phenomenon. For many parents, intense son preference and the introduction of sex-selective abortion—made possible by the legalization of abortion after 1979 and the introduction of ultrasound technology in the early 1980s—led to a "merger of Eastern philosophy and Western technology." As a consequence, cohorts born between 1980 and 2000 included 22 million more men than women, a phenomenon known as the "missing girls" of China. According to projections in this article, approximately 10.4 percent of the men in these cohorts can be expected to fail to marry.

The popular press is replete with predictions that the vast number of unmarried men will destabilize Chinese society and represent a "geopolitical time bomb."[2] Hudson and den Boer (2004) argue that the high sex ratios in China will be associated with an increase in crime, since most violent crime is committed by unmarried young men. They also suggest that the poor marital prospects for these men may lead to China taking a more aggressive stance in world affairs, as happened before. In the 18th century, the Qing dynasty government responded to the rising sex ratios brought about by high levels of female infanticide by encouraging single men to colonize Taiwan. And in the 19th century, poor economic conditions in Shandong province led to rampant female infanticide and a subsequent rebellion when the unbalanced cohorts matured and organized an uprising against the Qing dynasty (Poston and Glover 2004).

The relevance of such examples to modern China is unclear, since empirical evidence is lacking on the connection between large numbers of single men and social upheaval. The potential consequences of this gender imbalance has spurred research in several disciplines, including demography, political science, and economics, but more work on the direct causal links between high sex ratios and social disorder is warranted.[3]

High sex ratios at birth have several predictable consequences, which this article analyzes. It finds that the growing population of unmarried men will affect the prevalence of commercial sex activity and the transmission of sexually transmitted infections, including HIV. And men who fail to marry may be worse off economically and will not have children to support them in their old age.

2. Michael Fragoso, "China's surplus of sons: a geopolitical time bomb," *Christian Science Monitor*, October 19, 2007. Retrieved from www.csmonitor.com/2007/1019/p09s02-coop.html)

3. Edlund and others (2007), exploiting time variation in the introduction of China's one-child policy to estimate the impact of high sex ratios on crime rates, find that the rising sex ratio explains a third of China's recent increase in crime rates.

Understanding the social and economic consequences of high sex ratios in China is critical in light of the persistence of this phenomenon since the advent of the one-child policy. The high sex ratios of cohorts born in the past two decades have already altered the demographic destiny of China. The shortage of women lowers the reproductive potential of the population and accelerates the shrinking of the population in the 21st century, absent a return to replacement fertility rates (Cai and Lavely 2005). Recent Chinese government figures indicate that the female deficit has actually worsened since the 2000 Census, with the official sex ratio at birth reaching 120 boys for every 100 girls in 2008 (China National Population and Family Planning Commission 2009).[4] Unless action is taken to reverse this trend, the negative consequences appear all but inevitable.

This article is organized as follows. The first section presents background information on marriage and fertility and uses population simulations to assess how different scenarios for changes in the sex ratio at birth and the total fertility rate could affect the share of men who fail to marry in China over the next century. Section II discusses the expected consequences of the high sex ratios and the failure of men to marry for migration, commercial sex activity, and the prevalence of sexually transmitted infections, with a focus on HIV. Section III explores the implications of the sex imbalance on China's ability to care for its elderly in an aging population with a growing number of unmarried, childless men. Section IV briefly discuss the benefits of marriage using indicators of economic and physical well-being and examines the welfare impact of the failure to marry on health and financial outcomes. Section V briefly discusses current efforts by the Chinese government to address the consequences of the skewed sex ratio and summarizes several policy recommendations for China in light of the anticipated costs of this worrisome demographic pattern.

I. Demographic Consequences of China's "Missing Girls"

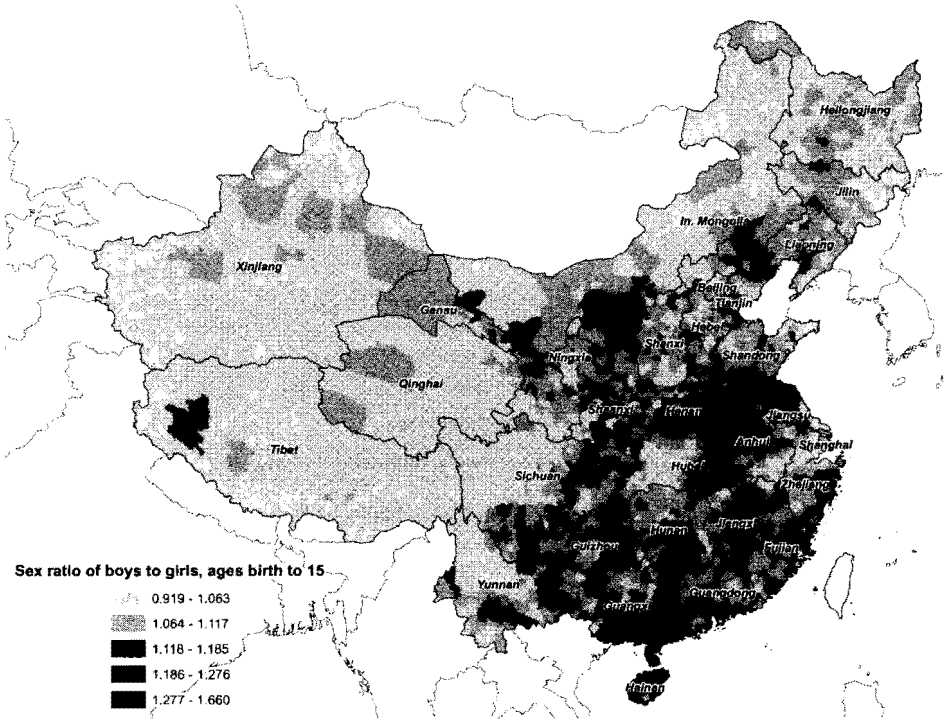
This section contains background information on marriage and fertility in China and presents several scenarios on how changes in the sex ratio at birth and the total fertility rate could affect the share of men who fail to marry in China over the next century

Marriage, Fertility, and Sex Ratios in China

The failure of men in China to marry because of a shortage of women is not an entirely new phenomenon. High sex ratios could be observed even in the 19th century, when missionaries reported that women they interviewed indicated very high rates of female infant mortality (Coale and Banister 1994). China's 1982 Census shows that nearly 6 percent of men born between 1935 and 1945 failed to marry, compared with less than 2 percent of the women

4. A discussion of alternative calculations of the sex ratio of new births is available in Goodkind (2008).

MAP 1. Sex Ratio of Children, Ages Birth to 15



Source: China National Bureau of Statistics (2000)

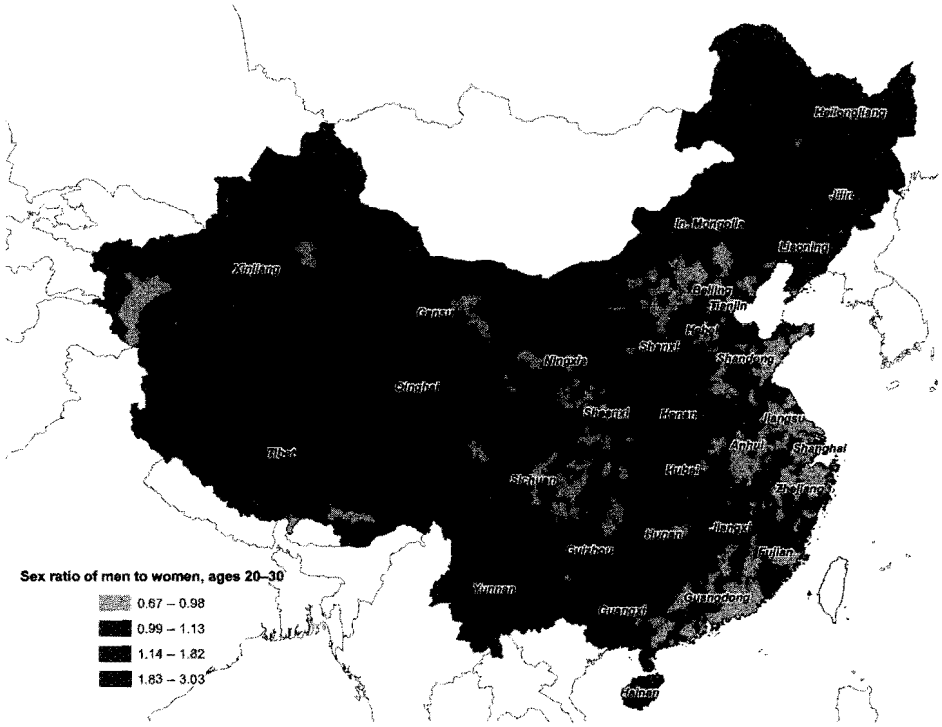
birth data are available.[6] Fertility has been falling in China for decades, for a number of reasons. Improvements in health have improved the survival rates of children to adulthood, greater economic competition has increased the level of investment necessary for each child, and government policy has encouraged family planning to various degrees.[7] This demographic transition, however, is made more profound by the policy climate in China, especially legislation regulating minimum age at marriage and the one-child policy. As birth cohorts age, they find that each successive generation is smaller than their own, giving rise to a kite-shaped age distribution in many Asian countries.

There is a discrepancy between the geographic areas with the highest sex ratios of children in China (map 1) and those with the largest shortage of women of marriageable age (map 2). The sex ratios of children—reflecting how strongly parents manifest a son preference—are highest in the Han majority areas of Eastern China. By contrast, the sex ratios at marriageable

6. In contrast, the 2008 revision of the UN *World Population Prospects* projection for China assumes that this level of sex ratio balance is not attained until 2050 (United Nations Population Division 2009).

7. Contraceptives, banned before 1953, became widely available after the government's first birth control campaign in 1957 (Hemminki and others 2005).

MAP 2. Sex Ratio of the Marriage Market, Ages 20–30



Note: The marriage market is defined as men ages 22–32 and women ages 20–30.
Source: China National Bureau of Statistics (2000)

ages are highest in the non-Han regions to the west, south and north. These are also the more remote and poor regions of China, where employment opportunities have grown far more slowly than in Eastern China. If men living in regions with better economic prospects are able to draw brides from poorer areas, it would appear to provide additional evidence for the suggestion made by many observers that Chinese society tends toward hypergamy (marriage with a person of a higher social class or position; Parish and Farrer 2000).

Projecting the Number of Unmarried Men in China over the Next Century

Projecting the number of unmarried men in China depends on sex ratios in future marriage markets, which in turn depend on the sex ratios at birth of future cohorts and population growth rates. This section describes the derivation and results of population simulations that capture the anticipated effect of high sex ratios on the number of unmarried men over the 21st century.

Decline in fertility could exacerbate the impact of the sex ratio imbalance, since future cohorts of men would be unable to find brides in younger and

smaller cohorts. But fertility rates in China are still a matter of scholarly debate.[8] The simulations presented here assume a total fertility rate of 1.45, based on China's National Bureau of Statistics (2005b) estimate from 2004 survey data, except where otherwise noted.[9]

The potential trajectories for the sex ratio at birth in China from 2006 to 2100 are summarized in four scenarios. The first scenario assumes an immediate correction in the sex ratio at birth to 1.06, which is overly optimistic but represents a lower bound for the analysis. The second scenario assumes that official policy such as the Care for Girls campaign is effective at stabilizing the sex ratio at birth at 1.09, a level identified as a government target, although there is no sign that this target will be achieved soon (Li 2007). The third scenario assumes that the sex ratio at birth in 2005 of 1.18 persists indefinitely, and the fourth scenario assumes a further deterioration of the ratio to 1.25.

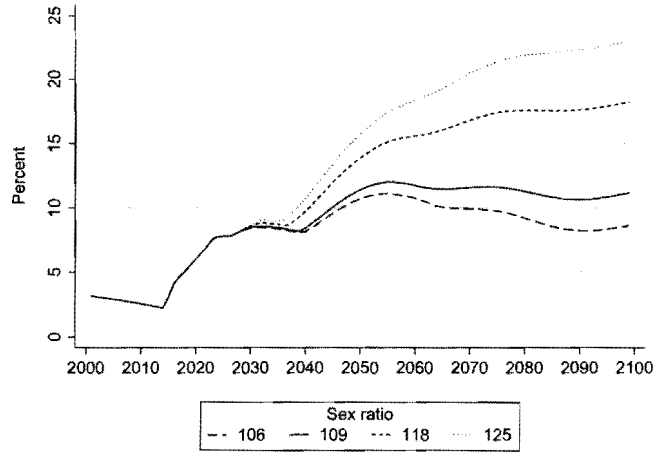
The simulation model allows for variations in fertility rates and the sex ratio of new births. The estimates here assume modest increases in fertility to 1.75 births per woman by 2010, although the choice of this date is not theoretically important. A return to replacement fertility without a concomitant adjustment in the sex ratio of new births will have only a minor effect in the long run on the percentage of the population failing to marry since it merely redistributes additional women to marginally older men (see Supplementary Appendix S1, at http://wber.oxfordjournals.org/, for additional fertility scenarios).

The simulations use age-specific mortality rates reported by Banister and Hill (2004) and essentially assume no improvement in life expectancy from 2000 onward. The marriage rule assumes that men marry all available women three years older or younger than they are until the supply of marriageable women is exhausted. Though a simplification of real marriage markets, the process nonetheless demonstrates the essential properties of a marriage market in which marriageable women become increasingly scarce because of both below-replacement fertility and an imbalanced sex ratio. The most realistic scenario that mitigates the serious consequences of the unmarried men

8. Data from the 2000 Census indicate a total fertility rate of 1.22 children in the prior year (China National Bureau of Statistics 2000). However, some argue that census officials were given misleading information out of a fear of punishment by parents who had violated the one-child policy (Retherford and others 2005). Such undercounting affects both fertility estimates and the observed sex ratio. However, Cai and Lavely (2005) found that 71 percent of the missing girls in the 1990 census were still missing in 2000. Also, the sex ratio of children ages birth to 4 in 2000 conforms well to the male to female ratio of children ages 5–9 in 2005 (1.19) from the China National Bureau of Statistics (2005a) One Percent Inter-Census Population Survey of China. While not decisive, these findings suggest that the undercounting issue is surmountable. Additional values for these parameters were included in the analysis here because of the remaining uncertainty about the extent of the undercount phenomenon. Cai (2008) summarizes the debate on China's total fertility rate and estimates a value of 1.5–1.6, in line with other third-party estimates.

9. These projections forecast a continuation of current trends, including modest increases in fertility at all ages. Many forecasts predict a rapid return to replacement fertility rates (Peng 2004). A supplemental appendix to this article, available at http://wber.oxfordjournals.org/, explores the sensitivity of the results to different fertility scenarios (table S1.2). The crucial assumption is not how population changes, but how the relative supply of men and women will change as fertility changes, which will be affected by the population size but will be less important than the sex ratio at birth.

FIGURE 3. Share of Men Ages 25 and Older Who Fail to Marry, under Four Scenarios, 2000–2100



Note: The technical assumptions underlying marriage formation for the simulations are outlined in detail in Supplementary Appendix S1 available at http://wber.oxfordjournals.org/. The shares of unmarried men are evaluated for four possible trajectories for the sex ratio at birth, ranging from an immediate correction to 1.06 to a further deterioration to 1.25.

Source: Authors' analysis based on data from China National Bureau of Statistics (2000).

phenomenon is one that addresses both the sex ratio and fertility. To the extent that marriage norms may change, the simulations overestimate the percentage of men who fail to marry. However, assortative mating constraints are not imposed, so the failure to marry rate is underestimated. In the end, these competing influences should largely cancel each other out. The details of the matching algorithm and alternative specifications testing the sensitivity of these results are described in Supplementary Appendix S1.

The results of the simulation are presented in figure 3. Under baseline assumptions, the share of men ages 25 and older who fail to marry will exceed 5 percent by 2020. As the cohorts born in recent years enter the marriage market and some share inevitably fail to marry, the population of unmarried men will rise well beyond this level. In the most optimistic scenario, where the sex ratio returns to normal immediately in 2006, the share of men who fail to marry will stabilize at just below 10 percent in 2060. In the second scenario, unmarried men will represent roughly 10–12 percent of men ages 25 and older. In the third and fourth scenarios, where the sex ratio at birth persists at either 1.18 or 1.25, the share of men who fail to marry will peak above 15 or 20 percent.

To some extent, these outcomes can be mitigated by realistic increases in both the age at marriage and the age gap between spouses.[10] This idea of demographic translation was introduced to describe the shift of the age-specific fertility distribution observed in the postwar baby boom era, but it also applies to the case of sex imbalance in marriage markets (Foster and Khan 2000; Ryder 1964). This view

10. Edlund (1999) demonstrates that son preference can account for increases in spousal age gaps and also the pattern of hypergamy.

holds that an excess of men over women in the marriage market can be fully compensated for by modest increases in men's age at marriage. Using an estimate of 15 percent excess men over women, it appears that the share of men who fail to ever marry can be kept close to the historical rate of 5 percent if the gap in age between spouses reaches eight years by 2050 (see also Supplementary Appendix S1).

This back of the envelope calculation neglects to consider that, because fertility rates are artificially held below natural replacement rates, each cohort of women entering the marriage market is smaller than the last. Indeed, the simulation results are highly sensitive to the assumption about the trajectory of fertility rates. With a return to a replacement fertility rate in the next decade, the impending problem of shortages of marriageable women can be averted, albeit by dramatic increases in both the age at marriage and the age gap between spouses.

However, there are few indications that the total fertility rate will rise to the natural replacement rate in the near future. The National Population and Family Planning Commission recently reaffirmed its intention to maintain the policy *status quo* for "at least another decade."[11] Moreover, the high sex ratios and smaller size of birth cohorts under the one-child policy imply that the age gap at marriage must increase until larger birth cohorts enter the marriage markets (some 25 years into the future, at the earliest), at which point any social upheaval associated with shortages of women and delay in marriage will already have occurred. In the more pessimistic scenarios, where the fertility rate remains around 1.45 and the sex ratio at birth remains above the natural rate, the age gap between spouses and age at marriage for men will necessarily rise *ad infinitum* as each cohort of men passes along the bride shortage to the next.

II. "Bare Branches," HIV, Prostitution, and Migration

In light of the large number of men who will delay marriage and who are anticipated to fail to marry, this section examines some of the potential negative impacts of high sex ratios.

In China during the early 1990s, growth in the number of people with HIV was concentrated among intravenous drug users and recipients of tainted blood transfusions. During the mid-1990s, however, HIV and AIDS began to spread to new regions and populations not previously considered at risk. As the population of single men rises, the transmission of HIV through risky heterosexual contact, particularly commercial sex activity, will become an increasingly severe problem.

Currently, the number of people who are HIV positive who contracted the disease through sexual contact is as large as the number who were infected through intravenous drug use. Individuals who contracted the virus from sexual activity represented half of all new infections in 2005 (China CDC and others 2006). The population that is HIV positive can be broken down into four groups. Intravenous drug users (90 percent of them concentrated in far western

11. Jim Yardley, "China sticking with one-child policy," *New York Times*, March 11, 2008. Retrieved from www.nytimes.com/2008/03/11/world/asia/11china.html?_r=1.

and southern provinces) account for 44.3 percent of infected people, and those infected through sex account for 43.6 percent (China CDC and others 2006).[12] The third group, those who donated or received blood from commercial blood donors, account for 10.7 percent, and the remaining 1.4 percent of infected people are those who were infected through mother-to-child transmission.

Considering the impending demographic pressures as heavily male birth cohorts enter adulthood and encounter shortages of marriageable women, female sex workers are an important at-risk group that has been understudied as an HIV vector. In the 1980s, sex workers represented a small share of the population, but between 1990 and 2000, prostitution expanded rapidly. Current estimates range from 1 million women whose primary income comes from commercial sex to up to as many as 10 million women engaging in paid sex of some kind.[13]

Recent evidence indicates that Chinese men are more likely than U.S. men to have paid for sex and that young Chinese men are more likely than older men to have visited a prostitute: 12.6 percent of men ages 21–30 and 8.8 percent of men ages 31–40 have been to a prostitute.[14] Moreover, Chinese men are less likely than their U.S. counterparts to report that they use condoms regularly, which places them at higher risk of sexually transmitted infection. While HIV rates among prostitutes are difficult to measure, the HIV prevalence rate among sex workers in Guangdong, Guangxi, and Yunnan provinces was as high as 11 percent in 2000,[15] and it seems reasonable to assume that the risky sexual practices of illegal sex workers place them at higher risk of exposure.[16]

While not all single men will patronize sex workers, and married men will also pay for sex, documenting the relationship between demographic change and commercial sex activity is important, as the population of single men will grow in the years to come.[17] Identifying specific groups of men who are more prone to patronize sex workers is also important because of the need to target public health interventions to the groups most at risk.

To analyze the relationship between numbers of men in at-risk groups and commercial sex activity, data from the Chinese Health and Family Life Survey were used to calculate the percentage of men reporting having paid for sex, for

12. The provinces with the highest levels of intravenous drug use (90 percent of it heroin) are Yunnan, Xinjiang, Guangxi, Guangdong, Guizhou, Sichuan, and Hunan. The share infected through sex includes those who contracted HIV from sex with a sex worker (19.6 percent of the total number of people infected with HIV), from an infected partner (16.7 percent), and from sex with men (7.3 percent).

13. Maureen Fan, "Oldest profession flourishes in China," *Washington Post Foreign Service*, August 5, 2007. Retrieved from www.washingtonpost.com/wp-dyn/content/article/2007/08/04/AR2007080401309.html.

14. Authors' calculation from the Chinese Health and Family Life Survey data (Population Research Center 2000). For comparable estimates, see Parish and Pan (2006).

15. This calculation is based on sex workers in detention centers, since prostitution is illegal in China (Settle 2003).

16. See Merli and others (2006) for an epidemiological model of sexual transmission of HIV in China.

17. To date, research has not been conducted on the relationship between the size of the single-male population and the supply of sex workers. While most researchers assume that the population of sex workers will increase as demand for their services increases, it could also be the case that the marriage squeeze for men may improve the marriage prospects of female sex workers and thereby take them off the sex market. This is a promising area for future research.

six regions (Population Research Center 2000). Paying for sex was most common in the coastal southern region, encompassing the provinces of Fujian and Guangdong, followed by the coastal eastern region including Jiangsu, Shanghai, and Zhejiang Provinces and the far northeastern provinces bordering the Democratic People's Republic of Korea and the Russian Federation. The majority of counties where a high percentage of men report having paid for sex tend to be counties with high percentages of single men. (Data on commercial sex activity are unavailable for Inner Mongolia, Tibet, and Xinjiang provinces.)

Among single men, young migrant construction workers make up a distinct at-risk population who are particularly likely to pay for services from low-cost female sex workers and are less likely to be educated about sexually transmitted infections and condom use (Garfinkel and others 2005). A pronounced relationship is found between the density of construction activity and the prevalence of commercial sex activity. In the urban provinces of Guangdong, Fujian, Jiangsu, Shanghai, and Zhejiang, more than 7 percent of men report having ever paid for sex. These and other areas of dense concentration of the construction industry, such as northern Shandong Province and the counties surrounding Beijing, merit particular attention from public health policy.[18]

The potential for an increase in HIV infection rates fueled by migrant workers has attracted the attention of many researchers. Tucker and others (2005) present compelling evidence that rising rates of sexually transmitted infection in cities are due to the sexual practices of migrant workers, who are demographically similar to the men who are projected to fail to marry: poor, uneducated, and single. Chen and others (2007, p. 1658) analyze HIV rates among a sample of patients being treated at 14 Guangxi clinics for sexually transmitted infections and conclude that "China's imbalanced sex ratios have created a population of young, poor, unmarried men of low education who appear to have increased risk of HIV infections." A multivariate analysis of factors that affect HIV status yields an odds ratio of 1.7 for single people relative to those who are married and 1.4 for men relative to women.

To determine how migration might affect the transmission of HIV, especially migration to China's growing urban centers, it is helpful to examine current and expected migration patterns. Comparing the geographic distribution of sex ratios at birth with the distribution of sex ratios among the current adult population reveals the regions from which migration is likely to occur in the future (see maps 1 and 2). Particular attention should be paid to counties where the sex ratio is abnormally high and where HIV prevalence is also high, such as the southwestern provinces of Guangdong, Guangxi, and Yunnan (Lu and others 2006). As the cohorts of men younger than 15 enter adulthood and experience demand–supply imbalances in marriage markets, the likelihood of commercial sex encounters and other risk-taking behavior increases. This dynamic is likely to be strongest in areas where the risk of contracting HIV is highest. At the same time, as women migrate

18. The results for men in the construction industry are included in Supplementary Appendix S2.

TABLE 2. Share of Men Ages 25 and Older Paying for Sex, and Simulated HIV Prevalence in the Entire Population, by Sex Ratio at Birth, 2000–30, 2050, and 2070 (percent)

	Sex ratio at birth										
						1.06		1.09		1.18	
Category		2000	2010	2020	2030	2050	2070	2050	2070	2050	2070
Paid for sex	6.28	6.92	7.78	8.35	8.36	8.26	8.42	8.40	8.59	8.76	
HIV prevalence	0.031	0.046	0.065	0.076	0.093	0.095	0.094	0.097	0.097	0.103	

Note: The simulations profile behavior based on the age, sex, and marital status of the population. Rates of having paid for sex in these groups are imputed using calculations from the 1999/2000 Chinese Health and Family Life Survey (Population Research Center 2000). The HIV simulations assume an odds ratio of 1.4 of men to women and a 1.7 odds ratio of single to married individuals (Chen and others 2007). The total count of HIV positive population in 2000–10 by this method is between the low and medium estimates of the Joint United Nations Program on HIV/AIDS (UNAIDS, various years). Results before 2030 do not differ appreciably by sex ratio at birth because of known characteristics of the population in 2000.

Source: Authors' analysis using data from China National Bureau of Statistics (2000).

to wealthier coastal cities to maximize their marriage prospects, these young men will also face pressure to migrate to cities, and both groups could bring HIV from the countryside to cities. Results by Yang (2006) confirm fears that male migrants experience elevated rates of HIV infection.[19]

The connections between cohort-specific sex ratios, prostitution rates, and HIV transmission are complex, but it is clear that these factors are all responsible for the rising HIV rates in China. Given the correlation between percentages of unmarried men and commercial sex activity, how will the increase in sex ratios and the ensuing failure of many men to find marriage partners affect markets for sex? The results of a simple simulation show how the incidence of prostitution might evolve (table 2). The simulation projects the share of men who pay for sex, assuming that the gender, marital status, and age-specific rates of having paid for sex found in 2000 persist during the 21st century. The Chinese Health and Family Life Survey finds that 14.7 percent of single men and 7.3 percent of married men admit to having paid for sex in 2000

19. A study by Parish and Pan (2006) found no significant difference in the risk of HIV contraction between urban men and low-status male migrants. If confirmed, this could mean a reduced likelihood that male migrants will carry HIV to cities, although female migrants may still play the same role. Many migrants may eventually marry, which could decrease the spread of HIV (by reduced prevalence of commercial sex or by containing the geographic spread of HIV if migrants return home to marry). Many men will lack the means to migrate to urban regions or will leave the city after a time with new wealth and marry at home. An anonymous reviewer noted that poor rural men have been less likely to migrate and that those that do migrate are still more likely to partner with women in their home region. Going forward, it can be expected that rural men who migrate to cities will be forced to compete with urban men for sex and mates and therefore will be more likely to visit prostitutes, presenting a problem even if these men eventually return home with wealth and marriage prospects. The conflicting results leave room for further study.

(Population Research Center 2000).[20] That information, plus the age profile of commercial sex activity, can be used to calculate a hazard rate of the chance of visiting a prostitute over the life cycle.

Although this calculation is admittedly imprecise, in that current rates of having paid for sex represent a lower bound on the future prevalence of prostitution (due to increased levels of future migration from rural to urban areas), the results show an increased demand for commercial sex among Chinese men. Assuming continuation of current behavior patterns, increases in the sex ratio at birth will create a modest increase in the share of men paying for sex. Changes in policy, income, or sexual culture will likely be more important in the future. Nevertheless, the simulations indicate that, almost immediately, demographic change alone will contribute to 2–3 percentage point increase in the share of men paying for sex in the next 30 years.

The simulations of how demographic change will affect China's HIV infection rate in the 21st century assume that the unknown hazard rate for HIV infection by age and sex generates 650,000 cases (the current estimated number of HIV cases in China) when applied to the population ages 22–40 in 2006. The share of the population that is HIV positive is then imputed to each cohort by sex, age, and marital status using the odds ratios from Chen and others (2007). Thus, these simulations attempt to model how HIV infection rates will change driven solely by changes in the demographic structure of China as cohorts with higher percentages of single men enter their sexually active years. The results indicate that the infected population will increase precipitously over the next 30 years and stabilize at a higher rate of infection. As with the results for patronage of commercial sex, the effect of variation in the sex ratio at birth on HIV transmission is limited. Variation in the sex ratio at birth between 1.06 and 1.25 (not shown) results in HIV infection rates in 2050 of 0.93–1.05 per 1,000. The greatest increase in HIV incidence, from 0.3 infections per 1,000 in 2000 to 0.76 per 1,000 in 2030, is a result of momentum from the known characteristics of the population in 2000.

While these projections do not incorporate increases in the probability of contracting the disease that might result as more people become infected, they also do not assume any improvement in preventive behavior. Since the Chinese government is beginning to respond to the impending HIV crisis, there is reason to hope that these projections are overly pessimistic. The central government and local authorities show signs of recognizing the growing role of sex workers in HIV transmission, and several pilot projects promoting safer sex (practices such as condom use) are in place in Beijing, Fujian, Hubei, Jiangsu, and Yunnan. Government budget allocation for HIV/AIDS efforts grew from approximately $12.5 million in 2002 to about $100 million in 2005 and $185

20. These percentages are derived from a regression of an indicator for having paid for sex on several demographic control variables, including marital status. See also the discussion of similar results for these data in Parish and Pan (2006).

million in 2006.[21] The government is also treating more cases of HIV, with projects such as the China Comprehensive AIDS Response (CARES) campaign, a program initiated in 2003 to supply domestically manufactured antiretroviral AIDS medication free to anyone who contracted the disease through tainted blood transfusions. The effectiveness of such efforts will be critical in containing the virus as the sex ratio rises and the percentage of those who are married falls among the sexually active population.

III. Support of the Childless Elderly

This section examines the impact of China's changing demographic structure, with a growing population of unmarried and potentially childless men, on its ability to care for its elderly. China's age distribution in 2000 exhibits two pronounced spikes, both emerging as a legacy of its demographic transition (figure 4). In the 1960s, the total fertility rate exceeded 6, and this baby boom resulted in a large cohort of people ages 30–40 in 2000.[22] The second baby boom occurred when these cohorts began to have children, and so the number of children born in the 1990s was also large. However, in the wake of government-mandated fertility control, each successive cohort in China has been smaller than the previous one.

Although China's population is more than four times that of the United States, it has less than three times as many births.[23] In 2030, the children born in the second baby boom of the 1990s will still be in their most productive working years and presumably will provide support (fiscal or otherwise) for the elderly. However, by 2050, the population forecast for China is far worse than that for the United States (see figure 4).[24] The elderly dependency ratios will be alarmingly high in China, with large numbers of people entering old age without young workers to replace them. In contrast, even without further immigration, the United States can anticipate a more favorable age distribution by 2050, with a relatively young workforce and very few baby boomers left in the population of elderly.

While retirement funding for social security programs in urban areas is receiving research and analysis, the looming problems among the population of rural peasants—who make up roughly 70 percent of China's 1.3 billion

21. "Spending on HIV/AIDS prevention set to double," *China Daily*, December 28, 2005. Retrieved from www.chinadaily.com.cn/english/doc/2005-12/28/content_507212.htm.

22. Some researchers identify this bulge in the population as one explanation for China's recent rapid economic growth. This phenomenon, when a large cohort of workers, preceded and followed by smaller cohorts reaches its most productive period in the labor force, is known as the "demographic dividend."

23. In China, only 10.6 million children were born in 1999 (and survived to 2000) compared with 3.8 million in the United States.

24. As projected in Alternative Scenario I of the 2007 Trustees Report by the U.S. Social Security Administration (U.S. SSA 2007).

population will exceed 35 percent of the overall population.[26] This aging of the population occurs against the backdrop of an emerging generation of unmarried, childless men.[27] China's traditional cultural assumption is that the elderly are cared for by their children, and living patterns and fertility decisions are predicated on the presumption of familial support. The state has made some effort to promote retirement homes (*yang lao yuan*), especially in rural areas, but these efforts have met limited social acceptance or private investment interest.[28]

China's population aging over the next 50 years has already been determined by the current age structure. It will coincide with the emergence of a new group of permanently unmarried men that will impose a large and increasing cost on Chinese society, especially in 2050 and beyond. This problem, common to all countries with a below-replacement fertility rate, is especially acute where selective abortions have altered the sex ratio. A preference for sons in China is at least partly economic, since sons have traditionally been the most important source of old age support. Increased acceptance of daughters could reduce welfare in old age if the additional girls are a couple's only child and if virilocality remains a social norm. In China, however, unlike in Italy or Japan, for example, the possibility of fertility returning to the replacement level seems much brighter because in China fertility may be significantly more responsive to public policy changes.[29] Actions taken today to allow Chinese to have larger families could improve the support ratio and might also allow more couples to have a son without resorting to sex selection, thus helping to reduce the number of unmarried men in these cohorts.

IV. Marital Status and Welfare

This section examines the relationship between welfare and marital status, documenting the greater poverty, poorer health, and shorter life expectancy among men who fail to marry, and possible developments in household bargaining between spouses.

The Census and the China Household Income Survey indicate that failure to marry is associated with lower income, less financial wealth, and poorer health (table 3 and Supplementary Appendix S2, table S2.1). The selection of healthier, higher earning men into marriage is partly responsible,[30] although there

26. The 2008 revision of *World Population Prospects* projects that 23.3 percent of the population will be 64 or older in 2050 (United Nations Population Division 2009). The comparable figures for the United States are 20.8 percent in 2050 and 21.6 percent in 2060 (U.S. SSA 2007, Scenario II).

27. Divorce or out of wedlock births are uncommon in China, so for most of these men, a failure to marry because of a shortage of women will imply a failure to have children.

28. "China vows to promote home care for elderly" *Xinhua News Agency*, February 22, 2008.

29. Supplementary Appendix S1 presents results of the model for several scenarios that assume a more rapid or slower pace of fertility growth, reaching replacement level at different dates.

30. Lillard and Panis (1996) present evidence that, in the United States, less healthy men marry earlier and remarry more quickly following divorce, suggesting that negative selection into marriage by health is also a potential confounding factor.

TABLE 3. Marital Status and 10-Year Mortality Rates of Men, by Age Groups

Age group	Ever-married men (percent)	Never-married men (percent)	Difference (percentage point)
55–59	14.3	15.2	−0.9
60–64	25.7	39.1	−13.4
65–69	41.3	51.3	−10.0
70–74	59.6	67.5	−7.9
75–9	77.1	86.1	−9.0

Source: Authors' analysis based on data from China Population and Information Research Center (1990) and China National Bureau of Statistics (2000).

is some evidence in other countries that men's wages rise after marriage, suggesting a causal link (Korenman and Neumark 1991).

Even after controlling for a respondent's age, education, ethnicity, and prefecture of residence, men in China who fail to marry have a third less income, live in households with an eighth less wealth, and are 11 percentage points less likely to describe themselves as being in good health than are men who marry. While the causal link between marriage and welfare outcomes has not been established in China in the period of interest,[31] marriage could theoretically improve health among married men through reductions in risky behavior and economies of scale in household welfare (Drèze 1997; Lanjouw and Ravallion 1995; Lillard and Panis 1996). A link between marriage and welfare is especially likely in China, because social insurance programs are limited and familial support is correspondingly critical to welfare.

The poor financial and health status of unmarried men observed in the survey particularly manifest in perhaps the most important measure of welfare—life expectancy. Implied mortality rates of men who married and those who did not between 1990 and 2000 were calculated by comparing the number of men in the 1990 and 2000 census data by marital status and calculating the survival of the artificial cohort (table 3).[32] Never-married and ever-married men who were ages 55–59 in 1990 had similar mortality patterns, but at older ages the never-married men had higher mortality rates. For example, among men ages 65–69, the mortality rate was 10 percentage points higher for never-married men, and less than half of the never-married men survived to the 2000 Census.

The welfare cost of poor health and high mortality for this population of unmarried men suggests that the high sex ratio at birth could indirectly reduce

31. For other countries, Hu and Goldman (1990) find significant mortality differentials by marital status (China is not included in their analysis).

32. This calculation assumes that men do not marry for the first time past the age of 55. First marriage beyond 50 is not observed among any of the respondents in the 0.1 percent sample of the 2000 Census (China National Bureau of Statistics 2000).

the quality and shorten the duration of the lives of never-married men. While the Chinese preference for sons results in high mortality rates for girls during pregnancy and infancy, if the relation between marriage and health proves to be causal, the outcome could be elevated mortality in later years for men unable to marry because of the shortage of women resulting from the earlier high mortality rate for unborn and infant girls.

It could also be the case that the shortage of female partners could lead to increased competition for brides, which could result in behaviors, including investment in education, that improve the health and well-being of men.[33] As the marriage market tightens, competition for scare women may increase the bargaining power of married women as well as single women. Evidence from outside China has shown that greater bargaining power of women, which can result from gender mismatch in the marriage market, can positively affect family health and welfare outcomes. These benefits, of course, would accrue to men who find marriage partners but not to those who remain single throughout their adult years.[34]

The evidence presented here suggests that China's demographic change in the 21st century will be dramatic and that difficulties in supporting China's large elderly population will be compounded by high sex ratios, which will deny childless men intergenerational support.

V. Policy Reponses to the shortage of females in China

This section briefly summarizes the Chinese government's policy response to the problems associated with the high sex ratio and discusses its consequences and possible alternatives.

When the one-child policy was introduced in 1979, China was only 20 years removed from the Great Leap Forward and the associated famine. Today, China is rapidly industrializing and experiencing the growth of a country that can easily feed its estimated 1.3 billion people. If current trends continue, the population is set to begin declining within the next 20 years. While overpopulation is no longer a pressing concern in China, the potential consequences of the legacy of missing girls is of immediate importance.

The alarming increases in sex ratios at birth revealed in the 2000 census spurred the Chinese government to action, and several programs were

33. An alternative strategy to reduce this uncertainty by identifying the causal direction for marriage and health and wealth involves finding a factor that affects marriage probability but otherwise has no influence on welfare. An instrument for marriage is difficult to find in China, since the factors affecting marital success are so closely related to factors that affect welfare. Panel data would also be useful in disentangling causality. The regression model presented in table S2.1 in Supplementary Appendix S2 includes controls that are important determinants of marital outcome and explain a good deal of variation in marital probability in reconstructed cohorts from cross-sectional data.

34. For details, see Lundberg and Pollack (1996) and Rao and Greene (1996).

implemented to address the female deficit. The government's response can be classified into two primary strategies: increasing the value of girls in the minds of parents and reducing the availability of sex-selection technology. The Care for Girls campaign identified 24 counties with extremely high sex ratios and provided incentives to reduce the female deficit, including free public education for girls. Preliminary indications are that these programs are having an effect. In a joint venture of the Ford Foundation and the United Nations Children's Fund (UNICEF), the Chaohu Experimental Zone Improving Girl-Child Survival Environment, established in 2000, succeeded in lowering the sex ratio at birth from 125 in 1999 to 114 in 2002 (Li 2007). The government is currently expanding the Care for Girls campaign to a national initiative. In 2004, President Hu Jintao declared that the campaign was a top priority and that the government would work strongly to stop any further rise in the country's sex ratio at birth over the next three to five years (Li 2007). Zhang Weiqing, director of China's population ministry, estimated that it would take 10–15 years to return China's sex ratio to natural level.[35]

In a second strategy, China is cracking down on sex-selective abortion. Several legislative initiatives aim to curb the practice and to punish offenders. The first statutory prohibition on sex-selective abortion came in 1989, and the most recent family planning law of 2002 bans the use of ultrasound or other technologies to determine fetal sex. If parents are caught aborting a child on the basis of sex, health professionals performing the operation are penalized and parents forfeit any right to have another child (Hemminki and others 2005). In 2006, the government shuttered several fertility clinics for violating the policy.[36] Despite these efforts, however, the sex ratio at birth was 1.18 in 2005, near the all-time high. Enforcement has been weak and uneven, possibly due to the overriding obligation of local governments to meet stricter population growth targets. The perceived need for a national policy campaign hints at an acknowledgment that sex-selective abortions have occurred, and the timing of higher parity births is further evidence that the practice has continued (Ebenstein forthcoming).

Efforts to improve funding for old-age security programs have been limited in scope and have focused on urban areas (Wang 2006). Very limited efforts have also been made to provide insurance in rural China, but they are insufficient for dealing with the looming old age crisis. In light of this concern, policy efforts should be made in two directions. First, China must acknowledge the implicit obligation to the large elderly rural population forecast for the next generation, since this generation's fertility has been too low to enable reliance on the traditional intrahousehold mechanisms of elderly support. Expanding

35. Interview transcript "Xinwenban jiu jiaqiang jisheng gongzuo he renkou fazhan zhanlv deng dawen," *Zhongguo zhengfu wang* of January 23, 2007. Retrieved from www.gov.cn/zhibo49/wzsl.htm.

36. Joseph Kahn, "China: crackdown on abortion of girls," *New York Times*, June 1, 2006. Retrieved from www.nytimes.com/2006/06/01/world/asia/01briefs-brief-003.ready.html?_r=5.

efforts to provide old age support and to collect the revenue to fund these initiatives is a top priority. Second, the Chinese government might want to consider revising its fertility policy. The simulations presented here suggest that the situation will deteriorate precipitously under the current policy, and higher fertility in the next decade would help smooth China's age distribution. Allowing extra births today will slow China's demographic decline and establish a larger supply of workers who could be taxed to fund the baby boom generation when they reach retirement.

The Chinese government's recent actions to provide contraception and care for those infected with HIV are promising developments, but actions to contain the spread of the disease must focus on the large and growing number of unmarried men who are at risk. China's legacy of missing girls will have a dramatic effect on Chinese society in the 21st century, with increased internal migration and rising demand for commercial sex all but unavoidable. Government action is unlikely to effectively reduce the prevalence of commercial sex, and so policy should aim to reduce the danger of this activity by raising awareness of the risk of contracting HIV and increasing the availability of condoms, especially in regions that attract unmarried men. Although China's HIV rates are still low, failure to act soon could prove costly, and HIV might be difficult to contain once it spreads to these unmarried men.

The future course of Chinese policy is yet to be determined. Central government planners, acknowledging the need to address the son preference, have chosen to do so through education campaigns, punishment for sex-selective abortions, and economic incentives for raising daughters. Although the one-child policy is subject to periodic review, its current fertility targets were recently reaffirmed despite the desirability of higher fertility for several reasons.[37]

The results presented here on some of the potential negative welfare consequences to having large numbers of men who fail to marry suggest at least two strategies: increasing fertility, thereby reducing the demand for sex-selective abortions and slowing population aging, and increasing legal and social incentives for raising daughters.[38] The discussion on revising the one-child policy has begun (Wang 2005). Many scholars have identified clear links between the one-child policy and the high sex ratio at birth over the last 20 years, and so an associated benefit of allowing higher fertility could be a mitigation of the costs presented here. The simulations presented here also suggest that an impending imbalance between working age and elderly cohorts in China could be offset somewhat by higher fertility rates. The simulations also indicate the need to act quickly. Even if action is taken immediately, China will still have to manage

37. Alexa Olesen, 2007, "China sticking to one-child policy," Associated Press, January 23, 2007. Retrieved from www.washingtonpost.com/wp-dyn/content/article/2007/01/23/AR2007012300398.html.

38. And reducing incentives for bearing sons, as might be expected to occur with increased institutional support for elderly and retired workers.

the highly skewed sex ratios in cohorts born over the last 20 years. Addressing this problem for the second half of the 21st century requires action today.

VI. Conclusion

The most significant unexpected consequence of China's one-child policy is the decline in the number of female children born to parents who are subject to strict fertility limits. In time, these missing girls will result in increasing tightness of the marriage market, with mixed consequences. This article attempts to establish the magnitude of the expected imbalance as boys born during the years of abnormally high sex ratios at birth and below-replacement fertility rates enter the marriage market and find a dearth of female partners. Three of the most important consequences of this phenomenon are the impact on prostitution, internal migration, and HIV transmission; the undermining of traditional old-age support mechanisms; and the impact on the health and well-being of men in the event of an increase in the failure to marry or, in demographic terms, in the lifetime celibacy rate.

As sons born during the years of skewed sex ratios reach adulthood and are unable to find marriage partners, the dangers associated with increased commercial sex may translate into higher HIV incidence. Simulations, using what is known about sexual preferences and practices, extrapolated increases in patronage of sex workers and the incidence of HIV. The imbalance in sex ratios of adults of marrying age will result in increased opportunities for women who migrate from rural areas to marriage markets in wealthier areas but will also put pressure on the men who are left behind to migrate to cities or to engage in risk-taking behaviors, such as drug use and commercial sex. The result could be the transmission of HIV from areas of high prevalence in southwestern and central China to urban centers that have been insulated so far. The share of men ages 25 and older who have paid for sex is projected to rise from 6.5 percent to 8–9 percent, and the HIV incidence rate is projected to rise from 0.3 per 1,000 to 0.8–1.1 per 1,000. Because of demographic changes already in motion, variation due to future fluctuation in the sex ratio at birth is likely to be minor compared with that due to government policies.

China has historically relied on family support systems for the elderly, with parents residing with their adult sons. Although the one-child policy might generate economic benefits in the short term, as a relatively larger group of young men are employed, in the longer run it means that a growing share of aging, never-married men will have no family to support them in their old age. The share of the population ages 65 and older is projected to peak between 2050 and 2060 at more than 35 percent. Without initiatives to fund the retirement of childless men, a large share of today's young men will face a tenuous existence as they age. There are additional concerns about the welfare of single men

before they retire, as research has found positive mental and physical health effects associated with marriage, regardless of marital fertility.[39]

Central government planners have tried to use incentives to encourage families to have daughters without increasing fertility. These measures will ease but cannot solve the problem of marriage market tightening for generations born since 1982. Without a suitable policy response, improvements in the sex ratio will not address the problems faced by retirees without family to provide support in advanced age. Of all of the demographic consequences of China's missing girls, the possibility of an AIDS epidemic has attracted the most attention among policy planners.[40] In the near term, major adjustments in marriage market matching behavior are likely, and absent a comprehensive policy response, a historically unprecedented population of men will likely suffer health and income setbacks as a result of their failure to marry.

This article finds considerable room for government policy to improve the likely effect of demographic trends on the spread of HIV. Also, vigorous efforts to reduce son preference are showing initial success (Das Gupta, Chung, and Li 2009) Without timely reform of elderly support systems to capitalize on the current surplus of working-age population, adjustments in total fertility and a substantial shift toward equalization of the sex ratio of new births will be crucial while these trends are still reversible, or the impending problems for China's *guang gun* will not be averted.

Supplementary Material

Two supplemental appendixes to this article are available at http://wber.oxfordjournals.org/.

References

Banister, Judith, and Kenneth Hill. 2004. "Mortality in China 1964–2000." *Population Studies* 58(1):55–75.

Bhrolchain, Maire Ni. 2001. "Flexibility in the Marriage Market." *Population: An English Selection* 13(2):9–47.

Cai, Yong. 2008. "An Assessment of China's Fertility Level Using the Variable-r Method." *Demography* 45(2):271–81.

Cai, Yong, and William Lavely. 2005. "China's Missing Girls: Numerical Estimates and Effects on Population Growth." *The China Review* 3(2):13–29.

39. Research has not demonstrated a causal relationship between marriage and health status, but wage growth for men has been found to respond positively to marriage in the U.S. context (Korenman and Neumark 1991).

40. Justin McCurry and Rebecca Alison, 2004, "40 m bachelors and no women ... the birth of a new problem for China," *The Guardian*, March 9, 2004. Retrieved from www.guardian.co.uk/china/story/0,7369,1165129,00.html.

Chen, X.-S., Y.-P. Yin, J.D. Tucker, G. Xing, F. Chang, T.-F. Wang, H.-C. Wang, P.-Y. Huang, and M. S. Cohen. 2007. "Detection of Acute and Established HIV Infections in Sexually Transmitted Disease Clinics in Guangxi, China: Implications for Screening and Prevention of HIV Infection." *Journal of Infectious Diseases* 196(11):1654–61.

China CDC (Center for Disease Control and Prevention), China Ministry of Health, UNAIDS (Joint United Nations Programme on HIV/AIDS), and WHO (World Health Organization). 2006. *2005 Update on the HIV/AIDS Epidemic and Response in China.* Beijing: National Center for AIDS/STD Prevention and Control.

China National Bureau of Statistics. 1982. "One per Thousand Sample of the 1982 China Population Census." Retrieved from the Integrated Public Use Microdata Series–International, Minnesota Population Center (https://international.ipums.org/international/).

———. 1990. "One Percent Sample of the 1990 China Population Census." Retrieved from the Texas A&M University China Archive (http://chinaarchive.tamu.edu/).

———. 2000. "One per Thousand Sample of the 2000 China Population Census."

———. 2002. "Rural Household Survey in China." Retrieved from National Statistical Society of China (www.nssc.stats.gov.cn/gjjws.asp?newsid=27).

———. 2005a. "One percent Inter-census Population Survey of China."

———. 2005b. "Sample Survey on Population Changes 2004." Retrieved from International Technology Associates (www.allcountries.org/china_statistics/).

China National Population and Family Planning Commission. 2009. "Main Population Data in 2008." Retrieved from China NPFPC (www.npfpc.gov.cn/en/detail.aspx?articleid=090428172413389282).

Coale, Ansley, and Judith Banister. 1994. "Five Decades of Missing Females in China." *Demography* 31(3):459–79.

Das Gupta, Monica, Woojin Chung, and Li Shuzhuo. 2009. "Is There an Incipient Turnaround in Asia's 'Missing Girls' Phenomenon?" World Bank Policy Research Working Paper 4846. World Bank, Washington, DC.

Drèze, Jean. 1997. "Widowhood and Poverty in Rural India: Some Inferences from Household Survey Data." *Journal of Development Economics* 54(2):217–34.

Ebenstein, Avraham. Forthcoming. "The Missing Girls of China and the Unintended Consequences of the One Child Policy." *Journal of Human Resources.*

Edlund, Lena. 1999. "Son Preference, Sex Ratios, and Marriage Patterns." *Journal of Political Economy* 107(6):1275–1304.

Edlund, L., H. Li, J. Yi, and J. Zhang. 2007. "More Men, More Crime: Evidence from China's One-Child Policy." Discussion Paper 3214. Institute for the Study of Labor, Bonn, Germany.

Foster, Andrew, and Nizam Khan. 2000. "Equilibrating the Marriage Market in a Rapidly Growing Population: Evidence from Rural Bangladesh." Working Paper. Philadelphia, PA: Economics Department, University of Pennsylvania.

Garfinkel, R., K. Longfield, Z. Zhang, J. Christian, and G. Zhang. 2005. "China(2005): HIV/AIDS TRaC Study Examining Condom Use among Construction Workers in Mengzi." First round. Washington, DC: Population Services International, Research Division.

Goodkind, Daniel. 2008. "Fertility, Child Underreporting, and Sex Ratios in China: A Closer Look at the Current Consensus." Paper presented at the Annual Meeting of the Population Association of America, April 17–19, New Orleans, La.

Hemminki, Elina, Zhuochun Wu, Guiying Cao, and Kirsi Viisainen. 2005. "Illegal Births and Legal Abortions—The Case of China." *Reproductive Health* 2(5). doi: 10.1186/1742-4755-2-5.

Hu, Yuangreng, and Noreen Goldman. 1990. "Mortality Differentials by Marital Status: An International Comparison." *Demography* 27(2):233–50.

Hudson, Valerie, and Andrea den Boer. 2004. *Bare Branches: The Security Implications of Asia's Surplus Male Population.* Cambridge, Mass.: MIT Press.

Korenman, Sanders, and David Neumark. 1991. "Does Marriage Really Make Men More Productive?" *Journal of Human Resources* 29(4):1027–63.

Lanjouw, Peter, and Martin Ravallion. 1995. "Poverty and Household Size." *Economic Journal* 105(433):1415–34.

Lee, Ronald. 1994. "Population Age Structure, Intergenerational Transfer, and Wealth." *Journal of Human Resources* 26(2):282–307.

Li, Shuzhuo. 2007. "Imbalanced Sex Ratio at Birth and Comprehensive Intervention in China." Paper prepared for 4th Asia Pacific Conference on Reproductive and Sexual Health and Rights, October 29–31, Hyderabad, India.

Lillard, Lee, and W.A. Panis. 1996. "Marital Status and Mortality: The Role of Health." *Demography* 33(3):313–27.

Lu, F., N. Wang, Z. Wu, X. Sun, J. Rehnstrom, K. Poundstone, W. Yu, and E. Pisani. 2006. "Estimating the Number of People at Risk for and Living with HIV in China in 2005." *Sexually Transmitted Infections* 82(Suppl III):iii87–91.

Lundberg, Shelley, and Robert A. Pollack. 1996. "Bargaining and Distribution in a Marriage." *Journal of Economic Perspectives* 10(4):139–58.

Merli, M. Giovanna, Sarah Hertog, Bo Wang, and Jing Li. 2006. "Modeling the Spread of HIV/AIDS in China." *Population Studies* 60(1):1–22.

Population Research Center. 2000. "1999/2000 Chinese Health and Family Life Survey." Retrieved from the Data Archive at the Social Science Research Computing Center at the University of Chicago (www.src.uchicago.edu/prc/chfls.php).

Parish, William, and James Farrer. 2000. "Gender and the Family," In Wenfeng Tang, and William Parish eds., *The Changing Social Contract: Chinese Urban Life During Reform*. New York: Cambridge University Press.

Parish, William, and Suiming Pan. 2006. "Sexual Partners in China: Risk Patterns for Infection by HIV and Possible Interventions." In Joan Kaufman, Arthur Kleinman, and Anthony Saich eds., *AIDS and Social Policy*. Cambridge, Mass.: Harvard University Asia Center.

Peng, Xizhe. 2004. "Is It Time to Change China's Population Policy?" *China: An International Journal* 2(1):135–49.

Poston, Dudley, and Karen Glover. 2005. "Too Many Males: Marriage Market Implications of Gender Imbalances in China." Paper presented at the 25th IUSSP World Population Conference, July 18–23, Tours, France.

Rao, V., and M. Greene. 1996. "Bargaining and Fertility in Brazil: A Qualitative and Econometric Analysis." Research Memorandum RM-153, Williams College Center for Development Economics, Williamstown, Mass.

Ryder, Norman. 1964. "The Process of Demographic Translation." *Demography* 1(1):74–82.

Retherford, Robert, Minja Kim Choe, Jiajian Chen, Li Xiru, and Cui Hongyan. 2005. "How Far Has Fertility in China Really Declined?" *Population and Development Review* 31(1):57–84.

Scharping, Thomas. 2003. "Birth Control in China 1949–2000." New York: Routledge Curzon.

Settle, Edmund. 2003. "AIDS in China: An Annotated Chronology 1985–2003". China AIDS Survey (http://hivaidsclearinghouse.unesco.org/search/resources/AIDSchron_111603.pdf).

Tucker, J.D., G.E. Henderson, T.-F. Wang, Y.Y. Huang, W. Parish, S.M. Pan, X.S. Chen, and M. S. Cohen. 2005. "Surplus Men, Sex Work, and the Spread of HIV in China". *AIDS* 19(6):539–47.

UNAIDS (United Nations Joint Programme on HIV/AIDS). Various years. *Report on the Global HIV/AIDS Epidemic*. (www.unaids.org/en/KnowledgeCentre/HIVData/EpiUpdate/EpiUpdArchive/Default.asp).

United Nations Population Division. 2009. *World Population Prospects: The 2008 Revision*. New York: United Nations, Department of Economic and Social Affairs.

U.S. SSA (United States Social Security Administration). 2007. *The 2007 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds*. Washington, D.C.: U.S. Government Printing Office.

Wang, Dewen. 2006. "China's Urban and Rural Old Age Security System: Challenges and Options." *China & World Economy* 14(1):102–16.

Wang, Feng. 2005. "Can China Afford to Continue its One-Child Policy?" *Asia Pacific Issues*, Analysis from the East-West Center 77. Honolulu, Hawaii: East West Center (www.eastwestcenter.org/fileadmin/stored/pdfs/api077.pdf).

Yang, Xiushi. 2006. "Temporary Migration and the Spread of STDs/HIV in China: Is There a Link?" *International Migration Review* 38(1):212–35.

Yu, Zhenpeng. 1980. "On China's Future Population Growth: Projections and Targets." [Translation]. *Population and Development Review* 6(2):343–48.

The Gender and Intergenerational Consequences of the Demographic Dividend: An Assessment of the Micro- and Macrolinkages between the Demographic Transition and Economic Development

T. Paul Schultz

The demographic transition changes the age composition of a population, potentially affecting resource allocation at the household level and exerting general equilibrium effects at the aggregate level. If age profiles of income, consumption, and savings were stable and estimable for the entire population, they might imply how the demographic transition would affect national savings rates, but there is little agreement on the impact of age composition. These age profiles differ by gender and are affected by human capital investments, whereas existing microsimulations are estimated from samples of wage earners that are not distinguished by sex or schooling and make no effort to model family labor supply behavior or physical and human capital accumulation. Considering these shortcomings of assessments of the "demographic dividend," a case study based on household surveys and long-run social experiments may be more informative. Matlab, Bangladesh, extended a family planning and maternal and child health program to half the villages in the district in 1977, and recorded fertility in the program villages was 15–16 percent lower than in the control villages for two decades. Households in the program villages realized health and productivity gains that were concentrated among women, survival and schooling increased among children, and after 19 years household physical assets were 25 percent greater per adult than in the control villages. These large gains in the wake of the program-induced demographic transition suggest reasons for designing new labor market and microcredit policies to help women during the demographic transition invest in productive skills; shift their time more efficiently from child care to home production, self-employment, and wage labor; and invest more in the human capital of their children. JEL codes: J13, J21, J68, O15

Several decades into a country's demographic transition, once its crude birth rate starts to decline steadily, the ratio of children (ages 0–14) to adults (ages

T. Paul Schultz is Malcolm K Brachman Professor, Emeritus, at Yale university; his email address is paul.schultz@yale.edu. The author appreciates comments from Mayra Buvinic, Monica Das Gupta, Andrew Foster, Maureen Lewis, Germano Mwabu, Shahid Yusuf, the journal editor, and three anonymous referees.

THE WORLD BANK ECONOMIC REVIEW, VOL. 23, NO. 3, pp. 427–442 doi:10.1093/wber/lhp015
Advance Access Publication October 31, 2009

15–59) declines, and for several more decades this decline in the youth ratio more than offsets the slow increase in the ratio of elderly (ages 60 and older) to adults. This intermediate stage in the demographic transition is associated with a temporary increase in the share of adults in the population that is referred to as the "demographic dividend." How does this change in the age composition of a population affect economic growth and the distribution of income by age and gender?

This article considers links proposed between the demographic transition and economic development that are sometimes assumed to operate through changes in the age composition of national populations. The demographic dividend literature emphasizes a period of high aggregate savings following the decline in fertility, but of potentially equal importance are the consequences for women's productivity and labor supply and the health of women who avoid unwanted childbearing. These life-cycle substitutions of family resources from childbearing activities to labor market activities may be facilitated by microcredit and labor market policies that ease the reallocation of women's time and bring family planning and reproductive health programs within reach of relatively immobile women in the rural South Asia. Such policies can reduce the gaps between the health and schooling of men and women and boost investment, economic growth, and labor force participation.

The article is organized as follows. Section I discusses the difficulty of reconciling the large aggregate estimates of life-cycle savings effects and the small and insignificant microestimates of age composition effects on household savings. Sections II and III review micro- and macrosimulation studies that concentrate on the expected consequences of changes in the age composition of the population and suggest their limitations due to omitted variable bias and misspecified production relationships at the individual and aggregate levels. Section IV considers the empirical evidence from a long-run social experiment in Matlab Thana, Bangladesh, suggesting how a village-level family planning program helped to reduce fertility and contributes to the reallocation of family resources. The program has spurred the labor productivity of married women, increased child survival, improved the nutritional health of women and daughters, increased the schooling of children, and added to the accumulation of physical capital, all consequences of the demographic transition that should accelerate development. Although it is only a single study (long-term social experiments with family planning and family health are rare), Matlab suggests that changes in the age composition of the population and the slowing of population growth are not the key mechanisms that translate the demographic transition into economic growth. Sections V and VI draw on the record from Matlab to suggest that the policy challenge is to find ways to assist women in using effective family planning and then to design labor and credit market policies for mothers who, with fewer children, will want to reallocate their time and family resources to improve their economic opportunities and to facilitate investments in the health, schooling, and migration of their children. Section VII discusses some directions for research.

I. Life-Cycle Savings Effects Associated with Age Composition

It is reasonable to imagine that changes in the age composition of a population, other things equal, should affect household demand for physical assets and human capital and thus influence life-cycle savings and asset prices. The portfolio of assets held by households might also change if assets complement the endowments of households that vary systematically over their life cycle, such as the labor of children.

Data for the United States and several other high-income countries have shown that the elderly do not dissave at the rate implied by the pure life-cycle savings model (Poterba 1994, 2004; Bernheim, Skinner, and Weinberg 2001). To maintain the core of the life-cycle savings hypothesis, economists introduce other motivations for savings, such as precautionary savings (wealth as insurance against unpredictable end-of-life expenditures and health crises) and a dynastic family consumption objective (the elderly are assumed to want to make bequests). Modigliani and Brumburg (1954) consider only adult consumption without reference to families or children. A third complication might arise if longer lifespans and longer retirement periods affected savings (Sheshinski 2006). Although these three extensions of the life-cycle savings model do not imply identical predictions, they are difficult to distinguish empirically from each other, as the life cycle becomes more multifaceted. Poterba (1994, 2004) reviews this literature and examines the empirical evidence, finding no close relationships between the age composition changes from 1950 to 2000 and financial market outcomes in the United States or convincing evidence from other countries or cross-country comparisons.

Most of the limited number of studies of low-income countries have problems establishing the magnitude of empirical relationships between age and income and savings at the household level or even across countries. Only in the 1990s do cross-country regressions begin to suggest that more rapid population growth and youthful age compositions are associated with lower physical savings rates and slower economic growth (Kelley and Schmidt 1996). This may be due to the inclusion for the first time of African countries, and many other factors could explain their slow growth, including political institutions, health crises, and civil conflict. Aggregate evidence across Asian countries reveals an association between savings rates and age composition, allowing for country fixed effects, but only if current savings is a function of lagged savings, and this lagged dependent variable is implausibly treated as though it were exogenous (Higgins and Williamson 1997). Whether the trends in declining mortality and fertility are causing the increased savings and economic growth within this sample of countries remains controversial (Deaton and Paxson 2000). When the lagged savings rate is treated as endogenous within countries and estimated using the authors' own list of instruments, the estimated age

composition effects on savings collapses and ceases to be significantly different from zero (Higgins and Williamson 1997).

There is no consensus on how to reconcile the larger aggregate estimates of the magnitude of life-cycle savings effects (Kelley and Schmidt 1996; Higgins and Williamson 1997) and the smaller and insignificant microestimates of age composition effects on savings based on household surveys (Poterba 1994, 2004; Deaton and Paxson 1997, 2000).

II. Microsimulation of the Aggregate Effects of Aging on Savings, Transfers, and Growth

Mason and others (2008) propose simplifying assumptions that permit them to impute production and consumption to individuals by age, based primarily on data from the United States and Taiwan, China. Given their age accounting system, which does not involve economic behavior in the form of human capital investment or labor supply, and ignores gender and schooling, an individual's age profile of savings leads to wealth accumulation and intergenerational private and public transfers by age. Household surveys are used initially to measure average earnings across age groups for all wage earners, and these synthetic age profiles of earnings are then adjusted proportionately to sum to national totals for wage income in the aggregate National Income and Product Accounts (NIPA). This imputation procedure assumes that all adults who are employed in the labor force (wage earners, self-employed, and unpaid family workers) work the same amount of time and contribute equally to national income regardless of gender or schooling, subject to the nonwage worker fitted income adjustment to the NIPA total (Lee 2003). Consumption is allocated by a variety of rules, many of which are country specific or imputed by arbitrary age–sex equivalence scales (Browning, 1992).

What are the conceptual problems with this methodology? Demographic outcomes that differ substantially by age in part due to biological factors, such as mortality or fertility, are forecasted as a function of changing age compositions. But when this approach is applied to income and consumption, transfers must occur between age groups that are then required to balance out surpluses in production minus consumption. These transfers may be financed in the private sector, within families or by charitable or religious welfare institutions, or in the public sector, notably through transfers to youth for schooling and to the elderly for health care and pensions. The problem in using this fixed age-matrix of economic outcomes for projecting income, consumption, savings, and transfers is that there is no behavioral or institutional mechanism hypothesized to equilibrate the consumption surpluses and deficits, balance the aggregate budget in each time period, or shift resources intertemporally, because there is no behavioral model for family formation, fertility, labor supply, human capital investments in children, consumption, savings, asset pricing, and wealth accumulation for retirement and bequests. The specific

problems with this demographic simulation approach that relies on age profiles without a behavioral model should be obvious.

III. Life-Cycle Savings and Their Effects on Countries

Models of behavior that are important for answering macroeconomic questions are sometimes hard to estimate with confidence from basic microeconomic data on individuals and households. One such case is the life-cycle saving hypothesis, in which consumption behavior in aggregate time series among countries is thought to be affected by age composition (Ando and Modigliani 1963; Modigliani 1970). Efforts to confirm the theory at the micro or household level have led to ambiguous empirical results.

Ideally, income and consumption would be observed for all individuals in a sample survey or census in order to compare savings rates by age and replicate the pattern with lifetime wealth data where savings and transfers can be measured to include capital gains and changes in stocks of consumer durables. Empirical problems arise because consumption is generally pooled and measured at the household level, and attribution of consumption by age restricts the analysis to single-person households, which constitute a small and unrepresentative subset of the population, especially among the young and old. This is a more serious problem in low-income countries, where a larger proportion of the population resides in intergenerational households headed by working parents or adult children and where self-employment is more common. Adult equivalence scales for consumption "requirements" of household members by age and sex are an unavoidable administrative tool for setting poverty lines and comparing welfare across households that are demographically and economically heterogeneous, but these scales should not be interpreted as derived from a conventional model of individual or household behavior (Browning 1992).

Those who work for pay in the market labor force are a selected sample. The number of hours they decide to work and contribute to household market income is also an endogenous decision that is determined by individual preferences that affect household composition. When savings rates are calculated for the elderly who remain the heads of their households, their savings is often positive and wealth continues to increase on average. An exception is the present discounted value of social security pensions or other annuities, which by definition decline with age unless augmented by other sources (Poterba 2004).

Intergenerational transfers are ignored in the pure life-cycle savings model and complicate the interpretation of age–wealth profiles. Modigliani suggests that intergenerational transfers are not important for understanding private wealth holdings. But Kotlikoff and Summers (1988) cite studies of transfers and bequests between living people in the United States and other high-income countries that conclude that transfers are a substantial factor in age profiles of

wealth (Bernheim, Skinner, and Weinberg 2001). Bequest motives within families offer the best available explanation for why so few elderly dissave or rely on annuity insurance to supplement life-cycle savings in the face of an uncertain and increasing life span (Kotlikoff and Spivak 1981). The expected magnitude of life-cycle savings is reduced when overlapping-generation models allow for intergenerational altruism and bequest motives to affect savings. The magnitude of savings that is then residually attributed to life-cycle consumption smoothing is modest, and this may be the most plausible explanation for the failure of microeconomic evidence to show much variation in household savings rates over the life cycle (Poterba 2004).

Household surveys from low-income countries are generally less well designed to document individual income by age and sex than are surveys in high-income countries. Given the fragile empirical basis and limited theoretical implications of more general life-cycle models, there is reason to view these frameworks as currently an unreliable forecasting tool.

Because of these shortcomings of cross-country regressions on changes in age composition, and the inadequacy of microsimulations built on rudimentary age profiles of wages or savings for assessing the "demographic dividend," more microeconometric analyses of household surveys and case studies are needed that define a counterfactual and explain how health and family planning programs affect the timing of the demographic transition and might thereby modify the behavior and development of families.

IV. A Social Experiment in Matlab, Bangladesh, and Its Effects

To estimate the causal effects of changes in the age composition of a population, it is necessary to specify factors that change mortality and fertility and thus affect the path of the demographic transition but do not otherwise affect the behavior or outcomes of interest. The approaches outlined in the previous sections do not identify an exogenous source of variation in fertility or mortality driving the demographic transition. They implicitly assume, therefore, that birth and death rates are determined outside the model and that any observed association between birth and death rates and economic change is therefore an indication of a causal relationship operating in a single direction from the demographic transition to development. Those working assumptions are not tenable.

Fertility and to some degree mortality respond to individual preferences and to household economic resources, as well as to other preconditions that affect economic development in many ways, such as institutions that raise the returns to investment and stimulate savings, increase women's education, and reduce fertility. Only when an exogenous shock occurs that reduces fertility can it be confidently inferred that subsequent changes can be attributable to the decline in fertility. To ensure this independence between population policies and

fertility change and economic development, the policy intervention should be designed as a social experiment. The goal is to show first that the population receiving the policy intervention has the expected lower fertility and slower population growth. Then, this program-associated voluntary reduction in births can be related to parent reallocation of time and resources from bearing children to other life-cycle activities that substitute for child labor and child support and care for their parents. Also, the impact on the quantity and quality of family labor supply might affect the regional labor market and influence the level of wages, as assumed by Malthus, and could influence the structure of wages between young workers and adults or between men and women.

It is widely observed that parents with fewer children devote more economic resources to each child, as often measured by their children's health and survival and years of completed schooling. The increase in the wage return to schooling in the 20th century has been attributed to the accumulation of complementary physical capital and to a skill bias in technical change that may motivate parents to increase their demands for child quality relative to child quantity. But there are as yet few empirical studies that account for the increase in schooling or health through exogenous increases in returns to human capital or through the decline in fertility.

Estimating the causal effects of exogenous fertility variation on family lifetime behavior and outcomes is a challenge for assessing the policy implications of the demographic dividend. At the individual level, the two instruments used most commonly to induce exogenous variation in fertility are twins and the sex composition of initial births (Schultz 2008). Twins are interpreted as an exogenous shock to fertility before there are drugs to treat subfecundity. But twins are not identical to singleton births, because twins have below average health endowments and birth spacing is altered for twins, an added burden on families, especially those that are credit constrained. The sex composition of initial births is even less useful as an instrument for estimating the consequences of exogenous variation in fertility, because in many low-income countries the estimated response arises because of the preference of some families for male offspring, which may be associated with other unobserved characteristics of those families, and the differential costs per child incurred by families rearing boys and girls.

A family planning and maternal and child health program implemented in a remote rural district of Bangladesh, Matlab Thana, was designed as a long-term social experiment. It was initiated in half of 141 villages that already had a reliable demographic surveillance system that registered all births, deaths, marriages, and population movements monthly. Under the family planning program outreach effort, begun in October 1977, female health workers contacted all married women of childbearing age every two weeks in their home, offering them various methods of birth control and, after 1982, a variety of additional maternal and child health services (Phillips and others 1982; Fauveau 1994). The program was maintained through 1996, when a household

survey was conducted that could be linked to background census data collected in 1974 and 1982 for the 141 villages (Rahman and others 1999).

No claim has been found that the villages were assigned randomly to the program and control areas. Program services were expected to influence behavior in neighboring villages, which they have, and these spillover effects could be reduced by clustering the program and control villages, as was done in Matlab. This regional cluster design also probably reduced the administrative and transportation costs of the program.

To assess whether the program and control areas differed before the program started, ratios of children ages 0–4 to women ages 15–49 from 1974 census data were compared in the program and control villages, and this indicator of surviving fertility did not differ significantly between the two types of villages. By the 1982 census, the surviving fertility levels were 16 percent lower in the program villages, according to a double-differenced population-weighted regression, and this difference remained 15 percent lower in the program than in the control villages after 19 years as shown in the 1996 follow-up survey (Moulton 1986; Joshi and Schultz 2007).

Population growth was more rapid in the control than in the program villages, but monthly wage rates did not differ significantly between the two village groups in 1996 for males or females ages 15–24 or for men ages 25–54. But for women ages 25–54, who in the program villages tended to have significantly fewer children by 1996, the monthly wage rates were 40 percent higher than in the control villages, though the participation of adult women in wage employment declined relatively in the program villages. Thus, the aggregate effects of population growth on wage rates that Malthus expected, because of diminishing returns to labor, are not evident in Matlab, whereas women who appear to have avoided unwanted and ill-timed births seem to have increased their productivity in the wage labor force (Schultz 2009). The Matlab family planning program can thus be viewed as a female-specific human capital investment program, raising adult women's wages about as much as would three years of additional schooling (Schultz 2009).

Other differences between the program and control villages confirm the tendency of the family planning and maternal and child health program in Matlab to be associated with increased schooling of children, measured by a Z-score normalized for age by sex. The nutritional health status of children, summarized by their body mass index Z-score, is significantly better for girls ages 1–11 in the program villages and for women ages 25–54 (Joshi and Schultz 2007).

Parents in the program areas reported 25 percent more lifetime assets by 1996 per adult residing in the household than did parents in the control areas. This pattern is consistent with parents treating physical assets as a substitute for children. The composition of household assets also differs between the program and control villages. Parents in the program villages reduced their value of livestock more rapidly than did parents in the control villages,

presumably because child labor is a critical input in caring for livestock. On the other hand, households in the program villages had 33 percent or more asset values than did control households in financial assets, ponds and orchards, homesteads, agricultural equipment, buildings and shops, jewelry, and consumer durables (Schultz 2009). The program-associated increases in women's wages, physical assets, and human capital are all expected to contribute to the economic development of villages in Matlab.

V. Household Credit, Human Capital Formation, and Financial Institutions

What can governments do to ensure that households are equipped to turn the decline in fertility into an economic dividend? In Matlab, the resulting household benefits do not appear to be due to the aggregate effects on the general wage labor market of the program's slowing of the population growth. The program benefits of improved control of reproduction are associated with households reallocating their time and financial resources as they reduce family size, realize health and productivity gains that are concentrated among women and children, and accumulate nonhuman capital to drive economic development.

If the returns to human capital rise, due perhaps to technical change in the world that complements the skills of more educated workers in production, parents with physical assets that are accepted as collateral, such as land, are in a more favorable position to respond to these new opportunities and borrow to invest more in the schooling of their children. Underinvestment in human capital by poor people may then occur, and public policies are needed to facilitate the schooling of children in poor households to prevent a widening gap in schooling between landed and landless classes. A variety of policy responses are discussed in the development literature: expanding local access to schooling, monitoring the quality of schools and making teachers more accountable to local parents in poor areas, providing fellowships for able students whose parents are relatively poorly educated and lack collateral to finance their children's schooling, and targeting cash transfers to poor mothers conditional on their children's enrollment and advancement in school.

Encouraging financial institutions to make loans to poorer parents may require subsidies and close monitoring to document that the loans reach the intended group and have the anticipated consequences on family resource allocation. An institutional alternative is joint lending to neighborhood groups or social networks, such as the prototypical Grameen Bank in Bangladesh, which is said to rely on social network pressures within a group of borrowers to substitute for the incentive effects provided by normal collateral to enforce loan repayment. Women often lack collateral because of their culturally weak property rights in the family, and women are consequently a prime beneficiary of some microcredit institutional innovations. These institutions could arguably

solve the problem of market failure for poor women who do not currently have access to the formal financial sector (Aghion and Morduch 2005).

But microcredit programs oriented toward poor women may still embody biases among types of investment activities and occupational careers that might discourage women from some favorable long-run choices. Physical capital investments may be favored over human capital investments. Self-employment of women may be favored over investments to enter the wage sector. Outputs from traditional home production activities may be less profitable in the long run than other types of off-farm production and employment. Self-employment activities might increase the marginal productivity of child labor and thus deter parents from investing more in the schooling and migration of their children. And most self-employed women work in productive activities at their home, which increases the likelihood that they can combine their work with their traditional responsibilities for child care, thereby lowering the opportunity cost of additional children and favoring larger family sizes, other things equal. Fertility may thus be increased by microcredit schemes targeted to poor women. If women reallocate their productive efforts outside their home and enter into wage employment, their lifetime productivity may increase as well as the human capital of their children.

Reorienting microcredit programs to facilitate women's transition to wage work should reduce a built-in bias of many programs that are oriented toward supporting self-employment for women in the home. Microcredit might be designed to help parents support the temporary or permanent migration of their daughters and sons to improve their adult employment opportunities, with remittances from the children to their parents helping to repay the loans, motivating parents and daughters to delay marriage and increasing daughters' influence in the choice of a mate as they become more economically empowered. Outmigration of children from poor rural areas might also encourage parents to first send their children to school for a longer period when urban jobs reward better educated workers more than do most manual rural jobs.

Finally, the products produced by participants in these microcredit programs for women might not represent the most promising lifetime opportunities. Traditional handicrafts (baskets, textiles, ceramics, and wood working) might not be commodities for which domestic demand is especially price and income elastic. Livestock, which are often acquired by women with the aid of microcredit programs, might increase the household's demand for child labor and thus discourage children's school attendance or outmigration. As noted earlier, these developments would reduce a woman's opportunity cost of having more children and could thereby sustain higher fertility. One evaluation of the consequences of microcredit in Bangladesh finds that after controlling for the heterogeneity of women who take loans from the village microcredit system, the program increased women's self-employment earnings, but the women were also more likely to then have additional births (Pitt and Khandkar 1998).

VI. Labor Market Reforms, Regulations, and Their Consequences for Women

Labor market regulations often restrict employment opportunities for low-wage groups, including women. How are these regulations modified so as to help rural women enter the wage labor force as their fertility declines? Evidence is accumulating, particularly from Latin America, that formal labor market regulations intended to raise wages, increase fringe benefits, fund social welfare programs, and increase job security for workers through employment regulations have one thing in common (Schultz 2000; Heckman and Pages 2004): they reduce employment opportunities for members of disadvantaged groups, who typically receive below average wages, presumably because they are less productive than the average formal sector worker. This includes inexperienced female entrants to the wage labor force, but also disadvantaged minority racial groups, such as indigenous groups in Latin America, lower castes in India, and remote ethnic and tribal groups in many regions. Mandatory regulations such as minimum wages and employee benefits may improve conditions for those who retain their jobs, if employers cannot shift their cost to workers, but labor market regulations tend also to exclude the less productive workers from entry-level jobs that might enable them to qualify over time for better jobs through on-the-job training (Mincer 1976).

Raising minimum wages reduces employment proportionately and lowers labor force participation rates among low-wage groups (Maloney and Nuñez 2004). In some provinces of Canada, for example, extending fringe benefits to women in the form of maternity leave reduced women's wage rates relative to men's and reduced the share of female employment in the provinces that added maternity leave (Gruber 1994). Where minimum wages are binding and coverage is enforced, it is anticipated that formal sector employment will be reduced among the less productive groups whose current output does not exceed what employers must pay for labor. Especially in South Asia and Africa, where schooling is substantially less for adult women than men, minimum wage regulations reduce women's employment opportunities in entry-level jobs (Bell 1997; Gruber 1997; Revenga 1997; Schultz 1988; Heckman and Pages 2004).

Labor market reform appears to be difficult to achieve directly, because of the political strength of vested interests, including unions, in maintaining the *status quo*. These restrictions in the labor market that restrain women's entry into the formal sector have, however, been indirectly eroded in some countries through lowering the barriers to international trade and encouraging foreign direct investment (Schultz 2000). Country studies have also found that women's employment is concentrated in export-oriented industries and that women's share of jobs in these industries increases as barriers to trade fall (see, for example, Ravenga 1997; Ozler 2000; Hanson 2003).

Nonetheless, even when employment growth is rapid, as in the Middle East and North Africa since 2000, and barriers to trade and capital mobility are

reduced, unemployment rates for women have risen relative to those for men, and this unutilized supply of women's labor is larger in many countries in the region for better educated women (Nabli, Fauregui, and de Silva 2007). The decline in the public sector employment may encourage more efficient labor allocation, but in some economies, such as Egypt before 1990, the public sector is a major employer of educated women. Reducing wages in this protected public sector could lower women's wages but thereby expand women's employment opportunities going forward and increase the benefit–cost ratio in the public sector provision of schooling and health services, delivered mainly by female employees.

VII. Conclusions and Research Directions

Following the demographic transition, the growth in the labor force and the increase in per capita productivity tend to be associated with the increase in women's labor force participation rates. With a microeconomic model of family labor supply that accounts for women's time allocation and their productivity in the wage sector, it should be possible to answer more generally the question that motivated this article. How do policies that affect the decline in fertility contribute to development through the increase in household income and to the accumulation of household human and physical capital? Changes in women's labor supply and household savings, in both human and physical capital, are major sources of per capita economic growth that may plausibly be linked to the demographic transition. With the growing availability of household panel survey data, microeconometric models of family labor supply, fertility, and consumption behavior could be estimated. The fertility transition could then be accounted for within a simultaneous equation behavioral model rather than by analyzing household consumption while treating fertility and family composition as though they were exogenous "control" variables.

Because the productive opportunities of people not in the wage employment are not observed, inferring the average productivity of all men and women requires a model that accounts for who participates in the wage sector, as well as the productive characteristics of individuals, such as their human capital and other resources (Heckman 1974a,b; Schultz 2009). Identifying such a sample selection model requires a variable that affects the individual's productivity of time in nonwage work or leisure but that is uncorrelated with the unobservable determinants of the market wage. The rural family's ownership of agricultural land is a possible exclusion restriction that is expected to raise labor productivity in home production and self-employment and increase the value of leisure, thereby reducing the likelihood that a wife, her husband, or their children will work outside the household for a wage (Schultz 2009). But land could also be correlated with unobserved factors such as ability, motivation, and family connections that might affect market wages.

This empirical approach underlies the findings on the Matlab district of Bangladesh, summarized above, that gains in productivity due to a program-induced decrease in fertility and slowing of population growth appear to have promoted development. No relative gain in wages of male or female workers ages 15–24 is detected in the villages where a family planning social experiment has reduced fertility and population growth for two decades, challenging a premise of the Malthusian framework. However, older women in the program villages, ages 25–54, who have reduced their childbearing, are observed to receive much higher wages than women in the control villages, holding constant for schooling and age. This empirical finding in Matlab confirms the hypothesis that an effective family planning and reproductive health program can enhance women's human capital and productivity. But the program effects on the time allocation of women benefiting from their avoidance of unwanted childbearing is difficult to predict *a priori*. In the Matlab case, women ages 25–54 work less in wage employment in the villages served by the program than in the comparison villages (Schultz 2009).

Microcredit institutions in many parts of the world have provided financial assistance for poor women seeking to enter the labor force as self-employed workers. Conditional cash transfers have also been widely adopted in Bangladesh and countries in Latin America as a public institutional mechanism to encourage poor mothers to invest in the schooling of their children, while also minimizing the leakages common in traditional transfer programs as a result of political corruption. It may be productive to reorient these microcredit institutions to also encourage poor families to invest in the human capital of their children as well as to provide loans to cover the costs of mothers entering the formal wage labor market. Changing the orientation of microcredit institutions from a focus on the self-employment of women in home-based cottage industries to one that also facilitates family human capital investments, migration, and wage work by women could extend and strengthen the benefits of microcredit for poor women and their children, especially following the demographic transition.

Reducing labor market regulations, such as minimum wages and mandatory benefits for workers in covered sectors, is one way to diminish the barriers to women's access to low-wage entry jobs that can enable them to improve their productivity through on-the-job experience and learning. Middle-age women are often denied employment because of a lack of experience. Unions understandably defend the employment benefits and prerogatives of the segment of the middle class they represent in low-income countries. In the public sector, unions can reduce the accountability and efficiency of the workforce assigned to produce essential public services in education and health care. In such cases, lowering wages for women entering the labor force or allowing more competitive entry of "uncertified" teaching assistants and auxiliary health workers in the public sector could reduce insider rents. But all stakeholders in the public sector might not support such reforms (Banerjee and others 2007).

Schultz, T.P. 1988. "Firm and Family Employment, Development, and Minimum Wages." *Estudios de Economia* 15(1):85–125.

———. 2000. "Labor Market Reforms: Issues, Evidence, and Prospects." In A.O. Krueger, ed., *Economic Policy Reform*. Chicago: University of Chicago Press.

———. 2006. "Does Liberalization of Trade Advance Gender Equality in Schooling and Health?" In E. Zedillo ed., *The Future of Globalization*. London: Taylor and Francis Books, Ltd.

———. 2008. "Population Policies, Fertility, Women's Human Capital, and Child Quality." In T.P. Schultz, and and J. Strauss eds., *Handbook of Development Economics*. Vol. 4. Amsterdam: Elsevier, B.V.

———. 2009. "How Does Family Planning Promote Development? Evidence from a Social Experiment in Matlab, Bangladesh—1977–1996." Yale University, Economic Growth Center, New Haven, Conn.

Sheshinski, E. 2006. *Longevity and Aggregate Savings*. CESifo Working Paper 1828. Munich, Germany: Munich Society for the Promotion of Economic Research.

Macroeconomic Stability and the Distribution of Growth Rates

Vatcharin Sirimaneetham and Jonathan R.W. Temple

It is often argued that macroeconomic instability can form a binding constraint on economic growth. Drawing on a new index of stability, threshold estimation is used to divide developing economies into two growth regimes, depending on a threshold level of stability. For the more stable group of countries, the output benefits of investment are greater, conditional convergence is faster, and measures of institutional quality have more explanatory power, suggesting that instability forms a binding constraint for the less stable group. Macroeconomic stability is also shown to dominate several other candidates for identifying distinct growth regimes. JEL codes: O23, O40

It is widely believed that economic growth requires macroeconomic stability. At the broadest level, stability could help to explain the sustained growth of East Asian countries between the early 1960s and the late 1990s. By contrast, Latin America and Sub-Saharan Africa have often endured both macroeconomic disarray and slow growth. Economic mismanagement could also help explain why some developing economies became heavily indebted, in which case the relatively slow growth of the 1980s and 1990s might be attributed to the macroeconomic policies of earlier decades.

Although macroeconomic stability could be important for growth, the strength of the empirical relationship remains uncertain. One argument is that the observed correlation between stability and growth is mainly due to a few countries with the very worst macroeconomic outcomes. Once a certain

Vatcharin Sirimaneetham is a consultant in the Poverty Reduction and Economic Management Network, East Asia and the Pacific Region, World Bank; his email address is vsirimaneetham@worldbank.org. Jonathan R.W. Temple (corresponding author) is a professor of economics at the University of Bristol; his email address is jon.temple@bristol.ac.uk. The authors are grateful to the journal editor, three anonymous referees, David Ashton, Holger Breinlich, Edmund Cannon, Huw Dixon, Jan Fidrmuc, Max Gillman, Andreas Leukert, and Patrick Minford for related comments or discussion as well as to seminar participants at the University of Bristol, Cardiff University, the 2006 Royal Economic Society Conference at Nottingham, and a Brunel University workshop at the Centre for Economic Development and Institutions on aspects of growth and macroeconomic policy. Temple thanks the Leverhulme Trust for financial support under the Philip Leverhulme Prize Fellowship scheme. A supplemental appendix to this article is available at http://wber.oxfordjournals.org/.

THE WORLD BANK ECONOMIC REVIEW, VOL. 23, NO. 3, pp. 443–479 doi:10.1093/wber/lhp008
Advance Access Publication September 16, 2009

threshold level of stability has been achieved, the marginal benefits of additional stability could be minimal. Another argument, which dates to at least Sala-i-Martin (1991), is that macroeconomic disarray could be a symptom of deeper problems. Recent research, especially after the work of Acemoglu, Johnson, and Robinson (2001), Acemoglu and others (2003), and Easterly and Levine (2003), argues that macroeconomic policies lack explanatory power relative to institutions. But this is far from a consensus, and Henry and Miller's (2009) case study of two Caribbean islands presents a different view.

This article revisits the growth effects of macroeconomic stability. As this is well-worked ground, a new article on this topic must work hard to justify its existence. One innovation is a composite index of macroeconomic stability. A more fundamental aim, however, is to sharpen the link between statistical modeling and informal commentary on policy and growth. Much of that commentary reduces to a simple idea: sound policy is a necessary but not sufficient condition for rapid growth, and bad policy may often be a sufficient condition for slow growth. Perhaps growth performance is only as strong as the weakest link in a set of policy outcomes.

Although the practical analysis of growth policy is often framed in terms of necessary and sufficient conditions, incorporating this idea into empirical models is not straightforward. Another approach—similar in spirit, but more general—frames the problem in terms of binding constraints, as in recent work by Hausmann, Rodrik, and Velasco (2008) and Rodrik (2007). If the marginal effects of policies and other growth determinants are not independent, one or more constraints on growth may be binding, with reforms elsewhere having limited benefits, at least until the key constraints are addressed. This contrasts with the linear regressions usually adopted in the empirical growth literature, which implicitly assume that different growth determinants smoothly substitute for one another.

With all this in mind, this article explores methods designed to close the gap between the vocabulary of policy analysis and the empirical models used to explain growth variation. First, direct comparisons of growth rate distributions are used, with countries divided into groups based on an index of macroeconomic stability. These distributions clearly show that macroeconomic instability is not always a binding constraint. In particular, even when a country ranks low in terms of macroeconomic stability, this is not a sufficient condition for slow growth. But the highest long-run growth rates are confined to countries with stable macroeconomic outcomes.

The article then examines how regressions can accommodate the binding constraints view. Standard regressions are used to quantify the effects of macroeconomic stability over 1970–99, restricting the sample to developing economies. These linear models assume that any adverse effect of instability can be offset by other factors and thus that instability is never a binding constraint. To allow for instability as a binding constraint, threshold estimation, based on Hansen (1996, 2000), is used. The results indicate that the sample can be split into two groups

by macroeconomic stability. For the more stable group of countries, the elasticity of steady-state output to the investment rate is greater, conditional convergence is faster, and the standard growth determinants of the Solow model (together with a measure of institutional quality) explain 75–90 percent of the cross-section variation in growth rates, a remarkably high proportion. For the less stable group, instability reduces growth, while the Solow variables have less explanatory power, investment is less effective, and the residual variance is much higher. Fundamentals such as good institutions are not strongly associated with growth unless macroeconomic stability is also in place. These results suggest that instability can indeed form a binding constraint on growth.

The analysis acknowledges an important criticism of past research: that policy outcomes are likely to be endogenous in both an economic and a statistical sense. Rodrik (2005) points out that observed policies are decision variables that must be endogenous to social and economic circumstances. The implication is that macroeconomic stability is not randomly assigned and will almost certainly be correlated with omitted country characteristics—and thus with the error term of the growth regression. When this problem arises in the microeconometric literature, the availability of control variables is often limited, but there may be plausible candidates for instrumental variables through "natural experiments." Growth researchers face almost the mirror image of that situation: there are many possible control variables but few plausible candidates for instruments.

This article uses two approaches. The first follows Barro (1996) in exploiting the observed association between French colonial heritage and macroeconomic stability, linked to the membership of many former French colonies in the CFA franc zone. This implies that French colonial heritage could be a suitable instrument, but it would not be difficult to criticize the necessary exclusion restriction. For example, French colonial heritage is likely to have influenced the legal system, with a variety of effects on development, a debate reviewed in La Porta, Lopez-de-Silanes, and Shleifer (2008).

The analysis therefore emphasizes an alternative approach that considers an unusually wide range of possible control variables, including various indicators of geographic characteristics and institutions. This comprehensive approach increases the chance of identifying controls that influence the extent of stability, in order to lessen the correlation between macroeconomic stability and the error term, even though macroeconomic stability is not randomly assigned. This relates to "selection-on-observables" from the treatment effects literature and is appropriate if the central endogeneity problem is omitted variables rather than simultaneity bias.[1] The approach is based on Bayesian methods for

1. Simultaneity bias is relevant if policy outcomes depend directly on growth outcomes, which may be plausible in the short run but less so over the 30 years considered here. It is more plausible that growth and policy outcomes are jointly influenced by other variables, such as institutions, hence this article's emphasis on the omitted variable problem rather than on simultaneity.

model averaging and thus addresses the model uncertainty problem highlighted by Levine and Renelt (1992). The evidence that stability matters varies with the sample of countries, but in the largest sample considered the estimated benefits of stability are robust across a wide range of specifications.

Finally, the results are used to construct counterfactual distributions of growth rates and steady-state levels of GDP per capita. These distributions indicate what might have happened had all developing economies achieved macroeconomic stability over 1970–99. To the extent that the estimated benefit of stability can be interpreted as a causal effect, the variation in stability exerts a major influence on the distributions of growth rates and steady-state GDP per capita. But it is important to acknowledge some major qualifications. As mentioned, macroeconomic instability may be a symptom of other problems. Instability may arise in the wake of conflict or relatively severe external shocks. The estimates are thus best interpreted as an upper bound on the importance of good macroeconomic management.

The article is organized as follows. Section I briefly reviews the literature on macroeconomic policy and growth and discusses the empirical analysis of binding constraints. Section II describes the new measure of stability. Section III looks at the relationship between stability and growth in a variety of ways, emphasizing threshold estimation. Section IV examines robustness using Bayesian methods. Section V uses the earlier growth regressions to generate counterfactual distributions of growth rates and steady-state levels of income. And section VI presents some implications of the findings.

I. The Literature on Macroeconomic Policy and Growth

Much of the literature on policy and growth has studied trade regimes and, more recently, such factors as entry barriers and regulation. But this article is about macroeconomic stability—not market-led development or the Washington Consensus. As initially summarized by Williamson (1990), the Washington Consensus reflected principles that went well beyond macroeconomic policies and included tax reform, financial and trade policy liberalization, openness to foreign direct investment, privatization, deregulation, and protection of property rights. Rather than investigate these, this article examines whether the Washington Consensus was right to emphasize the benefits of stable macroeconomic outcomes. Attempts to achieve stability can be controversial, especially when reductions in fiscal deficits are proposed. Moreover, it is rarely clear how much stability is "enough."[2]

2. This article does not address the subtler and much more difficult questions that relate to short-run policy activism such as demand management. The results concern macroeconomic outcomes (rather than policies) assessed over the long run and should be interpreted in that light; they do not imply, for example, that budget deficits must always be avoided.

Motivated by these considerations, empirical studies such as Bleaney (1996) and Fischer (1991, 1993) concluded that macroeconomic stability matters for sustained growth. More recent researchers are not so convinced. Macroeconomic policy outcomes have generally improved over time, while many developing countries grew more slowly during the 1980s and 1990s than they had previously. This led to the conclusion that the growth dividend of greater macroeconomic stability has been disappointing, an argument reviewed in Montiel and Servén (2006). The reasons behind the post-1980 growth collapse in developing economies are discussed in Easterly (2001b) and Rodrik (1999) and seem likely to go beyond macroeconomic policy decisions.

Other evidence casts further doubt on the role of stability. Improvements in policy indicators explain relatively few growth accelerations (Hausmann, Pritchett, and Rodrik 2005), and in general policy indicators are far more persistent than growth rates are, suggesting that policy will usually leave the medium-run variation in growth unexplained (Easterly and others 1993). Perhaps most fundamental, empirical studies such as Easterly and Levine (2003) have found that growth and policy variables are not robustly correlated in the cross-country data when controlling for institutional development. Easterly (2005, p. 1055) concludes that "the long-run effect of policies on development is difficult to discern once you also control for institutions." This highlights a problem in the empirical literature: that economic disarray usually extends across a range of outcomes. It can be hard to disentangle the effects of specific macroeconomic outcomes from one another and from other growth determinants. Perhaps bad macroeconomic outcomes are best seen as symptoms of deeper underlying problems, including institutional weaknesses and exposure to external shocks.

Although some claims for the importance of policy may have been exaggerated, a commonsense view commands wide support: there is likely a threshold level in the quality of macroeconomic management below which growth becomes difficult or impossible. Easterly (2001a) provides a clear and persuasive exposition of this view, indicating that governments may not be able to initiate growth, but they can destroy growth prospects with bad enough macroeconomic policies. He illustrates the consequences of policy errors using several historical examples, showing that the worst policy outcomes—hyperinflation, high black market premiums, and large budget deficits—are typically associated with slow growth or even collapses in output. None of this implies, however, that getting macroeconomic policy right is a sufficient condition for rapid growth. It is not difficult to find countries with sound macroeconomic policies and slow growth—Bolivia in the 1990s, for example, discussed in Kaufmann, Mastruzzi, and Zavaleta (2003).

The commonsense view dominates recent assessments of the role of policy but is largely absent from the empirical literature. Traditionally, cross-country

research on policy and growth uses simple linear models of the form

$$g = \eta + \alpha P + \beta' Z + \varepsilon \tag{1}$$

where g is the growth rate, P indicates the quality of macroeconomic policy, Z is a vector of other growth determinants, η and α are parameters, β is a parameter vector, and ε is an error term. This linear specification assumes that bad policies can be offset by other factors or, put differently, that the variables can smoothly substitute for one another. Yet many informal accounts of growth are phrased in terms of necessary conditions, which cannot be captured by a linear regression of the form in equation (1). There is surprisingly little research that considers necessary conditions in a formal way, with the exceptions of Hausmann, Rodrik, and Velasco (2008) on binding constraints and Hausmann, Pritchett, and Rodrik (2005) on the factors that instigate growth accelerations. A survey by Montiel and Servén (2006) also draws heavily on the binding-constraints perspective, and some additional discussion can be found in Temple (2009).

A simple way to address the problem is to examine the distribution of growth rates. If macroeconomic instability can form a binding constraint, unstable countries should have growth rates that are tightly distributed around a low mean, because instability is a sufficient condition for slow growth. By contrast, for stable countries growth rates should be more widely dispersed around a higher mean. Wide dispersion would arise because stable countries may lack other growth preconditions, leading to variation in performance across these countries driven by variation in other growth determinants (see figure 1, left panel, for hypothetical distributions of growth rates across countries).

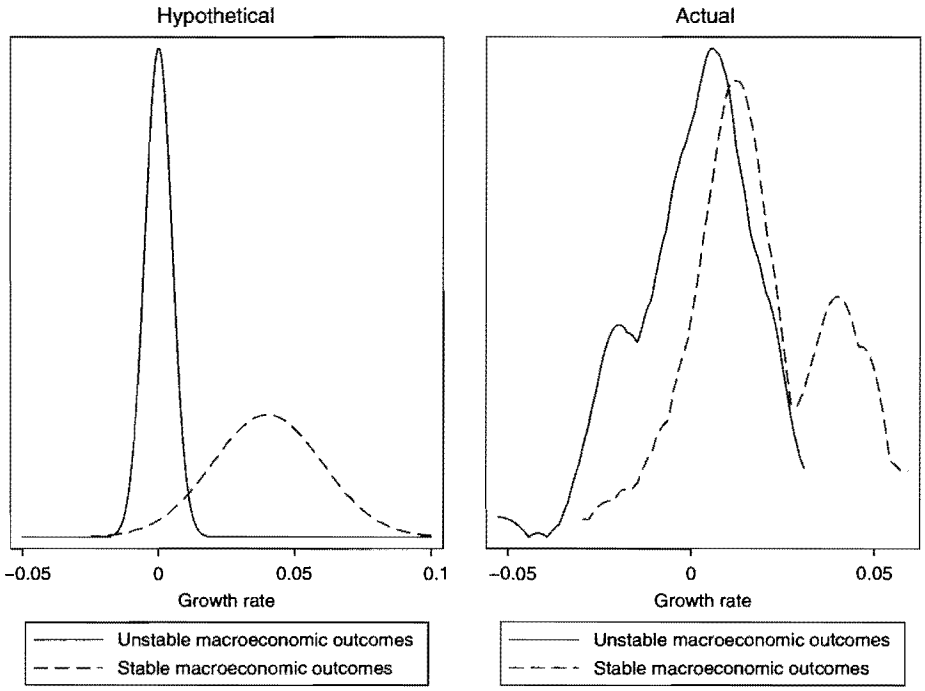
The binding constraints view also has implications for the specification of empirical growth models. One way to capture the idea is a simple nonlinear model with two regimes:

$$\begin{aligned} g &= \eta_1 + \varepsilon_1 && \text{if } P \leq \gamma \\ g &= \eta_2 + \alpha P + \beta' Z + \varepsilon_2 && \text{if } P > \gamma \end{aligned} \tag{2}$$

using similar notation to the previous example. The model implies that if the policy indicator P fails to exceed some threshold value γ, governments effectively destroy any prospect of growth, given low η_1 and a low variance of the error term ε_1 and regardless of other country characteristics. Section III uses Hansen's (1996, 2000) methods to estimate more general versions of equation (2) and shows that macroeconomic stability appears to be a more important threshold variable than other candidates, such as measures of geography and institutions.

The analysis here is based solely on cross-section variation, which has some advantages over a panel analysis for this research question. One drawback of panel data is that short-run deterioration in policy outcomes may be

FIGURE 1. Distributions of Growth Rates



Note: For clarity the distribution for intermediate macroeconomic outcomes is omitted.
Source: Authors' analysis based on data listed in table 1.

associated with a short-term growth slowdown, even if macroeconomic stability and growth are not associated in the long run (Bruno and Easterly 1998). The panel data approach could easily capture these short-run responses rather than genuine long-run effects on potential output. As Pritchett (2000) and Solow (2001) emphasize, models of growth are models of the evolution of potential output, and empirical analyses should be designed with this in mind.

Moreover, cross-section variation may be more informative than panel data about the effects of the ex ante prospects for stability, since a panel data analysis could be driven mainly by the effects of the realizations of outcomes. Given the spans of data currently available, there is a case for using cross-section data to identify the long-run impact of macroeconomic stability. A strong association between stability and growth in the international cross section would shift the burden of proof in the debate, placing new demands on those who argue that macroeconomic stability is largely irrelevant.

TABLE 1. Variables and Definitions

Variable	Description	Sources
ABSLAT	Absolute latitude (distance from the equator)	Hall and Jones (1999)
BD	Burnside-Dollar policy index	Burnside and Dollar (2000)
BMP	Log of (1 + mean black market premium)	Easterly and Sewadeh (2002)
ELR7097	Easterly, Levine, and Roodman update of Burnside-Dollar policy index	Easterly, Levine, and Roodman (2004)
ERATE	Variation of the Dollar real exchange rate measure	Dollar (1992)
EXPRISK	Protection against expropriation risk. Higher values mean lower risk. Mean value 1985–95.	Acemoglu, Johnson, and Robinson (2001)
FR	Log of a measure of natural openness to trade	Frankel and Romer (1999)
GOVKKM	A composite index of overall quality of governance that uses the mean of indexes for voice and accountability, political stability, and government during 1996–2000. Higher values indicate higher quality governance.	Kaufmann, Kraay, and Mastruzzi (2005)
INFLA	Log of (1 + median inflation rate based on GDP deflator)	World Bank (2004)
INVEST	Log of mean investment share in GDP, 1970–99	Heston, Summers, and Aten (2002)
LITERACY	Log of (100 − illiteracy rate of population ages 15 and older in 1970)	World Bank (2004)
MACRO	The first principal component from a classical principal components analysis of BMP, ERATE, INFLA, OVERVALU, and SURPLUS. Higher values indicate better policy outcomes.	See text
MACROOL	A macroeconomic stability index based on a classical principal components analysis that excludes Guyana, Nicaragua, and Sudan.	See text
OVERVALU	Log of mean overvaluation index. Dollar (1992) provides data for 1976–85. Easterly and Sewadeh (2002) update the data to 1999.	Dollar (1992); Easterly and Sewadeh (2002)
POLCON	A measure of the extent of political constraints in policymaking. A higher value implies stronger constraints. The mean value for 1970–99 is used.	Henisz (2000)
POLITY	A measure of the degree of democracy. The POLITY score is the democratic score minus autocratic score on a −10 to 10 scale, where higher values mean higher degree of democracy. The mean value for 1970–99 is used.	Marshall and Jaggers (2002)
POPG	Log of the average annual growth rate of the population ages 15–64 for 1970–99, plus 0.05.	World Bank (2004)
RGDP7099C	Log of real GDP per capita ("rgdpch") in 1999 minus the log of real GDP per capita for 1970. This is divided by 29, to obtain annual growth rates.	Heston, Summers, and Aten (2002)
RGDP7099W	Log of real GDP per worker ("rgdpwok") in 1999 minus that of 1970. This is divided by 29, to obtain annual growth rates.	Heston, Summers, and Aten (2002)

(Continued)

TABLE 1. Continued

Variable	Description	Sources
RGDPPC70	Log of real GDP per capita ("rgdpch") in 1970.	Heston, Summers, and Aten (2002)
RGNEAP	East Asia and Pacific regional dummy variable	Easterly and Sewadeh (2002)
RGNLAC	Latin America and the Caribbean regional dummy variable	Easterly and Sewadeh (2002)
RGNMENA	Middle East and North African regional dummy variable	Easterly and Sewadeh (2002)
RGNSA	South Asian regional dummy variable	Easterly and Sewadeh (2002)
RGNSSA	Sub-Saharan African regional dummy variable	Easterly and Sewadeh (2002)
RMACRO	The first principal component from a robust principal components analysis.	See text
SCHOOL70	Log of average years of schooling at all educational levels of population age 15 and older in 1970.	Barro and Lee (2001)
SURPLUS	Mean central government budget surplus as a share of GDP, 1970–99	World Bank (2004)

Source: Authors' construction.

II. A NEW WAY TO MEASURE MACROECONOMIC STABILITY

This section introduces a new index of macroeconomic stability that combines several indicators, and uses it to measure the average extent of stability over 1970–99. This combination has two advantages. From a statistical point of view, it lessens the outlier problems associated with skewed distributions. And from an economic point of view, it aims to capture an underlying latent variable, the quality of the macroeconomic decision-making process, rather than relying on more specific "symptoms" such as high inflation. Using several proxies for this latent variable reduces measurement error and makes sense if, as suggested by Sala-i-Martin (1991), macroeconomic disarray is associated with undesirable outcomes across a range of indicators. This approach acknowledges the difficulty in identifying the separate effects of fiscal discipline, inflation control, and exchange rate management in small cross-country data sets. Instead, it makes sense to reduce the dimensions of the problem and focus on a single index of policy outcomes. Arguably, there is more hope of answering questions about policy outcomes and growth when the relevant hypotheses are deliberately characterized in broad terms, given the limitations of the available data.

The composite measure is based on fiscal discipline, inflation, and exchange rate management. The preferred index is based on an outlier-robust version of principal components analysis, using Rousseeuw's (1984) minimum covariance determinant method. The empirical analysis discussed later focuses on developing economies with available data, excluding transition economies and countries with small populations (fewer than 250,000 people in 1970). The

main indicators were constructed from a sample of 78 countries; data availability means that the growth regressions discussed later use 60–70 observations, while the Bayesian model averaging in section IV uses 72 observations. See table 1 for definitions and sources of variables used in the analysis, and table 4 in section III for a list of countries.

The individual policy indicators are as follows. Fiscal discipline is measured using data on the average central government budget surplus as a share of GDP (*SURPLUS*) over 1970–99.[3] Some countries, notably Guyana and Sudan, have extreme negative values for this variable, reflecting persistently high budget deficits. The principal component analysis, and hence the later results, is robust to excluding these countries or replacing *SURPLUS* with the monotonic but bounded transformation, arctan (*SURPLUS*).[4]

Success in keeping inflation low is captured in the variable *INFLA*. This is the natural log of 1 plus the median inflation rate over 1970–99, computed from the GDP deflator. The median inflation rate is used to capture success in keeping inflation low on average. Relative to the more commonly used of the mean, this measure is less at risk of being dominated by short-lived episodes of hyperinflation.

Exchange rate management is measured in three ways: the black market premium (*BMP*), an index of currency overvaluation or real exchange rate distortion (*OVERVALU*), and a measure of the variability in exchange rate distortions (*ERATE*). The black market premium reflects departures of an illegal, market-determined exchange rate from the official exchange rate. To lessen outlier problems, *BMP* is defined as the natural log of 1 plus the mean value of the black market premium over the period.

Dollar (1992) introduced the variables *OVERVALU* and *ERATE*, whereas Easterly and Sewadeh (2002) extended *OVERVALU* forward and backward. *OVERVALU* is based on evaluating price levels in a common currency, after correcting for the possible effects of factor endowments on the prices of nontradables by using the component of price levels that is orthogonal to GDP per capita and its square, population density, and two regional dummy variables. A price level higher than predicted by these controls indicates that the domestic prices for tradables may be high; thus high values of *OVERVALU* could indicate a combination of real overvaluation and trade restrictions. The precise interpretation of this measure is discussed further in the appendix.

ERATE is a measure of variability in the overvaluation index for 1976–85 (see table A-1 in Dollar 1992) and can be seen as capturing instability in

3. An alternative would be the stock of central government debt relative to GDP, but *SURPLUS* is available for more countries.

4. This transformation is a natural choice, given that the variable is a ratio that can take on extreme values in either direction, positive or negative. The arctan(x) function maps x into the smallest or most basic angle with tangent x. When the angle is expressed in radians, the values of the arctan function will be restricted to the interval $(-\pi/2, \pi/2)$ and this will limit the effect of outlying observations. When the transformation is applied to *SURPLUS*, the lowest value is less than 1 standard deviation below the mean, compared with 5 standard deviations below in the raw data.

exchange rate management. Given the probable role of inflation in generating movements in the overvaluation index, it may also indicate more general forms of macroeconomic instability (Rodriguez and Rodrik 2000).

Although the analysis sometimes uses the five outcome indicators individually, they are usually aggregated into a composite index. The best known such index in the recent literature is that of Burnside and Dollar (2000), who construct an aggregate measure of policy quality based on three indicators: inflation, the budget surplus, and the Sachs and Warner (1995) indicator of openness to trade.[5] Since Burnside and Dollar's focus is a possible interaction between the growth effects of aid and the quality of policy, they weight the policy indicators using the coefficients in a simple regression of growth on the indicators and controls, including initial GDP, regional dummy variables, and proxies for political stability. This procedure is less suited to the aims of this study. In their procedure, growth will typically be correlated with the aggregate policy index by construction. But here the aim is to compare distributions of growth rates across countries with good and bad policy outcomes, which requires a composite policy index that does not use information on growth rates.

The five separate variables are aggregated using a principal components analysis. The first step is to check that the correlations between the variables are high enough to justify using principal components: in the extreme case, where the variables were all pairwise uncorrelated, a principal components analysis would not make sense. A likelihood ratio test can be used to examine that "sphericity" case, allowing for sampling variability in the correlations. This test comfortably rejects sphericity at the 1 percent level (for more details, see the supplemental appendix at http://wber.oxfordjournals.org/).

The first principal component is always normalized in such a way that high values indicate macroeconomic stability (table 2). In terms of standardized indicators (all with mean 0 and variance of 1) the first index can be written as

$$\begin{aligned} MACRO = {} & 0.334 * SURPLUS - 0.447 * INFLA - 0.585 * BMP \\ & - 0.347 * OVERVALU - 0.475 * ERATE. \end{aligned} \tag{3}$$

This index places most weight on the black market premium and the Dollar (1992) measure of variability in exchange rate distortions. The first principal component explains 42 percent of the total variance in the standardized data. According to this index, the governments that were most successful in achieving macroeconomic stability during 1970–99 were Singapore, Thailand, Malaysia, Panama, and Benin. By contrast, the analysis suggests that

5. Burnside and Dollar (2000) also experiment with government consumption as a share of GDP but find it to be negatively correlated with the budget surplus and insignificant when the budget surplus is included.

TABLE 2. Results of Principal Components Analysis

Variable	Expected sign	MACRO 1st principal component	MACRO 2nd principal component	RMACRO 1st principal component	RMACRO 2nd principal component	MACROOL 1st principal component	MACROOL 2nd principal component
SURPLUS	+	0.484	0.579	0.340	0.297	0.276	0.768
INFLA	−	−0.647	0.437	−0.744	0.172	−0.727	0.161
BMP	−	−0.848	0.184	−0.888	−0.034	−0.843	0.120
OVERVALU	−	−0.503	−0.633	−0.395	−0.951	−0.327	−0.654
ERATE	−	−0.688	0.232	−0.653	−0.164	−0.665	0.311
Number of countries		78		78		75	
Variance explained (percent)		41.94	20.29	41.27	24.00	37.29	23.10

Note: Values are the correlation between principal components and the corresponding variables. Numbers in bold indicate the highest correlations between a given principal component and corresponding variables. See table 1 for definitions and sources of variables.

Source: Authors' analysis based on data listed in table 1.

Nicaragua, Guyana, Sudan, Uganda, and Zambia were characterized by long-term instability.

A drawback of principal components analysis, especially in a small sample, is the inherent sensitivity to outlying observations. As Hubert, Rousseeuw, and Branden (2005) note, a classical principal components analysis maximizes the variance and decomposes the covariance matrix, both of which can be highly sensitive to outliers. This is an important concern when aggregating measures of macroeconomic outcomes. Easterly (2005) points out that the empirical distributions of macroeconomic outcomes are often heavily skewed, with a small number of countries experiencing outcomes that are unusually bad (several standard deviations from the mean) relative to other developing economies.

For this reason, the main focus of this article is an alternative index, based on an outlier-robust principal components analysis. The relatively small dimensions of the problem suggest the use of the minimum covariance determinant method, which identifies the particular subset of $h < n$ observations, among the many possible subsets of the total set of n observations, for which the classical covariance matrix has the smallest determinant (a method from Rousseeuw 1984; see also Rousseeuw and van Driessen 1999). The covariance matrix for these h observations can be used to represent the associations among the variables and to compute the eigenvectors associated with the principal components. The standard choice $h = 0.75n$ will be used, so that the method effectively discards the least representative 25 percent of the cases in estimating the correlations, building in a high degree of robustness.[6]

This approach to estimating correlations can then be used to extract outlier-robust principal components. The correlations between the first two of these new principal components and the individual policy indicators are shown in the *RMACRO* column of table 2. In terms of loadings on the individual variables, the robust index can be written as:

$$\begin{aligned} RMACRO = {} & 0.101 * SURPLUS' - 0.578 * INFLA' - 0.693 * BMP' \\ & - 0.219 * OVERVALU' - 0.357 * ERATE', \end{aligned} \tag{4}$$

where each variable has now been centered using a robust estimate of its location. Relative to the classical principal components analysis, the outlier-robust principal components analysis places less weight on *SURPLUS*, *OVERVALU*, and *ERATE* and more weight on *INFLA* and *BMP*. Although the weights in the two cases may look different, the simple correlation between *MACRO* and *RMACRO* is 0.98, reflecting high correlations between some of

6. The ROBPCA program can be used to implement the minimum covariance determinant approach. The simpler alternative of identifying outliers from bivariate scatter plots is flawed because it will not always detect observations that are outliers in a multidimensional space. Also, using an outlier-robust approach to principal components analysis does not preclude the possibility of extreme (and hence informative) observations in the final index. Rather, the idea is to limit the influence of small numbers of observations on the weighting scheme used in constructing the index.

TABLE 3. Correlations between GDP Growth and Various Policy Indexes

Policy index	RGDP7099C	MACRO	RMACRO	MACROOL	BD	ELR7097
RGDP7099C	1.000					
MACRO	0.471	1.000				
RMACRO	0.420	0.976	1.000			
MACROOL	0.409	0.995	0.991	1.000		
BD	0.673	0.666	0.623	0.585	1.000	
ELR7097	0.590	0.603	0.621	0.645	0.850	1.000

Note: See table 1 for definitions and sources of variables. Sample size varies between 64 and 78 countries, depending on data availability.

Source: Authors' analysis based on data listed in table 1.

the individual components. With the *RMACRO* index, the five best performers are Singapore, Thailand, Panama, Malaysia, and Togo, and the five worst performers are Nicaragua, Uganda, Ghana, Argentina, and the Democratic Republic of Congo.

An alternative approach would be to use the diagnostic plot suggested by Hubert, Rousseeuw, and Branden (2005), which can identify possible outliers that are then excluded from an otherwise standard principal components analysis. This method indicates that Guyana, Nicaragua, and Sudan might be anomalous observations. However, the *MACROOL* column of table 2 shows that this makes little difference. The proportion of the variance explained by the first principal component falls slightly, but the correlations between this component and the different indicators are similar to those reported in the *MACRO* and *RMACRO* columns.

The correlations between *MACRO*, *RMACRO*, the Burnside-Dollar index, and the updated Burnside-Dollar index for 1970–97 from Easterly, Levine, and Roodman (2004) are high (table 3), suggesting that the various indexes may be capturing an underlying latent variable. This is the case even though the Burnside–Dollar and Easterly–Levine–Roodman measures use a different weighting strategy as well as the Sachs-Warner measure of liberal policies, including trade policies. At the same time, the correlations clarify that the results in sections III and IV should not be interpreted too literally. A measure that is notionally of macroeconomic stability may capture other aspects of policy or equilibrium outcomes, especially when instability is a symptom of a dysfunctional policy environment or periods of conflict.

III. Is Macroeconomic Instability the Weakest Link?

The preferred index, *RMACRO*, is now used to examine how growth varies across countries with good and bad macroeconomic outcomes. Ordering the countries by *RMACRO* and splitting the sample at the 33rd and 66th percentiles yields three groups of countries (table 4). The distributions of growth rates

can then be compared across these groups. The growth rate is measured in annual terms, based on GDP per capita (chain weighted) over 1970–99, using data from version 6.1 of the Penn World Table (Heston, Summers, and Aten 2002).

The median growth rate is substantially lower for the relatively unstable group 1 than for groups 2 and 3 (see figure 2, left panel; group 1 is the least stable, group 3 the most stable). There is less support for the idea that macroeconomic instability always destroys long-term growth prospects, because even in group 1, the 75th percentile of the growth rate is 1.4 percent. The patterns are similar (not shown) when growth is measured using GDP per worker rather than GDP per capita and when classifying countries according to *MACRO* rather than *RMACRO*.

Kernel density plots can be used to summarize the same information in a slightly different way.[7] Stable countries have higher growth on average, but instability does not necessarily preclude growth (figure 1, right panel). There is substantial variation in growth across the countries with unstable outcomes, and a significant fraction display positive growth rates over the period. Nevertheless, there are no countries growing at more than 3.5 percent a year in the unstable group, whereas seven countries in the stable group grew at least this rapidly (Cyprus, Indonesia, Republic of Korea, Malaysia, Mauritius, Singapore, and Thailand). Based on this evidence, macroeconomic stability is a necessary condition for sustaining high growth rates over a long period.

An alternative method is to examine the box plots for all five individual indicators, *SURPLUS, INFLA, BMP, OVERVALU*, and *ERATE*. The patterns (not shown) are generally less supportive of the idea that stability promotes growth, suggesting that combining the indicators into an overall index is worthwhile. The evidence that stability matters is strongest when the Dollar (1992) index of exchange rate distortions (*OVERVALU*) and the black market premium (*BMP*) are used to group countries (see figure 2, right panel, for results using the black market premium).

Growth Regressions

This subsection uses growth regressions to examine the relationship between macroeconomic stability and growth. Conventional linear models are used, estimated by ordinary least squares and two-stage least squares, starting with Mankiw, Romer, and Weil's (1992) version of the Solow model. This is arguably the leading structural model in the literature, and it reduces arbitrariness in the choice of specification. The model is estimated using data for 1970–99 rather than for 1960–85 as in Mankiw, Romer, and Weil. Even conditional on the investment rate, population growth, initial income, and regional dummy

7. The samples are relatively small to apply these methods, and the choice of bandwidth becomes important. This is discussed in the supplemental appendix, available at http://wber.oxfordjournals.org/

TABLE 4. RMACRO Values and Grouping, by Country

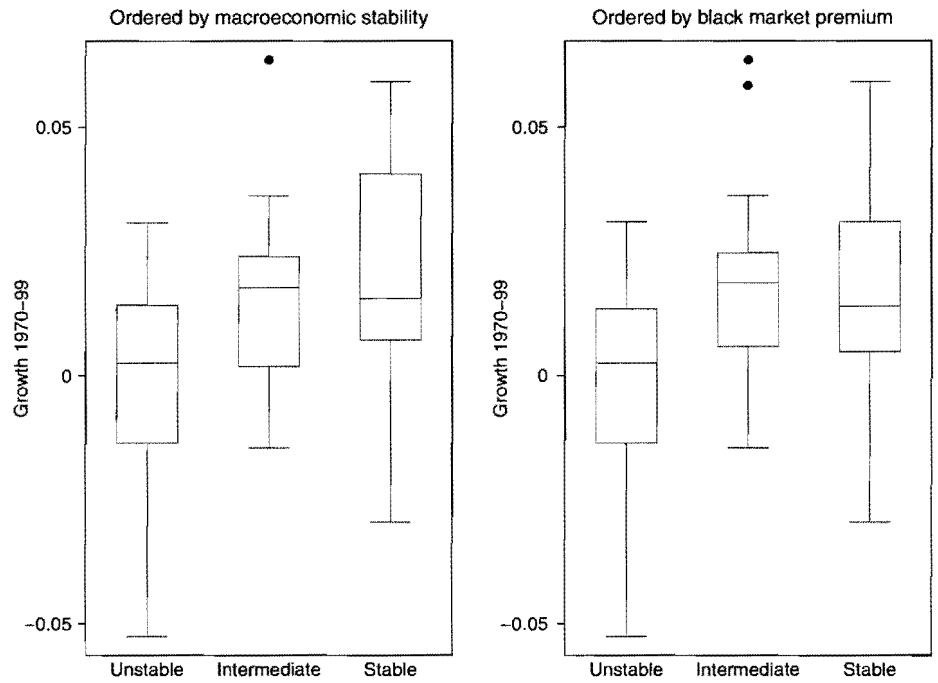
Number	Country	RMACRO	Group	Number	Country	RMACRO	Group
1	Nicaragua	−2.974	1	40	Ethiopia	0.161	2
2	Uganda	−2.009	1	41	Sri Lanka	0.165	2
3	Ghana	−1.680	1	42	Mexico	0.237	2
4	Argentina	−1.669	1	43	Madagascar	0.277	2
5	Congo, Dem. Rep.	−1.610	1	44	Lesotho	0.310	2
6	Guyana	−1.547	1	45	Colombia	0.325	2
7	Iran	−1.504	1	46	Kenya	0.348	2
8	Sudan	−1.476		47	Trinidad and Tobago	0.352	2
9	Sierra Leone	−1.463	1	48	Nepal	0.352	2
10	Somalia	−1.266		49	India	0.364	2
11	Zambia	−1.254	1	50	Botswana	0.371	2
12	Bolivia	−1.185	1	51	Pakistan	0.379	2
13	Brazil	−1.165	1	52	Nigeria	0.560	2
14	Peru	−1.115	1	53	Papua New Guinea	0.566	2
15	El Salvador	−1.061	1	54	Philippines	0.622	3
16	Liberia	−0.911		55	Indonesia	0.669	3
17	Niger	−0.696	1	56	South Korea	0.684	3
18	Algeria	−0.661		57	Tunisia	0.686	3
19	Uruguay	−0.655	1	58	Ecuador	0.763	3
20	Egypt	−0.555	1	59	Mauritius	0.790	3

21	Syria	−0.522	1	60	Congo, Rep.	0.832	3
22	Venezuela	−0.493	1	61	Morocco	0.842	3
23	Jamaica	−0.491	1	62	Mali	0.865	3
24	Yemen	−0.460		63	Chad	0.927	3
25	Zimbabwe	−0.454	1	64	Cameroon	0.954	3
26	Turkey	−0.441	1	65	Gabon	0.983	3
27	Mauritania	−0.399	1	66	Cyprus	0.989	3
28	Costa Rica	−0.371	1	67	Oman	1.124	
29	Paraguay	−0.360	1	68	Central African Rep.	1.126	3
30	Chile	−0.360	2	69	Burkina Faso	1.139	3
31	Malawi	−0.338	2	70	Senegal	1.153	3
32	Haiti	−0.311	2	71	Benin	1.210	3
33	Rwanda	−0.237	2	72	Fiji	1.219	3
34	Israel	−0.137	2	73	Jordan	1.246	3
35	Honduras	−0.123	2	74	Togo	1.371	3
36	Burundi	−0.026	2	75	Malaysia	1.607	3
37	Dominican Rep.	0.003	2	76	Panama	1.652	3
38	Guatemala	0.078	2	77	Thailand	1.742	3
39	Bangladesh	0.102	2	78	Singapore	1.837	3

Note: RMACRO is listed for the 78 countries used in table 2, ordered from worst to best. The 72 countries with a group number are for countries included in figures 1–4. Group indicator refers to the groups underlying the left panel in figure 2. The main 70 country regression sample is based on the same set of countries, minus Gabon and Sierra Leone, for which the literacy indicator was unavailable.

Source: Authors' analysis based on data listed in table 1.

FIGURE 2. Box Plots for Growth Rates



Note: The upper and lower limits of each enclosed box correspond to the 75th and 25th percentiles of the growth rate, while the horizontal line within each box corresponds to the median. "Unstable" refers to group 1 countries in table 4, "intermediate" to group 2 countries, and "stable" to group 3 countries.

Source: Authors' analysis based on data listed in table 1.

variables, a significant partial correlation is found between growth and macroeconomic stability.

The specification relates the log difference in GDP per capita to the log of the investment rate, the log of initial GDP per capita, the log of population growth plus 0.05, and a human capital variable, as in Mankiw, Romer, and Weil (1992). There are two main departures in the current specification. First, regional dummy variables are used to proxy for the initial level of efficiency, as in Temple (1998). Second, the regressions use a measure of the initial level of educational attainment rather than the rate of investment in human capital.[8] This will be the natural log of either the 1970 literacy rate (from World Bank 2004) or average years of schooling in 1970 (from Barro and Lee 2001). In both cases, the data refer to the population ages 15 and older.

Regression 1 excludes the policy indicators (table 5). The Mankiw, Romer, and Weil (1992) regression continues to work well over a different time period;

8. The use of a stock measure rather than a flow can be justified formally as a proxy for the steady-state level of educational attainment, as in equation (12) in Mankiw, Romer, and Weil (1992).

TABLE 5. Macroeconomic Stability and Growth Regressions

Variable	1 OLS	2 OLS	3 OLS	4 OLS	5 OLS	6 2SLS	7 OLS	8 OLS
Regime	All	All	All	All	All	All	1	2
Number of observations	70	70	70	70	60	70	42	28
RMACRO		0.71 (0.30)	0.49 (0.31)	0.64 (0.27)	0.64 (0.29)	1.35 (0.66)	0.70 (0.42)	−1.20 (0.39)
Initial income	−1.10 (0.37)	−0.26 (0.37)	−0.80 (0.37)	−1.04 (0.38)	−1.15 (0.42)	−0.98 (0.33)	−0.83 (0.43)	−1.26 (0.30)
Investment	1.07 (0.32)		1.10 (0.34)	0.83 (0.32)	0.84 (0.48)	0.56 (0.42)	0.45 (0.42)	1.52 (0.41)
Population growth	−0.21 (0.23)		−0.19 (0.22)	−0.12 (0.25)	−0.10 (0.28)	−0.02 (0.23)	−0.15 (0.29)	0.08 (0.15)
LITERACY	0.68 (0.31)			0.88 (0.34)		1.12 (0.32)	0.72 (0.36)	0.41 (0.35)
SCHOOL70					0.79 (0.27)			
GOVKKM							1.06 (0.98)	1.91 (0.44)
Investment elasticity	1.18		1.69	0.96	0.89	0.69	0.66	1.47
R^2	0.51	0.37	0.51	0.57	0.55	n/a	0.47	0.90
Regression standard error	1.56	1.75	1.57	1.47	1.58	1.47	1.38	0.84
Heteroscedasticity								
Breusch-Pagan	0.32	0.02	0.07	0.27	0.18	0.48	0.01	0.09
White	0.66	0.19	0.03	0.64	0.35	0.66	0.07	0.46
Ramsey RESET	0.90	0.58	0.02	0.68	0.24	0.97	0.01	0.84
Anderson-Rubin						0.02		

Note: OLS is ordinary least squares. 2SLS is two-stage least squares. The dependent variable is the annual growth rate over 1970–99 in percentage points. Numbers in parentheses are MacKinnon-White heteroskedasticity-consistent (hc3) standard errors, except for regression 6, for which numbers in parentheses are White heteroskedasticity-consistent standard errors. Constants are included but not reported. Regressions 1–6 include five regional dummy variables, for which the coefficients are not reported. The explanatory variables are standardized to have a standard deviation of 1 in the 70 country sample. Investment elasticity is the elasticity of the steady-state income level to the investment rate. Heteroscedasticity reports p-values associated with two tests for heteroscedasticity. Ramsey RESET (regression equation specification error test) is the p-value associated with this test. Anderson-Rubin is the p-value associated with the Anderson-Rubin test for the significance of the endogenous explanatory variable (RMACRO). See table 1 for definitions and sources of variables.

Source: Authors' analysis based on data listed in table 1.

the explanatory power is similar, although the effect of population growth is imprecisely estimated. The elasticity of steady-state income to the investment rate is 1.18, within the range spanned by Mankiw, Romer, and Weil's estimates. Regression 2 includes only initial income, regional dummy variables, and the new measure of stability, *RMACRO*. The stability measure is significant at the 5 percent level, and the association is strong: if interpreted as a causal effect, a 1 standard deviation improvement in stability would have raised the annual growth rate by 0.71 percentage point over the time period. Regression 3 controls for the effects of investment and population growth, as in Mankiw, Romer and Weil. The effect of *RMACRO* is slightly weaker, as might be expected, but significant at the 12 percent level. The reduction in the size of the coefficient indicates that macroeconomic stability may boost investment, an idea that will be explored later.

Regression 4 includes *LITERACY*, the log of the 1970 literacy rate, which increases the explanatory power of the model. *RMACRO* is once again significant at the 5 percent level. This result is robust to replacing the literacy rate with the log of average years of schooling in 1970, *SCHOOL70*, as in regression 5. This reduces the size of the sample by 10 observations, so regression 4 is the preferred specification in the discussion that follows.

The partial correlations between growth and macroeconomic stability do not appear to be driven by anomalous observations. The results are robust to the deletion of potential outliers, as identified by least absolute deviation regressions.[9] The findings are similarly robust to using single-case diagnostics such as DFFITS and DFBETA, which identify a similar set of outliers to the least absolute deviation method in this case.[10] Added-variable plots (not shown) were also used to identify potential outliers. When the Democratic Republic of Congo and Nicaragua are excluded, the results are slightly less strong, with *RMACRO* significant only at the 8 percent level. Finally, some simple diagnostic tests are supportive: the models do not suffer from omitted nonlinearities (based on the Ramsey RESET test) or heteroskedasticity (based on versions of the Breusch–Pagan and White tests) except in regression 3, which includes investment but not a measure of human capital.

Given the concern that macroeconomic stability is not randomly assigned, an instrumental variable approach might be preferable. One possible route exploits the observed association between French colonial heritage and macroeconomic stability, as in Barro (1996). Many former French colonies maintained a fixed exchange rate with the French franc, and this appears to have been associated with lower inflation rates. The sample contains 15 former French colonies, and for these countries the mean of *RMACRO* is 0.52 and the

9. Outliers were defined by least absolute deviation residuals more than two standard deviations from the mean value.

10. The results are available on request. See Cook and Uchida (2003) for a brief discussion of how DFFITS and DFBETA are computed and used.

standard deviation is 0.72. This compares favorably to a mean of 0.01 and standard deviation of 1.03 for former British colonies and, since *RMACRO* is standardized, to a mean of 0 and a standard deviation of 1 for the sample as a whole.

Regression 6 instruments *RMACRO* using a dummy variable for former French colonies. The significance of *RMACRO* is tested using the Anderson and Rubin (1949) statistic, which is optimal for models that are just-identified (Moreira 2003) and should be robust to weak-instrument problems. The p-value associated with the test is 0.02, so *RMACRO* is significant even in the two-stage least squares estimates. The two-stage least squares estimate assigns more weight to macroeconomic stability and less to investment than the ordinary least squares point estimate does. The finding that the two-stage least squares coefficient for *RMACRO* is considerably higher than the ordinary least squares coefficient could be due to measurement error or sampling variability, as Acemoglu, Johnson, and Robinson (2001) and Frankel and Romer (1999) have argued in other contexts. But it could also be due to a failure of the exclusion restriction, so these results should be treated cautiously. The small number of observations reinforces this point. To lessen endogeneity problems arising from the nonrandom assignment of policy, the approach used in the next subsection, namely a comprehensive search through a wide range of control variables and specifications, may be preferable.[11]

In summary, there is an association between macroeconomic stability and growth, even conditional on investment rates. Taking the results at face value, a 1 standard deviation improvement in stability translates into an annual growth rate that is 0.5–0.7 percentage point higher over 30 years. Increasing the annual growth rate by 0.7 percentage point would leave GDP per capita 23 percent higher after 30 years. A later analysis will consider the implications for the location and shape of the distribution of growth rates and the steady-state distribution of GDP per capita.

Threshold Estimation

This subsection uses Hansen's (1996, 2000) methods for sample splitting and threshold estimation to estimate nonlinear models of the following type:

$$\begin{aligned} g &= \eta_1 + \alpha_1 P + \beta_1' Z + \varepsilon_1 \quad \text{if } P \leq \hat{\gamma} \\ g &= \eta_2 + \alpha_2 P + \beta_2' Z + \varepsilon_2 \quad \text{if } P > \hat{\gamma}, \end{aligned} \tag{5}$$

11. Moreover, the applicability of instrumental variable approaches to cross-country growth data may have been exaggerated. When the instrument is correlated with the error term, even weakly, the inconsistency of the instrumental variable estimator can be worse than that of the ordinary least squares estimator, particularly if the instrument is not strongly correlated with the endogenous explanatory variable (see Cameron and Trivedi 2005). There are good reasons to doubt many of the exclusion restrictions adopted in the literature, since most candidates for instruments might be correlated with omitted growth determinants; see Durlauf, Johnson, and Temple (2005, 2009) for more discussion.

where $\hat{\gamma}$ is a threshold estimated jointly with the other parameters in the model and P could be an indicator of macroeconomic outcomes or some other variable, such as a measure of institutional quality. This specification nests the earlier example, equation (2), since the intercept and slope coefficients are allowed to vary across the two regimes. A particular strength of the Hansen approach is that alternative candidates for the threshold variable P can be compared on statistical grounds. Moreover, by comparing the models for different regimes, it is possible to see whether macroeconomic instability forms a binding constraint on growth. If so, instability should limit the benefits of favorable fundamentals, such as geographic and institutional characteristics.

It is possible to test for the existence of a threshold, and hence multiple regimes, using the Hansen (1996) bootstrapped Lagrange multiplier test. Hansen (2000) develops an asymptotic approximation to the least squares estimate of the threshold $\hat{\gamma}$, which allows construction of a (possibly asymmetric) confidence interval. These methods can therefore reveal the extent to which a proposed sample split is estimated with precision and whether the proposed nonlinearity is supported by the data.[12] As in the earlier analysis, the main limitation arises from the possible correlation between macroeconomic stability and the error term, which brings the risk that the assignment of countries across regimes could also be a function of the error term, and so the results should be cautiously interpreted.

Seven possible candidates for the threshold variable P are considered—*RMACRO* and six indicators of either geographic or institutional fundamentals—to determine whether differences in macroeconomic stability give rise to distinct growth regimes or whether fundamentals provide a better way to divide the sample. Two of the fundamentals considered are standard measures of geographic characteristics. The first variable, *FR*, is the log of the Frankel and Romer (1999) measure of natural openness to trade, which is based partly on proximity to large markets. The second variable, *ABSLAT*, is absolute latitude—that is, distance from the equator. In both cases, the data are taken from Hall and Jones (1999).

The remaining four candidates for threshold variables are all measures of institutional quality. These are *GOVKKM*, a composite index of the quality of governance for 1996–2000, from Kaufmann, Kraay, and Mastruzzi (2005); *POLITY*, the extent of democracy, based on the Polity IV database of Marshall and Jaggers (2002) and averaged over 1970–99; *POLCON*, a measure of the extent of political constraints from Henisz (2000) averaged over 1970–99; and *EXPRISK*, the measure of average expropriation risk for 1985–

12. Previous applications of these methods to growth regressions include Hansen (2000) and Papageorgiou (2002). In emphasizing institutions as a potential threshold variable, this article is especially close to the work of Minier (2007) but considers the role of macroeconomic stability in more detail.

95 used in Acemoglu, Johnson, and Robinson (2001). Several of these measures are based partly on observed outcomes rather than constraints. This may lead the benefits of good institutions to be overstated and the benefits of macroeconomic stability to be understated.[13]

For each of the six fundamental variables, a regression is used to relate growth to that variable, the Solow variables, and *RMACRO*. Regional dummy variables are omitted to avoid overfitting problems when the sample is subdivided. Hansen's approach is used to test for a threshold associated with *RMACRO* and alternatively with the fundamental variable.

It is immediately apparent that *RMACRO* dominates all the other candidates as a threshold variable (table 6). In all but one case the null of no threshold is rejected for *RMACRO* at the 10 percent level, while it is not rejected for any of the other six measures of fundamentals. These results suggest that the data are well described by two regimes, where the classification of countries into the two regimes depends on macroeconomic stability rather than geography or institutions. The estimated threshold for *RMACRO* is also reported in table 6, along with its 95 percent confidence interval (which may be asymmetric) and the number of countries in each subsample. Since *RMACRO* is normalized to have a mean of 0 and a standard deviation of 1 in the 70 country sample, it is clear that the threshold is precisely estimated and relatively stable across the various specifications.

An especially interesting result is that when the sample is divided using the estimated threshold, the standard growth variables have much higher explanatory power for the relatively stable countries. For this group, the model typically accounts for 75–90 percent of the variation in growth rates, while the R^2 for the less stable countries is typically 40–50 percent. This is consistent with the binding-constraints view: if macroeconomic stability is achieved, growth is well explained by a standard regression, but the Solow variables (and measures of geographic or institutional fundamentals) have less explanatory power when instability forms a binding constraint on growth, since this limits the benefits of favorable characteristics. The main departure from the earlier hypothesis is that the cross-section residual variance is higher, not lower, for countries that experience macroeconomic instability.[14]

Regressions 7 and 8 show the results for the two groups and are based on a model containing the Kaufmann, Kraay, and Mastruzzi (2005) measure (*GOVKKM*), so the candidate variables for a threshold were *GOVKKM* and *RMACRO*. As in the other cases, the Hansen (1996) test favored macroeconomic stability for splitting the sample. The estimated threshold for *RMACRO*, $\hat{\gamma} = 0.297$, is slightly above the mean and divides the sample into

13. See Glaeser and others (2004) on the general desirability of using measures of constraints or rules rather than measures closely related to outcomes.

14. This is consistent with a competing explanation for the results, namely that measurement errors in the data are more serious for unstable countries.

TABLE 6. Threshold Estimation

Z variable	FR	ABSLAT	GOVKKM	POLITY	POLCON	EXPRISK
RMACRO threshold	0.068	0.167	0.030	0.002	0.013	0.012
Z threshold	0.324	0.320	0.271	0.523	0.600	0.354
γ – RMACRO	0.309	0.180	0.297	−0.185	0.297	0.309
95 percent confidence interval						
Lower	−0.375	−0.300	−0.520	−0.375	−0.520	−1.241
Higher	0.309	0.309	0.714	0.309	0.618	0.324
N [RMACRO $\leq \gamma$]	43	36	42	29	42	36
N [RMACRO $> \gamma$]	27	34	28	40	28	19
R^2 [RMACRO $\leq \gamma$]	0.43	0.60	0.47	0.40	0.42	0.48
R^2 [RMACRO $> \gamma$]	0.80	0.76	0.90	0.77	0.78	0.82
Z p-value [RMACRO $\leq \gamma$]	0.90	0.00	0.29	0.57	0.85	0.49
Z p-value [RMACRO $> \gamma$]	0.29	0.09	0.00	0.07	0.51	0.04

Note: RMACRO threshold is the p-value for the Hansen (1996) test of a threshold in RMACRO, in a model that includes RMACRO, the Solow variables, and the Z variable. Z threshold is the p-value for the Hansen (1996) test of a threshold associated with the Z variable. The tests indicate that RMACRO can be used to divide the sample into two regimes. The lower rows show the threshold γ for RMACRO estimated using the Hansen (2000) procedure; the 95 percent confidence interval for the threshold (which need not be symmetric); the number of observations in the two regimes on either side of the threshold; the R^2 of the separate growth regressions for the two regimes; and the p-value of the Z variable for each of the two regimes. The growth regression always has the highest explanatory power in the subsample with greater macroeconomic stability; for an example based on GOVKKM, see regime 1 in regression 7 and regime 2 in regression 8 of table 5. See table 1 for definitions and sources of variables.

Source: Authors' analysis based on data listed in table 1.

42 unstable countries (regression 7) and 28 relatively stable countries (regression 8). Comparing these two sets of results shows that macroeconomic stability is clearly associated with a higher elasticity of steady-state output to the investment rate, faster conditional convergence, and perhaps stronger growth benefits of good institutions. Overall, the explanatory power of the growth regression is much higher and the specification tests more favorable for the stable group. By contrast, a Ramsey RESET test for the less stable group rejects the Solow specification.

Across the six specifications summarized in table 6, a less plausible result is that in the subsamples with relatively stable macroeconomic outcomes *RMACRO* often has a negative sign and is sometimes significant at conventional levels (see the results for regression 8 in table 5). The result that stability has a significantly negative effect in this particular group should be interpreted with caution. It does not arise when the control variable is *ABSLAT*, *POLITY*, or *EXPRISK*. Any significantly negative relationship that emerges may be related to a conditional convergence effect. By construction, all countries in the second regime must have achieved a certain degree of stability, but some may combine instability (relative to other members of the stable group) with strong potential for rapid growth. Simply including initial income as an explanatory

variable may not be enough to eliminate such effects. This interpretation of the evidence would be consistent with the idea that once a certain degree of stability has been achieved, the benefits of greater stability may be limited.[15]

Finally, the role of fundamentals (geography and institutions) is considered in more detail. The last two rows of table 6 report the *p*-values associated with these variables for the unstable and stable groups of countries. They show that the posited fundamentals usually lack explanatory power in the less stable countries but often emerge as significant for the more stable countries. Again, this supports an account in terms of binding constraints.

IV. Robustness

This section uses Bayesian methods to examine the robustness of the partial correlation between growth and macroeconomic stability. Levine and Renelt (1992) showed that partial correlations in the empirical growth literature may not be robust to changes in specification. This is a serious problem for growth researchers because the list of candidate predictors is long and it is not easy to rule out particular variables or specifications using prior reasoning. Put differently, there is a model uncertainty problem, and the standard errors in any specific regression will tend to understate the true extent of uncertainty about the parameters. To examine the robustness of the partial correlation, this section uses Bayesian model averaging, as in Brock, Durlauf, and West (2003), Durlauf, Kourtellos, and Tan (2008), Fernandez, Ley, and Steel (2001), Malik and Temple (2009), Raftery, Madigan, and Hoeting (1997), and Sala-i-Martin, Doppelhofer, and Miller (2004).[16]

The main ideas are discussed only briefly, drawing heavily on the original presentation in Raftery (1995). Bayesian approaches treat parameters as random variables and aim to summarize uncertainty about them using a probability distribution. The natural extension to model uncertainty is to regard the identity of the true model as unknown and to summarize uncertainty about the data-generating process using a probability distribution over the model space. By explicitly treating the identity of the true model as inherently unknowable but assigning probabilities to different models, it is possible to summarize the global uncertainty about parameters in a way that incorporates model uncertainty.

15. The difference in signs for *RMACRO* between the two regimes does not drive the evidence for the existence of a threshold. If the exercise is repeated and *RMACRO* is removed from the models, the *p*-value for a threshold based on *RMACRO* is generally similar to the results in table 6 except for *GOVKKM*, and even there the null of no threshold is still rejected at the 12 percent level.

16. More recently, Crespo Cuaresma and Doppelhofer (2007) and Eicher, Papageorgiou, and Roehn (2007) have developed approaches that allow joint consideration of model uncertainty and sample splits or thresholds. The application of these to macroeconomic stability would be an interesting area for further work, although in samples of the present size, it would be important to allow for outliers.

Consider the case of K possible models, assuming throughout that one of these models generated the observed data D. The models will be denoted by $M_1 \ldots M_K$ and their corresponding parameter vectors by θ_k. The Bayesian approach to model uncertainty is to assign a prior probability to each model, $p(M_k)$, as well as a prior probability distribution, $p(\theta_k|M_k)$, to the parameters of each model. Using this structure a Bayesian approach can then carry out inference on a quantity of interest, such as a slope parameter, by using the full posterior distribution. In the presence of model uncertainty, this distribution is a weighted average of the posterior distributions under all possible models, where the mixing weights are the posterior probabilities that a given model generated the data (see, for example, Leamer 1978).

To illustrate in the case of just two possible models, the full posterior distribution of a parameter of interest Δ can be written as

$$p(\Delta \mid D) = p(\Delta \mid D, M_1)p(M_1 \mid D) + p(\Delta \mid D, M_2)p(M_2 \mid D), \tag{6}$$

where terms in $p(\Delta \mid D,\ M_k)$ are the conventional posterior distributions obtained under a given model and terms in $p(M_k \mid D)$ are the posterior model probabilities—the probability, given a prior and conditional on having observed D, that model M_k generated the data. This approach requires the evaluation of posterior model probabilities. Briefly, as in Raftery (1995), Raftery, Madigan, and Hoeting (1997), and Sala-i-Martin, Doppelhofer, and Miller (2004), the Bayesian information criterion can be used to approximate the Bayes factors that are needed to compute the posterior model probabilities. This allows a systematic form of model selection and inference to be conducted in a way that acknowledges model uncertainty. For example, to investigate the hypothesis that a slope coefficient β_z is nonzero, the posterior model probabilities are summed for all models in which $\beta_z \neq 0$; this quantity is called a posterior inclusion probability.

As the list of candidate predictors grows, there quickly comes a point where estimation of all the possible models is not feasible, and attention must be restricted to a subset. The approach then follows Raftery, Madigan, and Hoeting (1997) in using a branch-and-bounds search algorithm to identify a subset of models with high posterior probability. For discussion and references, see Malik and Temple (2009) and Sirimaneetham and Temple (2006).

The analysis also draws on the more complex approach of Hoeting, Raftery, and Madigan (1996) because outliers could be a serious problem. In general, any procedure for dealing with model uncertainty or model selection may be influenced by outliers. Even if steps are taken to identify these observations, the final results can easily depend on the order in which model selection and outlier detection is carried out. Hoeting, Raftery, and Madigan suggest a procedure for addressing this issue. First, the full model, containing all the candidate predictors, is estimated by an outlier-robust method due to Rousseeuw (1984), and the standardized residuals are used to identify possible outliers.

Next, model averaging is carried out. As in Hoeting, Raftery, and Madigan, a model is now defined in terms of a set of predictors and a set of observations identified as outliers, where the set of observations identified as outliers include some or all of those identified in the initial stage. (This restriction on the number of candidate outliers is needed to keep the dimensionality of the problem manageable.) Then a Markov chain Monte Carlo model composition approach is used to approximate the posterior model probabilities.

Here, the question of interest is whether *RMACRO* is a robust determinant of growth. The list of candidate predictors will be taken from Sala-i-Martin, Doppelhofer, and Miller (2004), who seek to explain differences in growth rates over 1960–96 for 88 countries (developing and developed). This article instead measures growth over 1970–99 and replaces the Sala-i-Martin, Doppelhofer, and Miller measure of initial GDP for 1960 with a measure for 1970. Despite this change in time period, the same candidate predictors can be used, since the majority of the Sala-i-Martin, Doppelhofer, and Miller explanatory variables were chosen precisely because they are fixed over time or likely to change only slowly. In practice, to keep the application of Bayesian model averaging methods manageable, the analysis that follows focuses on the 31 variables in Sala-i-Martin, Doppelhofer, and Miller (2004, table 2) that have a posterior inclusion probability greater than 4 percent. One of these variables is Dollar's (1992) original index of real exchange rate distortions, measured for 1976–85. This has a low posterior inclusion probability, just 8.2 percent, in the main results of Sala-i-Martin, Doppelhofer, and Miller.

With this set of control variables, the effects of stability can be analyzed at the same time as a wide range of other hypotheses. For example, the Sala-i-Martin, Doppelhofer, and Miller (2004) variables include several measures related to geographic characteristics, including the share of land area in the tropics, the share of population in the tropics, population density, population density in coastal areas in the 1960s, and the prevalence of malaria in the 1960s. Other variables that are included in the Sala-i-Martin, Doppelhofer, and Miller data include regional dummy variables, the relative price of investment goods, life expectancy in 1960, indicators of religion, ethnic diversity, the relative importance of primary exports, and the share of public investment in GDP. The Sala-i-Martin, Doppelhofer, and Miller data span a wide range of the hypotheses investigated in the growth literature, and hence the robustness tests that follow are unusually systematic.[17]

For the purpose of Bayesian model averaging and given the high number of candidate predictors, there are benefits to including as many developing economies in the sample as possible. The measure *RMACRO* is available for

17. One change relative to Sala-i-Martin, Doppelhofer, and Miller is that some explanatory variables are transformed to reduce outlier problems: relative price of investment goods, population density in coastal areas in 1965, and overall population density in 1960, all of which have highly skewed distributions. In some of the analysis, the natural log of these variables is used in place of actual levels.

78 countries but, when merged with the Sala-i-Martin, Doppelhofer, and Miller (2004) data set, the sample is reduced to 63. Country coverage is extended by imputing missing values for a small number of variables in the Sala-i-Martin, Doppelhofer, and Miller data, increasing the number of countries to 72. The decision to impute missing values involves a tradeoff: it introduces measurement error, but it also brings to bear some additional information and lessens the biases that arise when data are missing in nonrandom ways. Here, the number of imputed values in the data matrix for the explanatory variables is just 21, or less than 1 percent of the total number of cells.

The evidence that policy has explanatory power is always much stronger in the 72 country sample than in the 63 country sample, as documented in Sirimaneetham and Temple (2006). The reason for this is clear based on the values of *RMACRO* for the 9 countries that are added to make the 72 country sample. These nine countries include four that are in the bottom decile for *RMACRO* (Guyana, Iran, Nicaragua, and Sierra Leone) and three that are in the top two deciles (Chad, Cyprus, and Fiji). Hence, moving to the larger sample increases the representation of countries at the extreme ends of the distribution of macroeconomic outcomes. This clearly adds identifying variation to the data set. At the same time, considerable faith is needed that policy outcomes and growth are reliably measured for these countries.[18]

The full Bayesian model averaging results are not reported, since the main focus is the posterior inclusion probability associated with *RMACRO*. This is the sum of the posterior model probabilities for all models in which the variable appears. When *RMACRO* is combined with the 31 variables from Sala-i-Martin, Doppelhofer, and Miller (2004), model averaging leads to a posterior inclusion probability of 100 percent, which implies that *RMACRO* appears in every model that is assigned nonzero posterior probability (the Raftery, Madigan, and Hoeting 1997 procedure effectively rounds posterior probabilities down to 0 for the weakest models). The relevant posterior mean—that is, the weighted average of the coefficients on *RMACRO* across all models, where the weights are the posterior model probabilities—is 0.51. This is close to the estimate found in the earlier growth regressions.

When the outlier-robust MC^3 approach of Hoeting, Raftery, and Madigan (1996) is used, the results are weaker, but still supportive. Dollar's (1992) original index of real exchange rate distortions has a high posterior inclusion probability of 99 percent. The evidence for a separate effect of *RMACRO* is weak

18. This is related to a more general debate about the appropriate response to "good" and "bad" leverage points, those observations with extreme values for the explanatory variables; see Dehon, Gassner, and Verardi (2009) and Temple (2000), for example. Here using the 72 country sample comes with the caution that it contains a number of leverage points, affecting inference and the posterior model probabilities.

but becomes much stronger when Dollar's index is excluded. The inclusion probability for *RMACRO* then rises to 69 percent.

Does macroeconomic stability matter, even conditional on institutions? This can be investigated by adding measures of institutional quality to the Bayesian model averaging exercises.[19] These are the same measures used earlier, namely *GOVKKM, POLITY, POLCON*, and *EXPRISK*. Initially, *EXPRISK* is excluded because it reduces the sample of countries substantially. When the other three measures are added to the previous Bayesian model averaging, the posterior inclusion probability of *RMACRO* is 97.4 percent. With the outlier-robust Markov chain Monte Carlo model composition approach, the inclusion probability of *RMACRO* falls to 53 percent. Incidentally, the results strongly support the hypothesis that growth and institutions are highly correlated. The measure *GOVKKM* dominates the others, with an inclusion probability of 100 percent. The inclusion probabilities for the extent of democracy (*POLITY*) and political constraints (*POLCON*) never exceed 35 percent.

When the expropriation risk measure is included, the sample is reduced to 56 countries. The posterior inclusion probability of *RMACRO* is high in this sample (96.8 percent), while *GOVKKM* continues to outperform the other measures of institutional quality. The *POLITY* and *POLCON* measures have inclusion probabilities in the 40–50 percent range, while expropriation risk adds little in terms of explanatory power, with an inclusion probability of just 0.1 percent.

To summarize, when considering a wide range of candidate growth predictors, the evidence that *RMACRO* matters is sensitive to the inclusion of leverage points. This explains why the results are much stronger for the larger sample based on imputed data. In that sample, there is always a high inclusion probability for either *RMACRO* or Dollar's (1992) index of real exchange rate distortions. Expressed differently, nearly all the best-performing models include at least one of these variables, regardless of how the rest of the specification varies. There is also some evidence that stability matters, even when controlling for institutional quality. This is a demanding test, given that some of these institutional measures are likely to reflect a wide range of outcomes, rather than simply rules and constraints.

V. Counterfactual Distributions

This section examines the role of macroeconomic stability in broader perspective by constructing counterfactual distributions for growth rates and

19. To keep the number of candidate predictors within feasible limits, some of the original Sala-i-Martin, Doppelhofer, and Miller (2004) variables have to be dropped. Those excluded will be the variables with relatively low posterior inclusion probabilities in the main results of Sala-i-Martin, Doppelhofer, and Miller.

steady-state levels of income. These distributions indicate what might have happened had all countries achieved the same level of macroeconomic stability over 1970–99. Regression estimates are used to construct the relevant counterfactuals[20] and to reveal where in the distribution the role of stability may have been especially important, information that is not directly apparent from regression estimates.

In counterfactual distributions, the effects—in terms of changes in the location and shape of the distribution—are rarely uniform throughout the distribution. For example, the changes in growth rates reflected in the shape of the counterfactual distribution depend on the full joint distribution of macroeconomic stability and the growth rate. If all countries with intermediate or better growth rates also had stable outcomes, but countries with low growth did not, then imposing macroeconomic stability throughout the sample would affect only the lower end of the distribution. Changes in the growth rate distribution cannot be summarized simply by a set of regression coefficients, and looking at the whole distribution can add useful information.

The first exercise considers actual and counterfactual distributions of growth rates. The basic goal is to determine what each country's growth rate would have been had all countries achieved the same level of macroeconomic stability over 1970–99. This starts from a growth regression similar to the regressions in section III relating growth to the Solow variables, regional dummy variables, and *RMACRO*. The coefficient on *RMACRO* in this regression is 0.64. The counterfactual growth rate g_i^* is then equal to

$$g_i^* = g_i + 0.64(M^* - RMACRO_i), \tag{7}$$

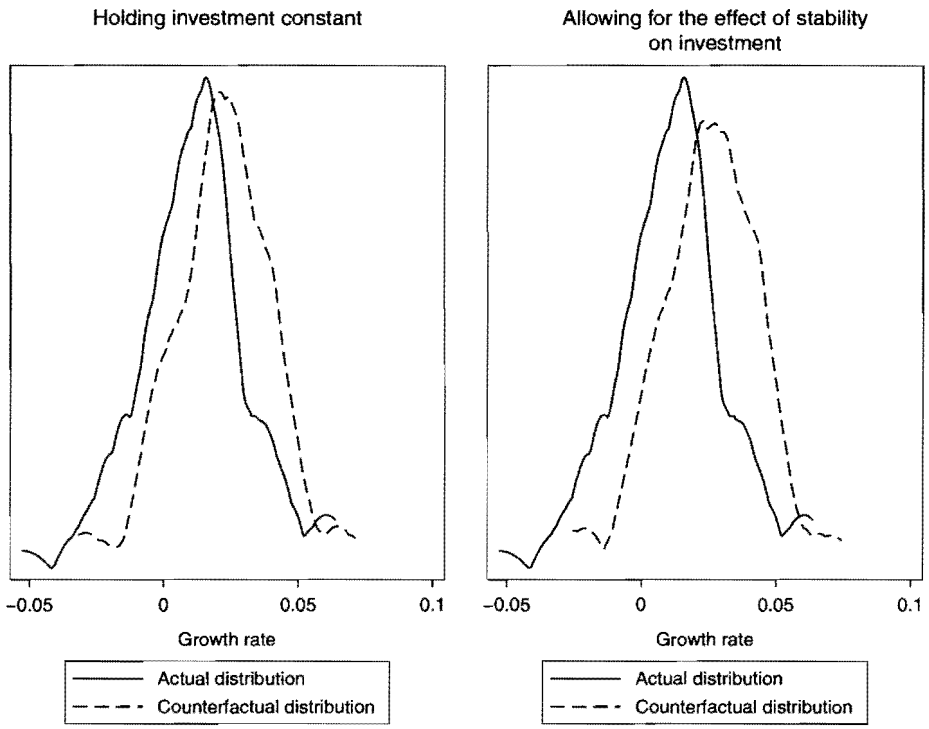
where g_i is the observed growth rate and M^* is the value of *RMACRO* at the 95th percentile in the sample, corresponding to the value for Malaysia.

The distribution of growth rates would have shifted to the right had macroeconomic stability been more widely achieved (figure 3, left panel). But this exercise holds the rate of investment constant and may therefore understate the benefits of macroeconomic stability. This can easily be examined by excluding investment from the growth regression used to construct the counterfactual distribution. The relevant counterfactual distribution now lies slightly further to the right (figure 3, right panel). The benefits of stability continue to be observed throughout the distribution.

The growth regressions include initial income and thus can be seen as modeling the level of the steady-state growth path, as in Mankiw, Romer and Weil (1992). Under the assumption that all countries grow at the same rate in the long-run steady state, the estimated coefficients for 1970–99 can be used to

20. Kernel density estimates of counterfactual distributions are associated in particular with the work of DiNardo, Fortin, and Lemieux (1996) on wage distributions. These methods have also been applied in growth economics by Beaudry and Collard (2006), Beaudry, Collard, and Green (2005), and Desdoigts (1996, 2004).

FIGURE 3. Actual and Counterfactual Distributions of Growth Rates



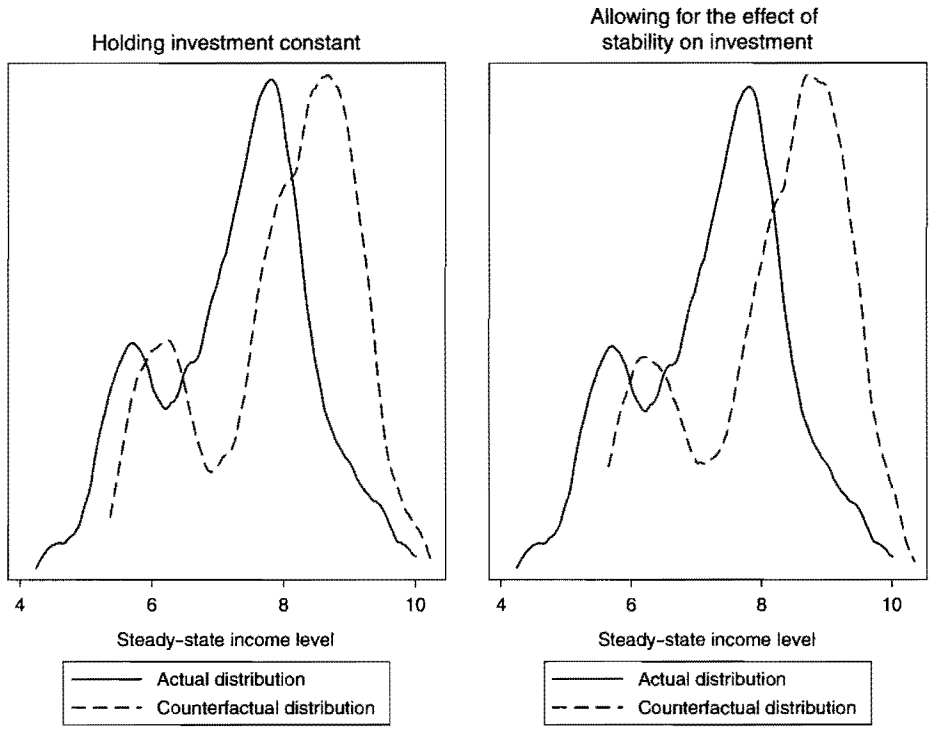
Source: Authors' analysis based on data listed in table 1.

compute the implied steady-state distribution of GDP per capita. Similarly, it is possible to construct a counterfactual distribution that would have been obtained under universal macroeconomic stability.

The actual distributions of the log of GDP per capita are not necessarily expected to have the familiar "twin peaks" pattern of Quah (1996) because the sample is restricted to developing economies. Better macroeconomic outcomes might have moved the distribution of steady-state income levels to the right, and the potential magnitude of this effect is clearly substantial (figure 4, left panel). The analysis is extended by taking into account the effect of *RMACRO* on investment.[21] The counterfactual distribution is slightly further to the right than in the left panel of figure 4, as would be expected if macroeconomic stability were associated with higher investment. Overall, the results indicate that

21. This is based on a simple regression of the log of the investment rate on initial income, initial human capital, regional dummy variables, and *RMACRO*, which is then used to calculate a set of (country-specific) counterfactual investment rates that would have obtained had all countries achieved macroeconomic stability. This is then used in the construction of the counterfactual steady-state distribution (figure 4, right panel).

FIGURE 4. Actual and Counterfactual Distributions of Steady-State Log Income



Source: Authors' analysis based on data listed in table 1.

macroeconomic stability could be a major influence on the steady-state distribution of income levels.

VI. CONCLUSION

This article examined the relationship between macroeconomic stability and growth in developing economies. It introduced a new index of the extent of macroeconomic stability, having aggregated five policy indicators using an outlier-robust version of principal components analysis. With this index, growth is found to be positively associated with macroeconomic stability in a sample of 70 developing economies. If this is interpreted as a causal effect, a 1 standard deviation improvement in the index would raise annual growth by roughly 0.5–0.7 percentage point over 30 years.

Consistent with previous work on this topic, the strength of the evidence depends on the sample of countries. In the largest sample considered, Bayesian methods indicate that the effect is generally robust across a range of

specifications. But as the discussion has emphasized throughout, the results are best interpreted as an upper bound on the benefits of good macroeconomic management. Unstable policy outcomes may sometimes reflect deeper institutional weaknesses, exposure to external shocks, or political instability and conflict.

One of the main contributions of this article is to close the gap between the vocabulary of policy analysis and the models used by empirical growth researchers. In particular, threshold estimation can be used to identify distinct growth regimes. Formal tests indicate that the stability index can be used to divide the sample into two groups. In the relatively stable group of countries, investment has a strong effect on output, and the standard growth determinants of the Solow model, together with a measure of institutional quality, can explain 75–90 percent of the cross-section variation in growth rates. In the less stable group instability clearly reduces growth, the Solow variables have less explanatory power, investment is less effective, and the residual variance is much higher. The results also suggest that good institutions are not strongly associated with growth unless macroeconomic stability is also in place. These patterns support the commonsense view that some degree of stability is a necessary condition for rapid growth, even when a separate role for institutions is taken into account. Viewed as a whole, the results indicate that the conclusion of some recent research—that macroeconomic stability is largely irrelevant—may be premature.

Appendix

This appendix briefly discusses Dollar's (1992) measure of exchange rate overvaluation, which can be interpreted in a variety of ways. One issue is whether Dollar's procedures can reliably control for the determinants of nontradables prices, which has been analyzed by Falvey and Gemmell (1998, 1999). They find that Dollar's approach can be a reasonable approximation on average, at least when GDP per capita is a good proxy for relative factor endowments.

Assuming for now that Dollar's (1992) procedure is effective in modeling nontradables prices, a remaining question is whether differences in tradables prices reflect trade restrictions or exchange rate policies. Exchange rate policies would be more relevant to this article. Rodriguez and Rodrik (2000) provide an especially useful discussion of the strict assumptions that are needed for Dollar's approach to capture trade restrictions. They argue that international variation in price levels will be driven partly by trade costs, which in turn could reflect geographic characteristics. They show that about half the variation in the original Dollar measure can be explained by a combination of the black market exchange rate premium, regional dummy variables, and two geographic indicators—one measuring the ratio of coastal length to land area and the other a dummy variable for tropical countries. Overall they conclude that the

cross-section variation in price levels is likely to be driven by a combination of nominal exchange rate policies and geographic characteristics rather than by variation in trade barriers.

This provides only partial support for the use of *OVERVALU* to measure macroeconomic policy outcomes. This article assumes that the cross-section variation in *OVERVALU* reflects primarily differences in national exchange rate policies. Given that other interpretations are possible, it is worth examining what happens when *OVERVALU* is omitted from the set of indicators developed in section II. Recalculating the principal components based on four indicators rather than five yields the following index:

$$\text{(A-1)} \qquad \begin{aligned} MACROND = {} & 0.332 * SURPLUS - 0.516 * INFLA \\ & - 0.615 * BMP - 0.495 * ERATE \end{aligned}$$

again in terms of standardized variables. This composite indicator is highly correlated with the preferred measures *MACRO* ($r = 0.97$) and *RMACRO* ($r = 0.98$). Hence, the main results are unlikely to be sensitive to omission of *OVERVALU* from the policy index. This robustness is likely to reflect, at least in part, the high correlation that Rodriguez and Rodrik (2000) note between *OVERVALU* and a variable with a much clearer interpretation, the black market exchange rate premium, *BMP*.

References

Acemoglu, Daron, Simon Johnson, and James Robinson. 2001. "Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91(5):1369–401.

Acemoglu, Daron, Simon Johnson, James Robinson, and Yunyong Thaicharoen. 2003. "Institutional Causes, Macroeconomic Symptoms: Volatility, Crises, and Growth." *Journal of Monetary Economics* 50(1):49–123.

Anderson, T.W., and Herman Rubin. 1949. "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." *Annals of Mathematical Statistics* 20(1):46–63.

Barro, Robert J. 1996. "Inflation and Growth." *Federal Reserve Bank of St. Louis Review* 78(3):153–69.

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

Beaudry, Paul, and Fabrice Collard. 2006. "Globalization, Returns to Accumulation and the World Distribution of Output." *Journal of Monetary Economics* 53(5):879–909.

Beaudry, Paul, Fabrice Collard, and David A. Green. 2005. "Changes in the World Distribution of Output Per Worker, 1960–1998: How a Standard Decomposition Tells an Unorthodox Story." *Review of Economics and Statistics* 87(4):741–53.

Bleaney, Michael. 1996. "Macroeconomic Stability, Investment and Growth in Developing Countries." *Journal of Development Economics* 48(2):461–77.

Brock, William A., Steven N. Durlauf, and Kenneth D. West. 2003. "Policy Evaluation in Uncertain Economic Environments." *Brookings Papers on Economic Activity* 34: 235–322.

Bruno, Michael, and William Easterly. 1998. "Inflation Crises and Long-run Growth." *Journal of Monetary Economics* 41(1):3–26.

Burnside, Craig, and David Dollar. 2000. "Aid, Policies, and Growth." *American Economic Review* 90(4):847–68.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics*. Cambridge, UK: Cambridge University Press.

Cook, Paul, and Yuchiro Uchida. 2003. "Privatisation and Economic Growth in Developing Countries." *Journal of Development Studies* 39(6):121–54.

Crespo Cuaresma, Jesus, and Gernot Doppelhofer. 2007. "Nonlinearities in Cross-Country Growth Regressions: A Bayesian Averaging of Thresholds (BAT) Approach." *Journal of Macroeconomics* 29(3):541–54.

Dehon, Catherine, Marjorie Gassner, and Vincenzo Verardi. 2009. "Beware of 'Good' Outliers and Overoptimistic Conclusions." *Oxford Bulletin of Economics and Statistics* 71(3):437–52.

Desdoigts, Alain. 1996. "Smoothing Techniques Applied to a Key Economic Issue: The 'Convergence' Hypothesis." *Computational Statistics* 11(4):481–94.

———. 2004. "Neoclassical Convergence Versus Technological Catch-Up: A Contribution for Reaching a Consensus." *Problems and Perspectives in Management* 3: 15–42.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semi-Parametric Approach." *Econometrica* 64(5):1001–44.

Dollar, David. 1992. "Outward-Oriented Developing Economies Really Do Grow More Rapidly: Evidence from 95 LDCs, 1976–1985." *Economic Development and Cultural Change* 40(3): 523–44.

Durlauf, Steven N., Paul A. Johnson, and Jonathan R.W. Temple. 2005. "Growth Econometrics." In Philippe Aghion, and Steven N. Durlauf, eds., *Handbook of Economic Growth*. Vol. 1A. Amsterdam: North-Holland.

———. 2009. "The Methods of Growth Econometrics." In Terence C. Mills and Kerry Patterson, eds., *Palgrave Handbook of Econometrics*. Vol. 2: Applied Econometrics. London: Palgrave Macmillan.

Durlauf, Steven N., Andros Kourtellos, and Chih Ming Tan. 2008. "Are Any Growth Theories Robust?" *Economic Journal* 118(527):329–46.

Easterly, William. 2001a. *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics*. Cambridge, Mass.: MIT Press.

———. 2001b. "The Lost Decades: Developing Countries' Stagnation in Spite of Policy Reform 1980–1998." *Journal of Economic Growth* 6(2):135–57.

———. 2005. "National Policies and Economic Growth: A Reappraisal." In Philippe Aghion and Steven N. Durlauf, eds., *Handbook of Economic Growth*. Vol. 1A. Amsterdam: North-Holland.

Easterly, William, and Mirvat Sewadeh. 2002. *Global Development Network Growth Database*. World Bank, Washington, DC.

Easterly, William, and Ross Levine. 2003. "Tropics, Germs, and Crops: How Endowments Influence Economic Development." *Journal of Monetary Economics* 50(1):3–39.

Easterly, William, Michael Kremer, Lant Pritchett, and Lawrence Summers. 1993. "Good Policy or Good Luck? Country Growth Performance and Temporary Shocks." *Journal of Monetary Economics* 32(3):459–83.

Easterly, William, Ross Levine, and David Roodman. 2004. "Aid, Policies, and Growth: Comment." *American Economic Review* 94(3):774–80.

Eicher, Theo S., Chris Papageorgiou, and Oliver Roehn. 2007. "Unraveling the Fortunes of the Fortunate: An Iterative Bayesian Model Averaging (IBMA) Approach." *Journal of Macroeconomics* 29(3):494–514.

Falvey, Rod, and Norman Gemmell. 1998. "Why Are Prices So Low in Asia?" *World Economy* 21(7):897–911.

———. 1999. "Factor Endowments, Nontradables Prices and Measures of 'Openness'." *Journal of Development Economics* 58(1):101–22.

Fernandez, Carmen, Eduardo Ley, and Mark F.J. Steel. 2001. "Model Uncertainty in Cross-Country Growth Regressions." *Journal of Applied Econometrics* 16(5):563–76.

Fischer, Stanley. 1991. "Growth, Macroeconomics, and Development." In Olivier Jean Blanchard, and Stanley Fischer, eds, *NBER Macroeconomics Annual 1991*. Cambridge, Mass.: MIT Press.

———. 1993. "The Role of Macroeconomic Factors in Growth." *Journal of Monetary Economics* 32(3):485–512.

Frankel, Jeffrey A., and David Romer. 1999. "Does Trade Cause Growth?" *American Economic Review* 89(3):379–99.

Glaeser, Edward L., Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2004. "Do Institutions Cause Growth?" *Journal of Economic Growth* 9(3):271–303.

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

Hansen, Bruce E. 1996. "Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis." *Econometrica* 64(2):413–30.

———. 2000. "Sample Splitting and Threshold Estimation." *Econometrica* 68(3):575–603.

Hausmann, Ricardo, Lant Pritchett, and Dani Rodrik. 2005. "Growth Accelerations." *Journal of Economic Growth* 10(4):303–29.

Hausmann, Ricardo, Dani Rodrik, and Andrés Velasco. 2008. "Growth Diagnostics." In Narcís Serra and Joseph E. Stiglitz, eds., *The Washington Consensus Reconsidered: Towards a New Global Governance*. Oxford: Oxford University Press.

Henisz, Witold. 2000. "The Institutional Environment for Economic Growth." *Economics and Politics* 12(1):1–31.

Henry, Peter Blair, and Conrad Miller. 2009. "Institutions Versus Policies: A Tale of Two Islands." *American Economic Review* 99(2):261–67.

Heston, Alan, Robert Summers, and Bettina Aten. 2002. Penn World Table Version 6.1. University of Pennsylvania, Center for International Comparisons, Philadelphia, Penn.

Hoeting, Jennifer, Adrian E. Raftery, and David Madigan. 1996. "A Method for Simultaneous Variable Selection and Outlier Identification in Linear Regression." *Computational Statistics and Data Analysis* 22(3):251–70.

Hubert, Mia, Peter J. Rousseeuw, and Karlien Vanden Branden. 2005. "ROBPCA: A New Approach to Robust Principal Component Analysis." *Technometrics* 47(1):64–79.

Kaufmann, Daniel, Massimo Mastruzzi, and Diego Zavaleta. 2003. "Sustained Macroeconomic Reforms, Tepid Growth: A Governance Puzzle in Bolivia?" In Dani Rodrik, ed., *In Search of Prosperity*. Princeton, NJ: Princeton University Press.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2005. "Governance Matters IV: Governance Indicators for 1996–2004." World Bank, Washington, DC.

La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2008. "The Economic Consequences of Legal Origins." *Journal of Economic Literature* 46(2):285–332.

Leamer, Edward E. 1978. *Specification Searches: Ad-Hoc Inference with Non-Experimental Data*. New York: John Wiley.

Levine, Ross, and David Renelt. 1992. "A Sensitivity Analysis of Cross-Country Growth Regressions." *American Economic Review* 82(4):942–63.

Malik, Adeel, and Jonathan R.W. Temple. 2009. "The Geography of Output Volatility." *Journal of Development Economics* 90(2):163–178.

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

Marshall, Monty, and Keith Jaggers. 2002. "Polity IV Project: Political Regime Characteristics and Transitions, 1800–2002." University of Maryland, College Park, Center for International Development and Conflict Management.

Minier, Jenny. 2007. "Institutions and Parameter Heterogeneity." *Journal of Macroeconomics* 29(3):595–611.

Montiel, Peter, and Luis Servén. 2006. "Macroeconomic Stability in Developing Countries: How Much Is Enough?" *World Bank Research Observer* 21(2):151–78.

Moreira, Marcelo J. 2003. "A Conditional Likelihood Ratio Test for Structural Models." *Econometrica* 71(4):1027–48.

Papageorgiou, Chris. 2002. "Trade as a Threshold Variable for Multiple Regimes." *Economics Letters* 77(1):85–91.

Pritchett, Lant. 2000. "Understanding Patterns of Economic Growth: Searching for Hills among Plateaus, Mountains, and Plains." *World Bank Economic Review* 14(2):221–50.

Quah, Danny T. 1996. "Twin Peaks: Growth and Convergence in Models of Distribution Dynamics." *Economic Journal* 106(437):1045–55.

Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research." In Peter V. Marsden, ed., *Sociological Methodology*. Cambridge, UK: Blackwells.

Raftery, Adrian E., David Madigan, and Jennifer E. Hoeting. 1997. "Bayesian Model Averaging for Linear Regression Models." *Journal of the American Statistical Association* 92(437):179–91.

Rodriguez, Francisco, and Dani Rodrik. 2000. "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence." In Ben S. Bernanke and Kenneth Rogoff, eds., *NBER Macroeconomics Annual 2000*. Cambridge, Mass.: MIT Press.

Rodrik, Dani 1999. "Where Did All the Growth Go? External Shocks, Social Conflict, and Growth Collapses." *Journal of Economic Growth* 4(4):385–412.

———. 2005. '*Why We Learn Nothing from Regressing Economic Growth on Policies*.' Cambridge, Mass: Harvard University.

———. 2007. *One Economics, Many Recipes*. Princeton, NJ: Princeton University Press.

Rousseeuw, Peter J. 1984. "Least Median of Squares Regression." *Journal of the American Statistical Association* 79(388):871–80.

Rousseeuw, Peter J., and Katrien van Driessen. 1999. "A Fast Algorithm for the Minimum Covariance Determinant Estimator." *Technometrics* 41(3):212–23.

Sachs, Jeffrey, and Andrew S. Warner. 1995. "Economic Reform and the Process of Global Integration." *Brookings Papers on Economic Activity* 1: 1–118.

Sala-i-Martin, Xavier. 1991. "Comments on Fischer (1991), 'Growth, Macroeconomics, and Development.'" In Olivier Jean Blanchard, and Stanely Fischer, eds., *NBER Macroeconomics Annual 1991*. Cambridge, Mass.: MIT Press.

Sala-i-Martin, Xavier, Gernot Doppelhofer, and Ronald I. Miller. 2004. "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach." *American Economic Review* 94(4):813–35.

Sirimaneetham, Vatcharin, and Jonathan R.W. Temple. 2006. "Macroeconomic Policy and the Distribution of Growth Rates." Discussion Paper 06/584. University of Bristol, UK.

Solow, R.M. 2001. "Applying Growth Theory across Countries." *World Bank Economic Review* 15(2):283–88.

Temple, Jonathan R.W. 1998. "Equipment Investment and the Solow Model." *Oxford Economic Papers* 50(1):39–62.

———. 2000. "Growth Regressions and What the Textbooks Don't Tell You." *Bulletin of Economic Research* 52(3):181–205.

———. 2009. "Review of *One Economics, Many Recipes: Globalization, Institutions and Economic Growth* by Dani Rodrik." *Economic Journal* 119(535):F224–F230.

Williamson, John. 1990. "What Washington Means by Policy Reform." In John Williamson, ed., *Latin American Adjustment: How Much Has Happened?* Washington, DC: Institute for International Economics.

World Bank. 2004. *World Development Indicators 2004*. CD-ROM. Washington, DC: World Bank.

The Effect of Male Migration on Employment Patterns of Women in Nepal

Michael Lokshin and Elena Glinskaya

What is the impact of male migration on the labor market behavior of women in Nepal? The instrumental variable full information maximum likelihood method is applied to data from the 2004 Nepal Household Survey to account for unobserved factors that could simultaneously affect men's decision to migrate and women's decision to participate in the labor market. The results indicate that male migration has a negative impact on the level of the labor market participation by women in the migrant-sending household. There is evidence of substantial heterogeneity (based on both observable and unobservable characteristics) in the impact of male migration. The findings highlight the important gender dimension of the impact of predominantly male migration on the well-being of sending households. Strategies for economic development in Nepal should take into account such gender aspects of the migration dynamics. JEL codes: O15, J21

A sharp increase in migration worldwide has fueled debate on the costs and benefits of international migration for sending communities (UNDP 2002). Remittances are considered a key means through which migration affects economic growth. Most microeconomic studies of migration and remittances focus on their role in reducing poverty and economic inequality. The impact of migration on the economic behavior of nonmigrating household members receives relatively little attention (Kanaiaupuni 2000).

Most research on the issue is sociological and demographic and finds that women spend more time working on home farms at least in part because of male migration (Crummet 1987; Deere and Leon de Leal 1987). Among the few economic studies of the labor market outcomes of members of households sending migrants, Funkhouser (1992) examines the effects of migration and remittances on the female labor market participation in Nicaragua. Itzigsohn

Michael Lokshin (corresponding author) is a senior economist in the Development Economics Research Group at the World Bank; his email address is mlokshin@worldbank.org. Elena Glinskaya is a senior economist in the Human Development Sector Unit of the Europe and Central Asia Region at the World Bank; her email address is eglinskaya@worldbank.org. The authors thank David McKenzie, Thomas Mroz, Martin Ravallion, and anonymous referees for constructive comments.

THE WORLD BANK ECONOMIC REVIEW, VOL. 23, NO. 3, pp. 481–507 doi:10.1093/wber/lhp011
Advance Access Publication November 18, 2009

(1995) assesses the effect of migrant remittances on the income and the labor market participation of members of low-income urban households in the Caribbean Basin. Rodriguez and Tiongson (2001) analyze the effect of migrants on the labor force participation of nonmigrants in the Philippines. Sadiqi and Ennaji (2004) study the impact of male migration from Morocco to Europe on the women left behind. Amuedo-Dorantes and Pozo (2006) and Hanson (2007) investigate how migration and migrant remittances affect the employment status and hours of work of others in the sending households in Mexico. Acosta (2006) looks at the relationships among remittances, labor supply, and school attendance in El Salvador. Cabegin (2006) and Yang (2008) examine the effect of overseas work-related migration on the market participation and labor supply behavior of spouses left behind in the Philippines. Kim (2007) studies the impact of remittances on labor supply in Jamaica. Görlich, Toman, and Trebesch (2007) consider the impact of migration on time allocation in migrant households in Moldova. The common finding of all these studies is that migration and remittances result in a decline in the labor force participation of household members left behind, in particular, of women.

This article examines the extent to which male migration affects the labor market participation of prime-age women in Nepal. This question is of interest for a country where 1 of 10 prime-age men works overseas and where in 2004 migrants sent back remittances valued at 17 percent of the GDP (World Bank 2005).

Work migration in Nepal, while predominantly a male phenomenon, occurs within a social framework. It affects families, households, and communities; changes the gender division of labor; and increases women's workload. Male migrants are gone for months and sometimes years at a time. When husbands are away, their wives not only continue to rear the children and take care of the usual household chores, but often also fill in for absent husbands on family plots or enterprises. Female heads of agricultural households have a particularly hard time when male labor is not available for tasks such as plowing, a taboo activity for women in certain areas of Nepal (Nandini 1999).

When men migrate, the well-being of sending households becomes increasingly dependent on the women, raising their status and strengthening their position in household decision-making. Women find themselves playing key roles as entrepreneurs in investing remittances or in running bazaar economies based on the sale of remittances in kind (Brown and Connell 1993). At the same time, however, social and traditional family norms and the structure of the Nepali labor market, which provides limited employment opportunities for women, reinforce husbands' objections to wives working away from home. Wives thus find it easier to work at home in order to maintain respectability in the eyes of neighbors and relatives.

This article models the household decisions on whether male household members are sent to migrate for work and then whether female household

members participate in labor market activities. Using data from the 2004 nationally representative survey of Nepali households, the full information maximum likelihood method is used to estimate the effect of male migration on the market participation of the women left behind. The method takes into account unobserved household characteristics that could simultaneously affect migration and decisions on the market participation. The results indicate that male migration has a negative impact on the level of the market participation by the women left behind.

This article contributes to the literature on the effects of migration and remittances in three important ways. First, this analysis is the only attempt known to the authors to estimate the impact of migration on the labor market behavior of household members of sending households in Nepal. Second, and new to this literature, a methodology is applied that controls for endogeneity and selection biases arising in the model. This econometric technique not only estimates the average effect of migration, but also shows for which types of women the effect of male migration matters more. Finally, the results highlight the important gender dimension of the predominantly male worker migration on the well-being of sending households.

The article is organized as follows. Section I describes the data and defines the main constructed variables. Section II presents the descriptive results, and section III discusses the theoretical model and the estimation methodology. The main findings are presented in section IV. Section V presents some policy implications of the findings.

I. Data

The data for this study are from the 2004 Round of the Nepal Living Standard Survey (NLSS-II), a nationally representative survey of households and communities conducted between April 2003 and April 2004 by the Nepal Central Bureau of Statistics, with assistance from the World Bank (Nepal Central Bureau of Statistics 2004). The sample frame used a two-stage method based on the 2001 Census (Nepal Central Bureau of Statistics 2003).[1] The NLSS collects data on the household consumption of a wide range of food and nonfood items; the sociodemographic composition of the household; the labor status, health, and education achievements of household members; and sources of household income, including income in kind and individual wages. Respondents also reported the amounts of any remittances their households received during the month of the survey and identified the age and migration destination of the remittance senders. This information was used to identify households with migrants.

The analysis here used a subsample of 3,528 households with information on 5,426 prime-age women (ages 18–60 years). The analysis focuses on the

1. For a detailed description of the sample frame and survey methodology, see World Bank (2005).

labor market behavior of these women, defining the labor market participation as engaging in wage-earning activities.[2] Data from the First Round of NLSS in 1996 and the Nepal Census of 2001 are used for the descriptive analysis and to construct the lagged indicators at ward and district levels.

Three groups of households could be misclassified under the definition based on the survey data. One group consists of households with migrants who are still in the process of establishing themselves or whose migrants bring rather than send the remittances home. The second group comprises households that do not report remittances because of fear of the tax consequences or for their own personal safety. The third group consists of households that receive remittances from nonhousehold members. Classifying these three groups of households as having no migrants would bias estimates of the impact of migration on household consumption.

To assess the extent of such misclassifications, the proportion of migrants in the total population from the 2001 Nepal Census was compared with the proportion of households with remittances in the NLSS data. The proportion of domestic migrants in the 2001 Census (4.8 percent) is statistically close to the proportion of migrants from households receiving domestic remittances in the NLSS (5 percent). The Census-calculated proportion of households with international migrants (14 percent) is lower than the NLSS proportion of household receiving remittances from abroad (18 percent). The official statistics report about 1 million prime-age men working outside Nepal. The equivalent NLSS figure is about 900,000. These relatively small discrepancies indicate that the bias resulting from misclassified households would most likely also be small.[3]

II. Migration and Female Labor Market Participation in Nepal

Migration has become a major factor in the economic development of Nepal over the last two decades. In 2004, close to 1 million Nepali migrants were working in India, countries of the Arab Gulf, South Asia, Western Europe, and North America. According to official sources, remittances to Nepali households from abroad reached $1 billion, overflowing foreign exchange reserves and affecting the exchange rate and inflation. Remittances coming through unofficial channels could be at least as large.

2. The focus is on female wage employment because an overwhelming majority of adult female respondents in the sample reported being self-employed in subsistence agriculture.

3. Households in which all members migrate together are omitted from the sample. The omission should have a negligible impact on the results. Kollmair and others (2006) show that only a small number of households migrate from Nepal to other countries and settle there. An analysis of the 2001 Nepal Census (Nepal Central Bureau of Statistics 2003) for this study indicates that only 1.78 percent of households changed their district of residence during the 5 years before the Census.

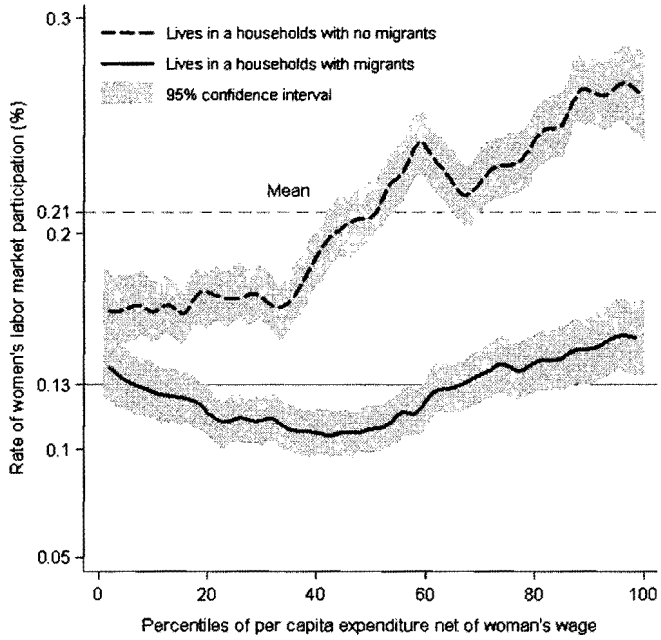
In 2004, 32 percent of households in Nepal had migrants and received remittances (World Bank 2005) averaging about 24,000 Nepal rupees in the year before the survey, or 16 percent of mean household yearly consumption. NLSS data reveal that almost all (97 percent) Nepali migrants are men, ages 15–44, and either the son or husband of the household members receiving the remittances.[4] Brothers make up about 10 percent of remittance donors. The propensity to migrate is higher among members of large households. Less than 2 percent of the households in the sample reported having two or more migrants. Most migrants come from rural areas. Only 13 percent of households in the capital city of Katmandu have migrants; more than twice as many do in rural areas. However, households in Katmandu and other urban areas receive remittances that average twice those received by rural households. The Newar and Janajati castes have the smallest proportions of households with migrants.

On average, 55 percent of men and 19 percent of women engaged in market wage-earning activities in 2004. Respondents, ages 20–35, made up the largest share of workers, with 58 percent of them men and 22 percent women engaged in wage-earning activities. Participation in market work declines with age for both men and women. The formal sector accounts for less than 8 percent of female employment in Nepal. More than 70 percent of female workers are self-employed or employed in low-wage activities in the informal sector. In urban areas, women are employed in a range of cottage industries—such as carpet-weaving, textiles, and handicrafts—and in occupations such as vending, petty trading, brewing, and vegetable selling (UNDP 2004). In rural Nepal, women often work as hired agricultural labor or manual labor in construction and forestry enterprises (Koolwal 2007). Nepalese women lag behind men in education attainment—the gap between male and female literacy rates is about 28 percentage points, and men receive almost twice as many years of schooling as women (World Bank 2005).

The level of market participation by prime-age women varied. On average, only 13 percent of women from households with migrants participated in the labor market, while 21 percent of women from households with no migrants did (figure 1). The gap between these two groups widened for households in the top percentiles of per capita expenditure net of women's market wages. Better-educated women had a higher propensity to work (figure 2). For all education categories except the highest, women from nonmigrant households had higher labor market participation rates. Participation was lowest among women with only 1–7 years of schooling.

4. The Nepal Foreign Employment Act of 1985 placed some restrictions on foreign work migration by women. It limited the overseas travel of single women, as well as women under age 35. The act prohibits the employment of women in foreign countries unless the women have permission from the Nepal government (Sanghera and Kapur 2000).

FIGURE 1. Rates of Women's Labor Market Participation by Percentiles of Per Capita Expenditure (Lowest regression)



Source: Authors' analysis based on data from Nepal Central Bureau of Statistics (2004).

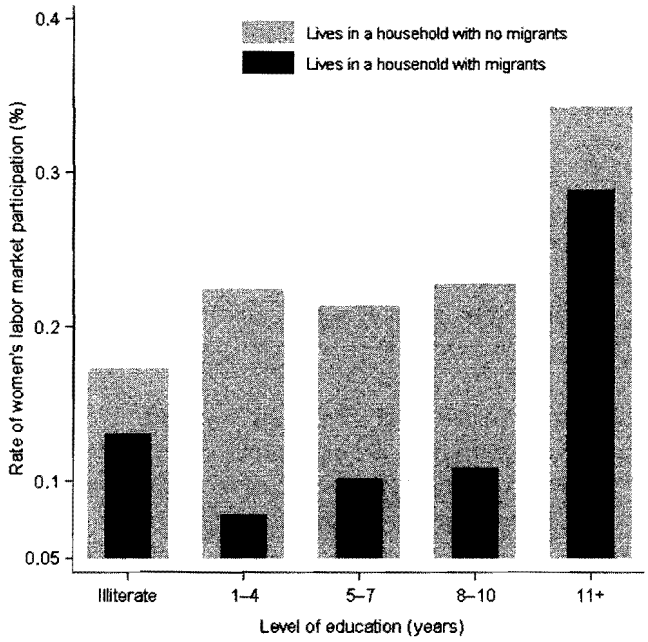
III. THEORETICAL FRAMEWORK AND EMPIRICAL STRATEGY

Before migration takes place, multiple arrangements need to be made. For international migration, for example, migrants have to obtain a passport, apply for a visa, and purchase a ticket. Costs include fees to the migration broker and travel costs, and often there is a contractual agreement between the migrant and the hiring agency (Bhatt and Bhattarai 2006). Thus, once the decision to migrate is made, reversing it can be costly for the household, so the worker usually has to migrate as planned.

Consider a two-period model of utility maximization by a household composed of a husband and wife.[5] Household utility depends on the leisure time of the spouses and the consumption of market goods and goods produced at home (Rosenzweig 1980). Spouses can allocate their time to leisure, market work, and home production. Assume, because of specialization, that the husband is more productive on the labor market and the wife is more productive at home. Assume also that the husband can earn a higher wage by migrating than in the domestic labor market. Under these assumptions, the husband always works on the market (at home or in another country) and the wife divides her time among home production, market work, and leisure.

5. The formal derivation of the theoretical model is available in Lokshin and Glinskaya (2008).

FIGURE 2. Years of Education by Females in Households With and Without Work Migrants



Source: Authors' analysis based on data from Nepal Central Bureau of Statistics (2004).

In period 1, the household compares its utility with and without migration, conditional on expected wages in period 2 (the actual wages in period 2 are unknown in period 1). The household decides that the husband will migrate if expected utility with migration exceeds expected utility without it. In period 2, the household observes the realized labor market outcomes: the migrant, now in the host country, informs the household about his wages, and wage conditions on the local market are known. With this information, the household decides whether the wife will participate in the labor market.

Standard testable hypotheses follow from this theoretical setup. A reduction in the costs of migration and higher expected returns from migration would be expected to increase the probability of the household choosing to send the migrant. The effect of the husband's migration on the wife's labor market behavior is determined by the interaction of income effects and the effect of changes in the wife's productivity at home caused by the migration of her husband. Remittances could be considered a source of household nonwage income. Following the standard assumptions of the theory of labor supply, an increase in nonwage income would raise the reservation wages of nonmigrating members of the household (Rosenzweig 1980). This, in turn, would have a disincentive effect on the wife's labor market participation.

If the inputs of spouses in home production are complements, the husband's migration would lower his wife's productivity at home.[6] In that case, the total effect of migration on the female labor market participation would be ambiguous: some women would enter the labor market and women who worked before their husbands migrated would work longer hours, while other women would spend more time on farm and household activities. If, however, the inputs of the husband and wife are substitutes, which is more likely in Nepal, where a large share of household production is in subsistence agriculture (Kniesner 1976; Leeds and von Allmen 2004), the husband's migration would make the wife's work at home more valuable, so she would reduce her participation in the labor market (Paris and others 2005). Some women would withdraw completely from labor market work. In those cases, lower levels of the market participation would be expected among women in sending households.[7]

The theoretical framework and empirical estimations do not differentiate between internal and international migrations. The impact of both types of migration on the labor market behavior of sending households should be similar and is transferred through two main channels: productive members leave their households, and remittances are transferred within a country or between countries. At the same time, an econometric model that would estimate the three-destination migration decision simultaneously with the market participation decisions of women left behind appears to be computationally infeasible in the full information maximum likelihood framework. Splitting migrant households into two groups would also create a small sample size problem.

Empirical Specification

Let the husband's propensity to migrate be expressed in linearized form as:

$$M_i^* = \gamma \mathbf{Z}_i + \mu_i \tag{1}$$

where subscript i denotes the individual, γ is a vector of parameters; $\mathbf{Z}_i$ is a vector that includes variables on the productive characteristics of a husband and a wife, household characteristics, local labor market characteristics, and the variables determining cost of migration; and μ_i is an error term. Then, the

6. Hiring a perfect substitute for the labor of a husband who migrates is assumed to be very costly to households (Pfeiffer and Taylor 2007).

7. Another channel through which remittances might affect the household labor supply is the removal of liquidity constraints. Remittances might allow liquidity-constrained households to open their own business, which will lead to an increase in household labor supply. There is also a theoretical possibility that access to the labor market differs across households where there are labor market rigidities or household restrictions to off-farm employment for women (such as a religious taboo; see, for example, Rodriguez and Tiongson 2001). Then, the migration decision can be affected by the labor supply of household members.

observed migration status of husband M_i can be expressed as:

$$M_i = 1[M_i^* \geq 0] = 1[\gamma Z_i + \mu_i \geq 0] \tag{2}$$

where 1[.] is an indicator function. The number of hours a wife spends on the labor market could be expressed in a linearized form as:

$$h_{ij} = \beta_j \mathbf{X}_i + \nu_{ij}, \quad j = 0, 1 \tag{3}$$

where β_j is a regime-specific vector of parameters; $\mathbf{X}_i$ is a vector of the individual characteristics of a wife, household characteristics, and locale characteristics; ν_{ij} is the regime-specific error term; and subscript j denotes the regimes (migrate/do not migrate).

Let R_{ij} be the observed labor market status of a wife in period 2, such that:

$$R_{ij} = 1[h_{ij} \geq 0] = 1[\beta_j \mathbf{X}_i + \nu_{ij} \geq 0], \quad j = 0, 1. \tag{4}$$

Error terms (μ_i, ν_{i0}, ν_{i1}) in equations (2) and (4) are assumed to be jointly normally distributed with a zero-mean vector and correlation matrix:

$$\Omega = \begin{pmatrix} 1 & \rho_{\mu 0} & \rho_{\mu 1} \\ & 1 & \rho_{01} \\ & & 1 \end{pmatrix} \tag{5}$$

where the $\rho_{\mu 0,1}$ terms are the correlations between ν_0, ν_1, and μ, and where ρ_{01} is the correlation between ν_0 and ν_1. Since R_{i1} and R_{i0} are never observed simultaneously, the joint distribution of (ν_0, ν_1) is not identified, and consequently ρ_{01} cannot be estimated. Then, the log-likelihood function for the simultaneous system of equations (2) and (4) is:

$$\begin{aligned} \mathrm{Ln}(\Im) = & \sum_{M_i \neq 0, R_i \neq 0} \ln\{\Phi_2(\mathbf{X}_i\beta_1, \mathbf{Z}_i\gamma, \rho_{\mu 1})\} + \sum_{M_i \neq 0, W_i = 0} \ln\{\Phi_2(-\mathbf{X}_i\beta_1, \mathbf{Z}_i\gamma, -\rho_{\mu 1})\} \\ & + \sum_{M_i = 0, R_i \neq 0} \ln\{\Phi_2(\mathbf{X}_i\beta_0, -\mathbf{Z}_i\gamma, -\rho_{\mu 0})\} + \sum_{M_i = 0, R_i = 0} \ln\{\Phi_2(-\mathbf{X}_i\beta_0, -\mathbf{Z}_i\gamma, \rho_{\mu 0})\} \end{aligned} \tag{6}$$

where Φ_2 is the cumulative function of a bivariate normal distribution.

This switching probit model in equation (6) (see, for example, Carrasco 2001; Cappellari 2002) can be used to generate the counterfactual probabilities for women in different regimes of migration and the labor market participation. The impact of migration on women's labor market participation is defined as a treatment effect, following the methodological framework developed by Aakvik, Heckman, and Vytlacil (2000). Then, the effect of migration on a working woman with characteristics x in sending households can be

interpreted as the effect of treatment on the treated (TT):

$$\begin{aligned} \mathrm{TT}(x) &= \Pr[R_1 = 1 | M = 1, \mathbf{X} = x] - \Pr[R_0 = 1 | M = 1, \mathbf{X} = x] \\ &= \frac{\Phi_2[\mathbf{X}\beta_1, \mathbf{Z}\gamma, \rho_{\mu 1}] - \Phi_2[\mathbf{X}\beta_0, \mathbf{Z}\gamma, \rho_{\mu 0}]}{F[\mathbf{Z}\gamma]} \end{aligned} \tag{7}$$

where F is the cumulative function of a univariate normal distribution. The TT is the difference between the predicted probability of the labor market participation for a woman currently residing in a household with a migrant and the probability of the labor market participation for that woman had the household decided not to send a migrant. The average treatment effect on the treated (ATT) is obtained from equation (7) by averaging TT(x) over the sample of women residing in households with migrants:

$$\mathrm{ATT} = \frac{1}{N_{M=1}} \sum_{M=1} \mathrm{TT}(x_i) \tag{8}$$

while the ATT for a subgroup of the population is an average of TT(x) for that subgroup (Heckman and Vytlacil 2000, 2005), for example:

$$\mathrm{ATT(Kathmandu)} = \frac{1}{n_k} \sum_{i=1}^{n_k} \mathrm{TT}(x_i) \tag{9}$$

where n_k is the number of households with a migrant that reside in Katmandu.

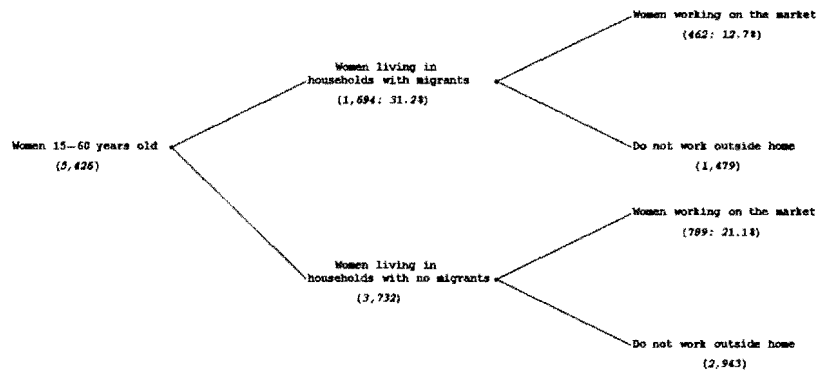
The effect of male migration on the probability of the market participation for a woman randomly drawn from the population of women with characteristics x can be expressed as the treatment effect (TE):

$$\mathrm{TE}(x) = \Pr[R = 1 | \mathbf{X} = x] - \Pr[R = 0 | \mathbf{X} = x] = F[\mathbf{X}\beta_1] - F[\mathbf{X}\beta_0]. \tag{10}$$

Similar to equation (8), the average treatment effect (ATE) is a sample average of TE(x).

The effect of male migration on the female market participation can vary by observed household characteristics $\mathbf{X}$ and unobserved characteristics μ. To account for the unobserved heterogeneity, the marginal treatment effect (MTE) is estimated, using the framework introduced by Bjorklund and Moffitt (1987) and developed by Heckman and Vytlacil (1999, 2000, 2001, 2005). The MTE identifies the effect of male migration on households induced to change the working status of female members because of migration. The MTE can be

FIGURE 3. Sample Selection Diagram (Number of observations and percentage of the sample in groups)



Source: Authors' analysis based on data from Nepal Central Bureau of Statistics (2004).

expressed as:

$$\begin{aligned} \mathrm{MTE}(x,\mu) &= \Pr[R_1 = 1|\mathbf{X} = x, \mu = \bar{\mu}] - \Pr[R_0 = 1|\mathbf{X} = x, \mu = \bar{\mu}] \\ &= F\left[\frac{\mathbf{X}\beta_1 + \rho_{\mu 1}\bar{\mu}}{\sqrt{1-\rho_{\mu 1}^2}}\right] - F\left[\frac{\mathbf{X}\beta_0 + \rho_{\mu 0}\bar{\mu}}{\sqrt{1-\rho_{\mu 0}^2}}\right]. \end{aligned} \tag{11}$$

The schematic diagram of the switching probit model of male migration and the female market participation is shown in figure 3.[8]

The literature on estimating the impact of male migration on the female labor market participation uses three main econometric techniques: instrumental variable regressions, bivariate probit estimators, and matching type estimators. The advantage of the matching estimator is that it requires no assumptions about the distribution of the error terms in equations (1)–(4). The matching estimators assume that the heterogeneity of the effects of migration could be captured by controlling for observable characteristics. The theoretical model indicates that both the husband's migration decision and the wife's labor market participation decision depend on unobservable (by the researcher) characteristics, such as the properties of the home production function.

Both the instrumental variable regression and the bivariate probit estimation rely on normality assumptions, as does the switching probit method used here. However, the switching probit method has several advantages: it relaxes the assumption of equality of coefficients of the labor market participation equations in two regimes and thus is more efficient than either the instrumental variable regression or the bivariate probit estimation. The instrumental variable

8. The likelihood function (equation 6) and the corresponding treatment effect are estimated with Stata command switch_probit (Lokshin and Sajaia forthcoming).

regression, when applied to estimation of the binary choice models with binary endogenous regressors, performs well only in cases of approximately equal probabilities of binary groups; in the analysis here, the proportions of working women and of households with migrants are quite low (Altonji, Elder, and Taber 2005). Finally, the switching probit framework permits defining the effect of male migration on the female labor market participation in terms of impact evaluation and, in particular, enables measuring the ATT and MTE. The instrumental variable estimation can recover only the local treatment effect (Angrist 1991).

Identification Strategy

The system of equations (2) and (4) is identified by nonlinearities even if the variables in **X** and **Z** overlap completely. To make the estimates more robust to alternative functional assumptions, stronger identification restrictions are imposed on the model by including variables that are believed to influence the household's migration decision but not to directly affect the labor market participation decision.

Information from the 2001 Nepal Census (Nepal Central Bureau of Statistics 2003) is used to construct two instrumental variables: the proportion of internal and international migrants in a ward (village) in 2001. In Nepal, labor markets are segmented by gender, and the rates of female employment are only marginally affected by changes in labor market conditions for men (Acharya 2003; Bhatt and Bhattarai 2006). Thus, women's decisions to participate in the labor market in 2004, after controlling for current conditions on the local labor market, should not be directly affected by migration networks formed as early as the 19th century.

Historically, the extent of migrant networks differs by regions in Nepal. In the early 19th century, Gorkhas from the northern hills of Nepal migrated to cities like Lahore in Pakistan, and joined the Indian and British armies. In the 1920s and 1930s, mountain people from the Solu Khumbu area of Nepal migrated to Darjeeling, India, where they were employed as porters. The development of the tea estates in northern India in the early 20th century attracted migrants from southern Nepal (Seddon, Jagannath, and Gurung 2001). Migration from these regions was determined by factors exogenous to local economic conditions.

For this study, the proportion of migrants in a ward in 2001 was used as a proxy for the ward-level networks that help new migrants (see Carringon, Detragiache, and Vishwanath 1996; Munshi 2003). On arrival in the host country, Nepali migrants develop extensive social and information networks that link them with relatives and friends in the home country (Yamanaka 2000). Such networks lower the cost of migration for others in the same wards. Nepali migrants tend to follow their co-villagers and migrate to the same destinations (Thieme 2005). They are also likely to fill a similar niche in the labor market of the country to which they migrate. Woodruff and Zenteno (2007)

and McKenzie and Rapoport (2005) applied the identification strategies that use migrant networks as instruments for migration decisions in studies of migration in Mexico.

The instruments used here could be criticized on the grounds that some lagged ward-level unobserved characteristics can be correlated with the female labor market participation. It would be preferable to use information on older networks, but such data are not available. This criticism is addressed by including in the empirical model a wide set of ward characteristics. Nevertheless, given the limitations of the data, it cannot be established with certainty that the instruments capture no unobserved characteristics of wards that are correlated with labor market outcomes. To some degree, the identification of the system used here comes from functional form assumptions. While the instruments used have these potential weaknesses, they are at least as valid as those used in previous studies. Several diagnostics tests are used to ascertain the validity of the instruments.

Explanatory Variables

The explanatory variables in this model contain the characteristics that determine a woman's market productivity: age (experience) and education-level dummy variables; variables that could affect the home productivity of household members, such as demographic composition and the size of the household's plot; variables describing ethnicity, religion, and household nonwage income;[9] and variables describing regional and ward characteristics, including labor market conditions, distance from the Indian border, and poverty and inequality.

The descriptive statistics for the main explanatory variables are reported in table 1. The characteristics are similar for women living in migrant-sending and nonsending households, with some differences—for example, women in nonsending households are better educated. Migrant-sending households are, on average, smaller, have a higher share of men (adjusted for the number of current migrants), possess larger land plots, and have higher nonwage incomes than do nonsending households. Women working in the labor market are better educated, less likely to be married, and reside in smaller households, with smaller proportions of women, than do women not working in the labor market. Compared with other castes, Brahmin, Chhetri, and Newar households are less likely to have women participating in the labor market. Land ownership has a strong negative effect on the probability of a woman participating in the labor market: women living in landless households are almost three times more likely to work for wages than women in households with less than 1 hectare of land. The gap is

9. Nonwage income is defined as the sum of all government and private transfers, such as, pensions and scholarships, that are exogenous to household migration and labor force participation decisions; it excludes interhousehold transfers, donations, and other private transfers that may respond to the household's migration and labor supply decisions. An alternative approach, following Blundell and Macurdy (1999), would be to include in the empirical specification the household expenditure instrumented with nonwage income.

TABLE 1. Descriptive Statistics for the Main Variables

Variable	Women from household with migrants		Women from household without migrants		Women participating in labor market		Women not participating in labor market	
	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Participate in wage work	0.127	0.008	0.211	0.007				
Live in household with migrants					0.214	0.412	0.335	0.472
Women's characteristics								
Age	34.542	12.825	34.521	11.799	34.045	10.565	34.637	12.453
Married	0.806	0.010	0.812	0.006	0.782	0.013	0.817	0.006
Illiterate	0.614	0.012	0.612	0.008	0.537	0.016	0.630	0.007
1–4 years of schooling	0.100	0.007	0.102	0.005	0.109	0.010	0.100	0.005
5–7 years of schooling	0.099	0.007	0.084	0.005	0.090	0.009	0.089	0.004
8–10 years of schooling	0.152	0.009	0.135	0.006	0.159	0.012	0.136	0.005
11+ years of schooling	0.035	0.004	0.066	0.004	0.106	0.010	0.045	0.003
Household characteristics								
Household size	5.835	2.952	6.267	3.189	5.536	2.537	6.268	3.226
Share of adult men	0.325	0.003	0.277	0.002	0.282	0.005	0.294	0.002
Share of elderly	0.320	0.003	0.337	0.002	0.347	0.005	0.328	0.002
Share of women	0.152	0.003	0.157	0.003	0.137	0.005	0.160	0.002
Share of children ages 0–6	0.165	0.004	0.192	0.003	0.199	0.006	0.180	0.003
Share of children ages 7–15	0.033	0.002	0.036	0.001	0.033	0.003	0.036	0.001
Male-headed household	0.643	0.012	0.903	0.005	0.779	0.013	0.831	0.006
Landless households	0.372	0.012	0.471	0.008	0.694	0.015	0.383	0.007
Own less than 1 hectare	0.377	0.012	0.314	0.008	0.234	0.013	0.357	0.007
Own 1–2 hectares	0.159	0.009	0.141	0.006	0.050	0.007	0.169	0.006
Own more than 2 hectares	0.091	0.007	0.073	0.004	0.022	0.005	0.092	0.004

Household ethnicity and nonwage income								
Brahman/Chhetri	0.355	0.012	0.285	0.007	0.212	0.013	0.329	0.007
Dalits	0.084	0.007	0.068	0.004	0.074	0.008	0.072	0.004
Newar	0.065	0.006	0.150	0.006	0.233	0.013	0.098	0.004
Terai Madhesi Caste	0.255	0.011	0.247	0.007	0.210	0.013	0.259	0.007
Muslim, other	0.241	0.010	0.250	0.007	0.271	0.014	0.242	0.006
Hindu	0.829	0.009	0.817	0.006	0.808	0.012	0.824	0.006
Household nonwage income	0.609	4.585	0.508	3.065	0.305	1.708	0.593	3.912
Regional and ward characteristics								
Katmandu	0.038	0.007	0.144	0.004	0.226	0.005	0.082	0.004
Other urban areas	0.185	0.009	0.187	0.006	0.221	0.013	0.178	0.006
Rural Western Hills	0.239	0.010	0.150	0.006	0.097	0.009	0.196	0.006
Rural Eastern Hills	0.169	0.009	0.193	0.006	0.148	0.011	0.194	0.006
Rural Western Terai	0.117	0.008	0.118	0.005	0.064	0.008	0.130	0.005
Rural Eastern Terai	0.250	0.011	0.208	0.007	0.244	0.014	0.216	0.006
Percentage of migrant population	0.139	0.004	0.091	0.002	0.091	0.004	0.109	0.002
Number of observations	1,694		3,732		1,004		4,422	

Source: Authors' analysis based on data from Nepal Central Bureau of Statistics (2004).

even larger between women in landless households and those in households that own larger land plots. Women living in Katmandu and other urban areas of Nepal are more likely to work for wages than are women in rural areas.

IV. Results

In the joint estimation of equations (2) and (4),[10] the coefficients on the main explanatory variables affecting household migration and woman's labor market participation decisions correspond well with the predictions of the theoretical model (table 2). Households in wards with higher proportions of migrants in 2001 were more likely to send their male members to migrate for work.

Overall, the observed household characteristics, particularly the geographic and ward characteristics, are more important in determining the level of the labor market participation of women in nonmigrant-sending households than of women in migrant-sending households. While a household's human and productive capitals have a strong effect on women's labor market participation in households without migrants, these factors become less important for households that have sent migrants (where remittances contribute a significant share to the household budget).

The level of market participation increases with age for women in both sending and nonsending households. Married women and women with 11 or more years of education are more likely to work for wages. Household nonwage income negatively affects the likelihood of market employment of women from nonsending households. The effect of nonwage income on the market participation of women in sending households is insignificant. Household demographic composition seems to affect the market participation of women in nonsending households. Relative to other ethnic groups, women in Dalit and Muslim households have a higher probability of working. These results are consistent with those of other studies that demonstrate that Hindu women in Indo-Aryan communities that are disposed toward patriarchy are less likely to work for pay than women in primarily Buddhists, Tibeto-Burman, and Muslim communities, which offer women greater social and economic mobility (Raghuram 2001; Koolwal 2007). Women in households with large plots are less likely than those in households with small or no plots to work

10. By the likelihood-ratio (LR) test criterion, the specification that assumes independence of the error terms in equations (2) and (4) (see Lokshin and Glinskaya 2008 for details) is rejected in favor of the full information maximum likelihood estimation.
The Wald tests show that the estimated $\rho_{\mu 0}$ is statistically significant with ($\chi^2(1) = 4.03$), and $\rho_{\mu 1}$ is statistically significant with ($\chi^2(1) = 3.88$); two ρ's are jointly significant.
The LR test on the equality of the coefficients in the equations determining the female market participation in sending and nonsending households rejects the null hypothesis that the effects of the regressors on the female market participation are the same in both regimes ($\chi^2(31) = 54.95$; $\Pr > \chi^2 = 0.0051$).

TABLE 2. Full Information Maximum Likelihood Estimation of the Endogenous Switching Probit Model

	Women's labor market participation decision					
	Households with no migrant		Households with migrant		Migration decision	
Variable	Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error
Women's characteristics						
Age and marital status						
Age	0.063***	0.017	0.113***	0.031	−0.042***	0.013
Age squared/100	−0.092***	0.022	−0.155***	0.041	0.053***	0.017
Married	−0.090	0.080	−0.373***	0.137	0.202***	0.064
Education (reference: illiterate)						
1–4 years of schooling	0.107	0.085	−0.290*	0.172	0.005	0.071
5–7 years of schooling	0.012	0.094	0.088	0.168	0.071	0.075
8–10 years of schooling	0.070	0.085	0.019	0.166	0.149**	0.072
11+ years of schooling	0.321***	0.119	0.830***	0.267	0.026	0.120
Currently in school	−0.437***	0.142	−0.791***	0.306	0.035	0.119
Household characteristics						
Household size	0.007	0.030	0.068	0.084	−0.082***	0.022
Household size squared	−0.000	0.001	−0.004	0.005	0.002**	0.001
Share of adult men	0.120	0.327	−0.601	0.830	−1.041***	0.256
Share of elderly	1.685***	0.426	0.839	1.903	−4.700***	0.244
Share of women	0.487	0.297	−0.100	1.019	−2.503***	0.186
Share of children ages 0–6	1.099***	0.262	0.025	1.010	−2.449***	0.173
Share of children ages 7–15	0.074	0.169	0.081	0.554	−1.468***	0.059
Male-headed household	−0.242***	0.093	−0.440	0.297	0.690***	0.067
Land ownership (reference: landless households)						
Own less than 1 hectare	−0.425***	0.079	−0.684***	0.126	0.021	0.065
Own 1–2 hectares	−0.822***	0.111	−1.137***	0.200	−0.003	0.078
Own more than 2 hectares	−1.101***	0.155	−0.859***	0.219	0.193**	0.094
Household nonwage income and ethnicity income						
Household nonwage income	−0.051***	0.014	−0.016	0.017	0.002	0.005

(Continued)

TABLE 2. Continued

	Women's labor market participation decision					
	Households with no migrant		Households with migrant		Migration decision	
Variable	Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error
Ethnicity (reference: Brahman/Chhetri)						
Dalit	0.229**	0.108	0.398**	0.167	−0.095	0.083
Newar	0.484***	0.088	0.499***	0.186	−0.256***	0.085
Terai Madhesi Caste	0.406***	0.080	0.207	0.140	−0.116**	0.059
Muslim, other	0.281***	0.082	0.361**	0.144	−0.143**	0.062
Hindu	0.020	0.069	0.252*	0.133	−0.109*	0.058
Regional and ward characteristics						
Regional dummy variables (reference: Kathmandu)						
Other urban areas	−0.105	0.127	0.282	0.444	0.777***	0.119
Rural Western Hills	0.002	0.164	0.573	0.544	0.813***	0.147
Rural Eastern Hills	0.399***	0.145	0.928**	0.433	0.593***	0.135
Rural Western Terai	0.123	0.174	0.967*	0.500	0.706***	0.150
Rural Eastern Terai	0.508***	0.171	1.210**	0.553	0.930***	0.141
Ward characteristics						
Percent illiterate	−1.083***	0.183	−0.880***	0.296	0.161	0.136
Percent in wage employment	1.539***	0.310	0.475	0.522	0.204	0.243
Percent self-employed	0.506**	0.221	−0.080	0.367	0.465***	0.151
Ward inequality (Gini)	0.537	0.452	−0.904	0.777	−0.769**	0.336
Ward poverty rate	0.049	0.103	0.095	0.191	0.118	0.082
Distance to India	−0.013	0.043	0.151**	0.074	−0.019	0.033
Percent of international migrants					0.863***	0.155
Percent of domestic migrants					0.213	0.238
Constant	−3.425***	1.019	−3.983**	1.737	1.878**	0.803
Number of observations	5,426					
Log-likelihood	−4847.91					

*Significant at the 10 percent level; **significant at the 5 percent level; ***significant at the 1 percent level.

Note: The standard errors are adjusted for clustering on a ward level.

Source: Authors' analysis based on data from Nepal Central Bureau of Statistics (2004).

outside the home, regardless of the migration decisions of male household members, likely because economies of scale in agriculture increase women's productivity when they work on larger plots. Compared with women living in Katmandu, women residing in other urban areas of Nepal and in rural Western Terai have a lower propensity to participate in market work.

Finally, certain local conditions are significantly correlated with levels of women's market participation. Women in nonmigrant households living in wards with a high proportion of illiteracy are significantly less likely to participate in market work than are women in wards with better-educated populations. Higher shares of wage and self-employment in a ward have a positive impact on women's labor market participation in households with no migrants. The effects of local labor market conditions on the market participation of women residing in migrant-sending households are insignificant.

Various diagnostic tests were run to determine the validity of the instruments. The Sargan's (1958) test on a linearized form (linear instrumental variable regression) of the system of equations (2)–(4) confirms that the excluded instruments are uncorrelated with the error terms ($\Pr > \chi^2(1) = 0.353$) and correctly excluded from equation (4). The test proposed by Stock and Yogo (2005) was used to investigate the potential of a weak instruments problem. The Cragg–Donald (CD) Wald *F*-statistic was calculated by regressing a woman's market participation on a set of her characteristics, an instrumental variable, and an endogenous dummy variable for having a migrant from the household. The hypothesis of weak instruments was rejected with a CD *F*-statistic of 20.39 and critical values of the Stock–Yogo test of 19.93 for 10 percent size of the Wald test. The Wald test on the joint significance of the excluded instruments of $\chi^2(2) = 28.86$ could be interpreted as further evidence for rejecting the weak instruments hypothesis. Finally, a "naïve" test of the validity of the instruments was conducted by including instruments in the labor market participation decision equations. This estimation, identified through nonlinearity, shows that both instruments are insignificant in the labor market participation decision equation and significant in the migration decision equation.

Simulations

The impact of male migration on women's labor market participation was simulated according to equations (7)–(10). Women living in migrant-sending households had a 5.3 percentage point (bootstrap standard error of 1.7) lower probability of participating in the labor market compared with the counterfactual scenario of women living in nonsending households; this is the ATT. The effect of male work-related migration on the market participation of a woman randomly selected from the population was positive and statistically not different from zero; this is the ATE. By comparison, the raw difference in rates of the market participation was −8.4 (standard error of 1.1; ($\Pr(W_1 \mid$

$M = 1) - \Pr(W_0 \mid M = 0) = -8.4))$, suggesting that controlling for selection appeared to be important in these data.

Next, the results of these simulations are compared with the results from the estimation techniques used in the literature on the effect of migration on the labor market participation of women left behind.[11] For the specification that included the migration dummy variable directly in the market participation equation, the ATT is −4.8 percentage points (standard error of 1.8). The magnitude of these effects is similar to that found by Kim (2007) for Jamaica. The bivariate probit of the migration and market participation equations was estimated replicating the methodology of Görlich, Toman, and Trebesch (2007), which uses the same set of explanatory variables and instruments as the preferred model here. This specification assumes joint normality of error terms in the migration and market participation equations and, compared with the switching probit model, imposes a restriction of the equality of coefficients in the market participation equation (4) for women in households with a male migrant and for those in households without a migrant. The derived ATT indicates that, relative to the counterfactual scenario of no migration, women in households with migrants are 5.2 (standard error of 1.1) percentage points less likely to participate in work outside the home. Finally, a propensity score matching technique similar to that of Esquivel and Huerta-Pineda (2006), which assumes selection on observables only, results in a −7.5 percentage point reduction in the market participation of women in households with migrants.

These results are consistent with those from methods used in previous studies. This could boost confidence in the findings of these other studies. Or the similarity of the results across methods could indicate that the full information maximum likelihood model is as biased as the other approaches and was not able to solve the selection issues that plague the migration literature. But the theoretical arguments in favor of the identification strategy, the empirical tests of the validity of the instruments, and the robustness of the results to different econometric specifications and assumptions increase the confidence in the estimates presented here of the impact of male migration on women's labor market participation.

The heterogeneity of the effect of male migration on the female market participation can also be simulated by observable characteristics, as in equation (10). These simulations are shown in the first column of table 3. The largest negative impact of male migration is on women ages 25–35, whose level of the labor market participation would rise by 6.5 percentage points if male migrants were to stay at home. The dampening effect of male migration on the female market participation increases with a woman's education. The market participation by women with 11 or more years of schooling is 15.3 percentage points lower than it would be in the counterfactual scenario. Male migration

11. The complete set of results for these estimations is available in Lokshin and Glinskaya (2008).

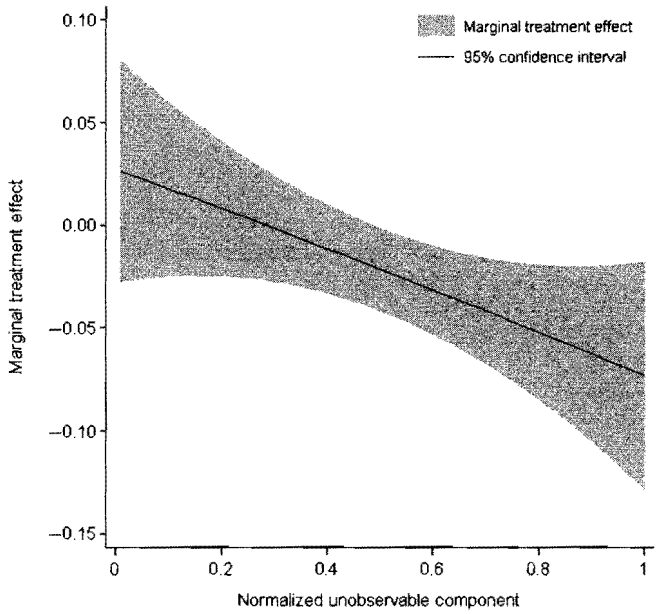
TABLE 3. Simulated Effect of Migration on Women's Labor Market Participation in Migrant-Sending Households (by characteristics of women and households)

	Average treatment effect on the treated (ATT)		Average treatment effect (ATE)	
Variable	Estimate	Standard error	Estimate	Standard error
Age				
18–25	−5.495	1.719	2.127	3.931
25–35	−6.507	2.504	2.354	4.511
35–45	−3.475	1.739	4.178	3.831
45–60	−1.417	1.922	2.502	3.638
Education				
Illiterate	−5.495	1.719	2.127	3.931
1–4 years of schooling	−6.507	2.504	2.354	4.511
5–7 years of schooling	−3.475	1.739	4.178	3.831
8–10 years of schooling	−1.417	1.922	2.502	3.638
11+ years of schooling	−15.313	5.573	9.388	8.760
Landholding				
Landless	−9.236	3.131	4.991	5.799
Own less than 1 hectare	−3.465	1.740	1.239	3.573
Own 1–2 hectares	−2.904	1.204	0.252	2.666
Own more than 2 hectares	−2.302	1.533	3.736	3.125
Ethnicity				
Brahman/Chhetri	−2.499	1.509	2.897	3.476
Dalit	−1.676	2.660	6.894	4.839
Newar	−17.818	4.240	2.606	7.067
Terai Madhesi Caste	−7.270	2.290	−0.722	4.409
Muslim, other	−5.787	2.499	4.620	4.447
Region and ward				
Katmandu	−19.353	5.941	−4.879	9.294
Other urban areas	−9.496	2.886	2.700	5.837
Rural Western Hills	−0.647	1.350	2.582	3.326
Rural Eastern Hills	−6.366	2.070	1.257	4.002
Rural Western Terai	−1.360	2.091	4.864	3.932
Rural Eastern Terai	−4.950	2.493	3.961	4.342
Total	−5.319	1.874	0.061	4.899

Note: The standard errors of the predicted probabilities are calculated by bootstrapping.
Source: Authors' analysis based on data from Nepal Central Bureau of Statistics (2004).

has a greater impact on the work participation of women residing in Katmandu and other urban areas of Nepal and of women living in landless households than it does on that of women living in rural areas or in households with large land holdings. Such differences might be explained by differences in the technology of home production. In households with large plots, women might be able to substitute, to some extent, hired labor for the inputs of men who have migrated, thus lowering the impact of male migration on their productivity at home. The home production of landless households is likely to be

FIGURE 4. Heterogeneity in the Effect of Migration on Women's Labor Market Participation by Unobserved Component (Estimated MTE at population means)



Source: Authors' analysis based on data from Nepal Central Bureau of Statistics (2004).

related to child-rearing and the tending of elderly household members—activities for which finding a paid substitute is difficult.

Heterogeneity in the effect of migration based on unobservable characteristics can be investigated using the MTE framework. Figure 4 plots the MTE against the normalized values of unobservables (μ) at the population means for **X**'s according to equation (11). The estimate of MTE is monotonically decreasing in μ, indicating that households that are more likely to send a male member to migrate for work are also more likely to withdraw their female members from the labor market. The fact that the MTE is not flat confirms the presence of unobservable heterogeneity in the impact of migration on women's labor market participation.[12]

The estimated correlations of error terms in equations (2) and (4) demonstrate the perverse selection on unobservable characteristics: for households sending migrants, unobservable characteristics that positively affect the probability of sending a migrant for work have a negative impact on the probability of a woman's participating in the labor market ($\text{Corr}(\mu, \nu_1) = -0.290$). At the same time, for households with no migrants, the unobservable characteristics

12. The structure of the MTE is determined, to a large extent, by the normality assumptions imposed on the error structure of the empirical model.

promoting migration are positively correlated with women's employment ($\text{Corr}(\mu, \nu_0) = -0.256$). Thus, higher values of μ are correlated with lower value of ν_1 and with higher values of ν_0, so that the impact of male migration on women's labor force participation is lower for households with high μ (who are more likely to have a working migrant).

Qualifications

There are several qualifications concerning the results of this study. First, the results were obtained using cross-sectional data of the year 2004. Without panel data, there are no instruments to control for possible household- or community-level endogeneity. In this sense, the estimations of the impact of work-related migration are valid only to the extent that the variables included in the empirical specification capture unobserved family and community characteristics. Second, the effect of male migration might differ with the relationship of the migrant (husbands, fathers, brothers, other relative) to the women of the sending households. The analysis fails to capture this heterogeneity. Finally, the analysis looks only at the direct impact of male migration on the labor market behavior of women in sending households. Male migration for work could also affect aggregate labor market conditions in the sending communities. Accounting for the general equilibrium consequences of work-related migration might reduce its estimated impact on the labor market participation of women in the household.

V. Conclusion

This article examined the extent to which male migration affects the labor force participation of prime-age women in migrant-sending households in Nepal, using nationally representative household survey data. The theoretical model developed here predicts that male migration could have two main effects on the labor market participation of women. First, the increase in household income from remittances could lead to a reduction in the labor market participation by women. Second, depending on the properties of the home production function, male migration could increase or decrease women's home productivity, thus having an ambiguous effect on their labor market participation. The overall effect of male migration on women's labor market participation therefore depends on the interaction of these factors. The article compared the observed rates of the labor market participation of women in households that had sent migrants with simulated rates under a counterfactual scenario of no migration. To construct these counterfactuals, a model of household male migration and the female labor market participation decisions was estimated that identified observed and unobserved differences in the returns to characteristics based on migration status.

The migration of male household members was found to reduce women's rates of labor market participation by 5.3 percentage points. The effect was

strongest for women ages 25–35 and for women with 11 or more years of education. The income effect of remittances from migrants and the substitutability of male and female time inputs in home production might explain the stronger impact of male migration on women residing in landless households and in urban areas of Nepal. The effect of male migration on the labor market participation of women living in households with large landholdings is weaker, suggesting that men and women in these households complement each other in home production. There is evidence of substantial heterogeneity (based on both observable and unobservable characteristics) in the impact of male migration.

Neoclassical micro theory sees the differentials in wages and employment opportunities between sending and host countries as major driving forces of migration. The inflow of migrants increases the supply of labor in receiving countries and could tighten labor supply in sending countries, thus lowering wage differentials. The withdrawal of women from market work because of male migration might accelerate this process of wage equalization. If particular types of jobs are held by women, the decrease in labor supply to those jobs could drive up the wages of people who still hold these jobs.

The policy implications of these results depend to a large extent on what women in migrant households are doing instead of working. If they are taking on farming tasks previously borne by their husband, that could imply a need to improve the wage labor market in rural areas to allow households to hire workers to replace those who migrate. Detailed information on time use is not available in the Nepal survey data and is rarely available elsewhere. Nepal and other countries should collect more information on both migration and time use to better understand the impact of male migration.

Migration is already high in Nepal and will likely continue to rise in response to the economic incentives offered by neighboring countries. The findings here highlight the gender dimension of the impact of predominantly male migration on the well-being of sending households. The effect of male migration on the work patterns of nonmigrating women has important implications for women's social status and could influence outcomes for other household members, particularly children. Thus, strategies for economic development in Nepal should take into account such gender aspects of migration dynamics.

References

Aakvik, A., J. Heckman, and E. Vytlacil. 2000. *Treatment Effect for Discrete Outcomes When Responses to Treatment Vary Among Observationally Identical Persons: An Application to Norwegian Vocational Rehabilitation Programs*. Technical Paper 262. Cambridge, Mass.: National Bureau of Economic Research.

Acharya, M. 2003. *Efforts at Promotion of Women in Nepal*. Kathmandu: Tanka Prasad Acharya Memorial Foundation.

Acosta, P. 2006. "Labor Supply, School Attendance, and Remittances from International Migration: The Case of El Salvador." Policy Research Working Paper 3903. World Bank, Washington, D.C.

Altonji, J., T. Elder, and C. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113(1):151–84.

Amuedo-Dorantes, C., and S. Pozo. 2006. "Migration, Remittances, and Male and Female Employment Patterns." *American Economic Review* 96(2):222–6.

Angrist, J. (1991). *Instrumental Variables Estimation of Average Treatment Effects in Econometrics and Epidemiology.* NBER Technical Working Paper 0115. Cambridge, Mass.: National Bureau for Economic Research.

Bhatt, S., and E. Bhattarai. 2006. "WTO Membership and Nepalese Women." *South Asian Journal* 13(July–September), www.southasianmedia.net/Magazine/Journal/previousissues13.htm.

Bjorklund, A., and R. Moffitt. 1987. "The Estimation of Wage Gains and Welfare Gains in Self-selection Models." *Review of Economics and Statistics* 69(1):42–49.

Blundell, R., and T. Macurdy. 1999. "Labor Supply: A Review of Alternative Approaches." In O. AshenfelterD. Card eds., *Handbook of Labor Economics.* Amsterdam: Elsevier.

Brown, R., and J. Connell. 1993. "The Global Flea Market: Migration, Remittances and the Informal Economy in Tonga." *Development and Change* 24(4):611–47.

Cabegin, E. 2006. *The Effect of Filipino Overseas Migration on the Non-migrant Spouse's Market Participation and Labor Supply Behavior.* Discussion Paper Series 2240. Bonn, Germany: Institute for the Study of Labor.

Cappellari, L. 2002 "Do the 'Working Poor' Stay Poor? An Analysis of Low Pay Transitions in Italy." *Oxford Bulletin of Economics and Statistics* 64(2):87–110.

Carrasco, R. 2001. "Binary Choice with Binary Endogenous Regressors in Panel Data: Estimating the Effect of Fertility on Female Labor Participation." *Journal of Business and Economic Statistics* 19(4):385–94.

Carrington, W., E. Detragiache, and T. Vishwanath. 1996. "Migration with Endogenous Moving Costs." *American Economic Review* 86(4):909–30.

Crummet, M. 1987. "Rural Women and Migration in Latin America." In C. Deere, and M. Leon de Leal eds., *Rural Women and State Policy: Feminist Perspectives on Latin America Agricultural Development.* Bolder, Colo.: Westview Press.

C. Deere, and M. Leon de Leal eds. 1987. *Rural Women and State Policy: Feminist Perspectives on Latin America Agricultural Development.* Bolder, Colo.: Westview Press.

Esquivel, G., and A. Huerta-Pineda. 2006. *Remittances and Poverty in Mexico: A Propensity Score Matching Approach.* Washington, D.C.: Inter-American Development Bank.

Funkhouser, E. 1992. "Migration from Nicaragua: Some Recent Evidence." *World Development* 20(8):1209–18.

Görlich, D., M. Toman, and C. Trebesch. 2007. *Explaining Labour Market Inactivity in Migrant-Sending Families: Housework, Hammock, or Higher Education?* Working Paper 1391. Kiel, Germany: Kiel Institute for the Working Economy.

Hanson, G. 2007. "Emigration, Remittances, and Labor Force Participation in Mexico." *Integration and Trade Journal* 27(July–December):73–103.

Heckman, J., and E. Vytlacil. 1999. "Local Instrumental Variable and Latent Variable Models for Identifying and Bounding Treatment Effects." *Proceedings of the National Academy of Sciences* 96(April):4730–34.

———. 2000. "Local Instrumental Variables." In C. Hsiao, K. Morimune, and J. Powells eds., *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, Cambridge: Cambridge University Press.

———. 2001. "Policy Relevant Treatment Effects." *American Economic Review* 91(2):107–11.

———. 2005. "Structural Equations, Treatment, Effects and Econometric Policy Evaluation." *Econometrica* 73(3):669–738.

Itzigsohn, J. 1995. "Migrant Remittances, Labor Markets, and Household Strategies: A Comparative Analysis of Low-Income Household Strategies in the Caribbean Basin." *Social Forces* 74(2):633–55.

Kanaiaupuni, S. 2000. *Sustaining Families and Communities: Nonmigrant Women and Mexico–U.S. Migration Process.* Working Paper 2000–13. Madison: Center for Demography and Ecology, University of Wisconsin.

Kim, N. 2007. "The Impact of Remittances on Labor Supply: The Case of Jamaica." Policy Research Working Paper 4120. World Bank, Washington, D.C.

Kniesner, T. 1976. "An Indirect Test of Complementarity in a Family Labor Supply Model." *Econometrica* 44(4):651–69.

Kollmair, M., S. Manandhar, B. Subedi, and S. Thieme. 2006. "New Figures for Old Stories: Migration and Remittances in Nepal." *Migration Letters* 3(2):151–60.

Koolwal, G. 2007. "Son Preference and Child Labor in Nepal: The Household Impact of Sending Girls to Work." *World Development* 35(5):881–903.

Leeds, M., and P. von Allmen. 2004. "Spousal Complementarities in Home Production." *American Journal of Economics and Sociology* 63(4):795–812.

Lokshin, M., and E. Glinskaya. 2008. "The Effect of Male Migration for Work on Female Employment Patterns in Nepal." Policy Research Working Paper 4757. World Bank, Washington, D.C.

Lokshin, M., and Z. Sajaia. Forthcoming. "Impact of Interventions on Discrete Outcomes: Maximum-likelihood Estimation of the Binary Choice Models with Binary Endogenous Regressors." *Stata Journal.*

McKenzie, D., and H. Rapoport. 2005. "Migration Networks, Migration Incentives, and Education Inequality in Rural Mexico." *Paper presented at the Inter-American Development Bank Conference on Economic Integration, Remittances, and Development*, Washington, D.C., February.

Munshi, K. 2003. "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market." *Quarterly Journal of Economics* 118(2):549–99.

Nandini, A. 1999. *Engendered Mobilization—The Key to Livelihood Security: IFAD's Experience in South Asia.* Rome: International Fund for Agricultural Development.

Nepal Central Bureau of Statistics. 2003. *Population Census 2001: National Report.* Kathmandu: Central Bureau of Statistics.

———. 2004. *Nepal Living Standard Survey 2003/2004: Statistical Report.* Kathmandu: Central Bureau of Statistics.

Paris, T., A. Singh, J. Luis, and M. Hossain. 2005. "Labor Outmigration, Livelihood of Rice Farming Households and Women Left Behind: A Case Study in Eastern Uttar Pradesh." *Economic and Political Weekly* 40(25):2522–9.

Pfeiffer, L., and E. Taylor. 2007. "Gender and the Impacts of International Migration: Evidence from Rural Mexico." In A. Morrison, M. Schiff, and M. Sjöblom eds., *The International Migration of Women.* Washington, D.C.: World Bank and Palgrave Macmillan.

Raghuram, P. 2001 "Caste and Gender in the Organization of Paid Domestic Work in India." *Work Employment and Society* 15(3):607–17.

Rodriguez, E., and E. Tiongson. 2001 "Temporary Migration Overseas and Household Labor Supply: Evidence from Urban Philippines." *International Migration Review* 35(3):709–25.

Rosenzweig, M. 1980. "Neoclassical Theory and the Optimizing Peasant: An Econometric Analysis of Market Family Labor Supply in a Developing County." *Quarterly Journal of Economics* 94(1):31–55.

Sadiqi, F., and M. Ennaji. 2004. "The Impact of Male Migration from Morocco to Europe on Women: A Gender Approach." *Finisterra* 39(77):59–76.

Sanghera, J., and R. Kapur. 2000. *Trafficking in Nepal: Policy Analysis: An Assessment of Laws and Policies for the Prevention and Control of Trafficking in Nepal.* Kathmandu: Asia Foundation.

Sargan, J. 1958. "The Estimation of Economic Relationships Using Instrumental Variables." *Econometrica* 26:393–415.

Seddon, D., A. Jagannath, and G. Gurung. 2001. *The New Labures: Foreign Employment and the Remittance Economy of Nepal.* Kathmandu: Nepal Institute of Development Studies.

Stock, J., and M. Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." In D.W.K. AndrewsJ.H. Stock eds., *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg.* Cambridge: Cambridge University Press.

Thieme, S. 2005. *Social Networks and Migration: Far West Nepalese Labor Migrants in Delhi.* Münster, Germany: LIT Verlag.

UNDP (United Nations Development Programme). 2004. *Nepal Human Development Report 2004: Empowerment and Poverty Reduction.* Kathmandu: United Nation Development Programme.

Woodruff, C., and R. Zenteno. 2007. "Migration Networks and Microenterprises in Mexico." *Journal of Development Economics* 82(2):509–28.

World Bank. 2005. *Resilience Amidst Conflict: An Assessment of Poverty in Nepal 1995–96 and 2003–04.* Washington, D.C.: South Asia Poverty Reduction and Economic Management, World Bank.

Yamanaka, K. 2000. Nepalese Labour Migration to Japan: From Global Warriors to Global Workers. *Ethnic and Racial Studies* 23(1):62–93.

Yang, D. 2008. "International Migration, Remittances, and Household Investment: Evidence from Philippine Migrants' Exchange Rate Shocks." *Economic Journal* 118(528):591–630.

Political Accountability and Regulatory Performance in Infrastructure Industries: An Empirical Analysis

Farid Gasmi, Paul Noumba Um, and Laura Recuero Virto

The relationship between the quality of political institutions and the performance of regulation has recently assumed greater prominence in the policy debate on the effectiveness of infrastructure industry reforms. Taking the view that political accountability is a key factor linking political and regulatory structures and processes, this article empirically investigates its impact on the performance of regulation in telecommunications in time-series–cross-sectional data sets for 29 developing countries and 23 developed countries during 1985–99. In addition to confirming some well-documented results on the positive role of regulatory governance in infrastructure industries, the article provides empirical evidence on the impact of the quality of political institutions and their modes of functioning on regulatory performance. The analysis finds that the impact of political accountability on the performance of regulation is stronger in developing countries. An important policy implication is that future reforms in these countries should give due attention to the development of politically accountable systems. JEL codes: L51, H11, L96, L97, C23

Farid Gasmi is a professor at the Toulouse School of Economics, Atelier de Recherche Quantitative Appliquée au Développement Economique (ARQADE) and Institut d'Economie Industrielle (IDEI); his email address is gasmi@cict.fr. Paul Noumba Um (corresponding author) is a lead economist in the Sustainable Development Department of the Middle East and North Africa Region at the World Bank; his email address is pnoumbaum@worldbank.org. Laura Recuero Virto is an economist at the Organisation for Economic Co-operation and Development (OECD); her email address is laura.recuerovirto@oecd.org. An earlier version of this article was presented at the European Network for Training in Economic Research (ENTER) Jamboree, University of Mannheim, 2007; the Research Team on Markets, Employment, and Simulation (ERMES) seminar of the Université Panthéon-Assas Paris II, Paris, 2007; the European Conference on Competition and Regulation, Corfu, 2007; the conference of the European Association for Research in Industrial Economics, Valencia, 2007; the Télécom ParisTech conference on Telecommunications Infrastructure and Economic Performance, Paris, 2008; and the conference on Infrastructure Regulation, What Works, Why, and How Do We Know? Hong Kong, 2009. The authors thank the participants at these events for comments. They are grateful to Jean-Paul Azam, Aida Caldera, Antonio Estache, Frannie Léautier, Wilfried Sand-Zantman, the editor of the journal, and three anonymous referees for useful suggestions. They also thank Luis Hernando Gutierrez, Randeep Rathindran, and Lixin Colin Xu for help in constructing the data. Part of this research was undertaken during the summer of 2006 while Farid Gasmi and Laura Recuero Virto were visiting researchers at the World Bank Institute. These authors thank the members of the institute, in particular, Gabriela Chenet-Smith, for their warm hospitality.

THE WORLD BANK ECONOMIC REVIEW, VOL. 23, NO. 3, pp. 509–531 doi:10.1093/wber/lhp010
Advance Access Publication October 26, 2009

The last two decades have witnessed a worldwide wave of economic reforms affecting the market structure and the institutions in infrastructure industries, including high-technolgy sectors such as telecommunications and electricity and more traditional domains such as water and postal services. In developed countries, reforms have sought to improve the functioning of industries traditionally organized as—what has come to be recognized as—ill-performing monopolies. The policy task has been to redesign the legal and regulatory frameworks to produce "proper" economic incentives, to induce operators to enhance their offerings, in particular, their cost efficiency, service quality, and tariffs.

While the reforms in developing countries have been grounded on similar principles, in practice they have differed markedly in at least two respects. First, even though there was clearly room for improving the performance of infrastructure industries in developed countries, service was typically available in those countries, whereas it was absent in many parts of developing countries.

Second, and more important, the task of institutional design was far more challenging in developing countries. Developed countries needed to modernize an existing fabric of institutions with a complex system of operating rules built over a long history of political and economic administration of market economies. In most cases, although for different reasons, such crucial experience was lacking in developing countries. Beyond having to establish new institutions to regulate the reformed industries, a challenge in itself due to the scarcity of human capital, developing countries had to deal with inefficient administrative rules.

Following liberalization and privatization of some segments of infrastructure industries and the creation of regulatory authorities, developing countries had to devote considerable effort to improve the efficiency of the new regulatory authorities, by ensuring regulatory independence, adequate human resources capacity, and sound regulatory governance. Meanwhile, theoretical work on designing optimal regulatory institutions and empirical work on measuring their performance suggest that these three policy goals must be considered in the context of governance of the economy as a whole. This article investigates the relative weight of these sector-specific and economywide determinants of regulatory performance in the telecommunications sector, using econometric analysis of separate data sets for developing and developed countries.[1]

The determinants of regulatory performance have been discussed in both the theoretical and the empirical streams of the literature on infrastructure industry

1. This article considers the cellular and the fixed-line telecommunications segments and analyzes the developing and developed country data sets separately. These two segments were chosen because they have typically required regulatory intervention, often in controlling retail and wholesale prices or setting service targets. Developing and developed countries are analyzed separately because the potential gains in estimate efficiency associated with larger data sample size were considered less important than the potential inconsistency in estimates associated with greater data heterogeneity. To avoid adding econometric complexity, the fact that the data from developed and developing countries are not always directly comparable is dealt with in the discussion of the final results.

regulation. For this analysis, two approaches are distinguished. A first approach, conceptual in nature and inspired by political science, argues that it is political governance that is the relevant determinant of regulatory performance (Spiller and Tommasi 2003). Another, more empirical approach emphasizes the impact of regulatory governance on performance (Cubbin and Stern 2005b). The analysis in this article views the relationship between political and regulatory structures and processes as critical in assessing regulatory performance. It seeks to merge both approaches, inserting some empirical elements into the debate on the relationship between political and regulatory institutions that has so far taken place mainly at a conceptual level.

To do this, a series of econometric significance tests are run, with special attention to variables that capture the degrees of political accountability in the economy. How politically accountable an economic system is depends on how well implemented is the "pro-active process by which public officials inform about and justify their plans of action, their behavior and results, and are sanctioned accordingly" (Ackerman 2005, p. 6). The analysis considers political accountability to be fundamental to the link between political structures and regulatory processes and hence views its (political-game) equilibrium level as an important determinant of the performance of regulatory processes. With that in mind, a testing procedure is established for the hypothesis that, all things being equal, more political accountability should enhance the performance of regulation. In addition to testing the significance of political accountability, the analysis gives some empirical substance to the conjecture that the effect of political accountability is even stronger in developing countries.[2]

The article is organized as follows. Section I summarizes some of the recent theoretical and empirical arguments on the design of institutions and on the evaluation of regulatory performance in infrastructure industries. Not meant to be exhaustive, this section argues the need to merge these two streams of the literature on regulatory institutions. Section II describes the data and some of their general properties. Section III presents the empirical analysis of the relationship between political accountability and regulatory performance. The article concludes with some policy implications of the empirical findings. The appendix contains some summary statistics on the data.

I. Design of Institutions and Regulatory Performance: The Need for an Integrated Empirical Approach

Recent contributions to the theory of the design of institutions and empirical work on measurement of their performance have exposed the issue of the

2. From a normative perspective, with better regulatory performance expected to improve social welfare, this suggests that the marginal social benefit of political accountability is higher in developing countries.

evaluation of regulatory performance. Laffont (2005) meditates on the design of regulatory institutions in developing countries.

Two approaches have been used to examine the determinants of regulatory performance and outcomes. One approach is conceptual and analyzes the role of political structures and processes. Another approach, more empirical, emphasizes the impact of the quality of regulatory governance. This section briefly reviews the main arguments of these two approaches and highlights the need to develop a unified analytical framework. The rest of the article is an empirical effort in that direction.

The theoretical approach analyzes the relationship between political structures and processes and regulation by emphasizing the need to open the black box of the organization and functioning of governments (Estache and Martimort1999; North 2000).[3] In an analysis of the link between politics and regulation in the United States, McCubbins, Noll, and Weingast (1987) argue that by reducing the costs of monitoring and by sharpening sanctions, administrative procedures can give rise to an equilibrium in which compliance with the preferences of political agents is greater than it otherwise would be.[4] This relationship is further explored in the telecommunications sector by Levy and Spiller (1994), through case study analysis. In particular, they evaluate the potential for political agents to manipulate the regulatory process. They find that sector performance can be satisfactory under a wide range of regulatory procedures as long as arbitrary administrative decisions can be restrained.

The link between the political and regulatory spheres is further analyzed in Spiller and Tommasi (2003), through the impact that political environments have on the ability of political agents to achieve cooperation over time. They argue that long-term political cooperation is likely to lead to stable and flexible regulatory policies and thus to effective regulation. This is especially the case when the agents with decision power have strong intertemporal relationships, policy and political moves are widely observable, good enforcement technologies are available, and the short-run payoffs from noncooperation are not high. They argue that less efficient regulatory rules resulting from a rigid regulatory context may provide incentives for investment, whereas regulatory discretion may lead to arbitrary outcomes if institutional endowments are low.

Heller and McCubbins (1996) argue that incentives for investing in infrastructure industries are not credible within a given regulatory structure without a political context that makes them sustainable. Regulatory predictability is crucial to credibility, and political institutions play an important role in

3. By emphasizing the political game, this approach fits within the new institutional economics paradigm, which is grounded in the precepts of transaction cost theory and positive political economy. This paradigm constitutes an important departure from the standard normative approach to public economics.

4. Bottom-up "fire-alarm" monitoring through external agents affected by regulatory policies is a good example of a method that can reduce the information costs of monitoring the activities of agencies (McCubbins and Schwartz 1984).

enhancing this predictability. The higher the quality of the political and institutional environment, the harder it is to change regulatory structures and procedures. In particular, the greater the number of political players with effective authority and veto power, the easier it is to block policy change. The main argument of this line of policy research is that the more established the political structures and processes, the higher the cost of institutional change and the more efficient the conduct of regulation.

The fundamental belief motivating much of the empirical approach that emphasizes the role of regulatory governance in infrastructure industries is that good regulatory governance is a prerequisite to the proper functioning of the positive relationship between regulatory incentives and regulatory performance. This belief is based on the conjecture that "regulatory agencies with better governance should make fewer mistakes, have their mistakes identified and rectified better and more quickly, so that good regulatory practice is more readily established and maintained" (Cubbin and Stern 2005a, p. 3).

The basic empirical implications of these hypotheses is that the structure and practice of regulation thereby entailed—an independent regulator making transparent regulatory decisions—mean that better regulatory governance increases supply capacity and enhances productive and allocative efficiency. In telecommunications, these implications are typically tested with data collected for a set of developing countries observed during a given period. Regulatory performance is measured by mainline coverage rates or mainlines per employee, and regulatory governance is captured by an index that aggregates a set of aspects related to the structure and internal organization of regulation (Gutierrez 2003b).[5] Overall, when applied to telecommunications (Gutierrez 2003a) and electricity (Cubbin and Stern 2005a), the methodology yields a positive impact of regulatory governance on such regulatory output measures. For a survey of empirical studies on regulatory governance and performance in developing countries, see Cubbin and Stern (2005b).

A typical contribution to this line of research starts with the global conceptual view that "institutional quality is the dominant determinant of variations in long-term growth performance" (Cubbin and Stern 2005a, p. 2; Rodrik, Subramanian, and Trebbi 2004). However, the research often accounts only for the micro dimensions of institutional quality embodied in what is referred to as the "quality" of regulatory governance. This approach could gain substantially in richness by drawing lessons from the literature on the design of institutions, discussed earlier in this section. The analysis here takes a step toward a unified approach that explicitly incorporates variables linking political and regulatory structures and processes when evaluating regulatory

5. These studies and this one use outcome variables to measure regulatory performance. A more rigorous assessment of regulatory performance entails conducting surveys to capture the quality of regulators' decisions that ultimately affects sector outcomes (see Correa and others 2008; Brown and others 2006). Such surveys do not exist but would, if undertaken, provide a better indication of the performance of regulation in infrastructure industries.

performance, in addition to specifying variables of regulatory governance. The impact of political accountability is captured through variables accounting for macro dimensions of institutional quality that are seen as affecting the level of political accountability in the economic system.

The approach taken here rests on the belief that limiting the use—and sanctioning the abuse—of political power should help in disentangling regulatory processes from the opportunistic behavior of political agents.[6] The election mechanism should, in principle, ensure political accountability since citizens select the representatives who hold bureaucrats and members of the judiciary system accountable for their behavior. However, this property of elections is hard to satisfy since the electoral process suffers from important information asymmetries between elected politicians and citizens and from lack of accountability of politicians for their past actions. Privatization of government monopolies, liberalization of markets, and the application of private management principles to state-owned entreprises have been demonstrated to improve political agents' accountability much more directly. However, while it is important to consider such pro-accountability reforms, the independence of the regulator, and other factors related to the sector's regulatory governance when analyzing regulatory performance, it is also important to consider other pro-accountability factors related to governance of the economy as a whole—as in the empirical analysis that follows.

II. The Data

The data set on developing countries includes Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Côte d'Ivoire, Dominican Republic, Ecuador, El Salvador, Ghana, Guatemala, Honduras, India, Jamaica, Jordan, Kenya, Malawi, Malaysia, Morocco, Pakistan, Panama, Peru, South Africa, Sri Lanka, Tanzania, Thailand, Uganda, and Venezuela. The data set on developed countries includes Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom, and the United States. For each country, data were collected on variables regrouped into four categories: regulatory performance, local accountability, global accountability, and other variables (table 1). (For detailed definitions of these variables and their data sources, see Gasmi, Noumba, and Recuero Virto 2006.)

As indicated, regulatory performance is measured by the level of output (mainline coverage or cellular subscriptions), efficiency (mainlines per employee), or price (fixed-line residential service, cellular service).[7] To match

6. As Spiller and Tommasi (2003) note, opportunistic behavior by politicians can be expected in infrastructure industries because the economic stakes are large.

7. These outcome variables are indirect measures of regulatory performance based on objective data on regulated firms rather than on direct measures based on subjective data reported by surveyed regulatory agencies.

TABLE 1. Variables and Their Designation

Variable	Designation
Regulatory performance	
ml	Mainline coverage
cel	Cellular subscription
eff	Mainlines per employee
p_res	Price of monthly subscription to fixed-line service
p_cel	Price of cellular service
Local accountability	
reg	Regulatory governance index
Global accountability	
corruption	Corruption
bureau	Bureaucracy
law	Law and order
expropri	Expropriation
currency	Currency risk
institutional	Institutional environment index
checks	Checks and balances
Other variables	
priva	Privatization
comp_fix	Competition in fixed line
comp_cel	Competition in cellular
rural	Rural population
density	Population density

Note: For defintions of variables, see Gasmi, Noumba, and Recuero Verto (2006).
Source: Authors' analysis based on data described in text.

the conceptual framework discussed in the previous section, variables were regrouped into local and global accountability categories, representing the quality of regulatory governance in the sector and political governance at the economywide level. Therefore, local accountability is captured in variables reflecting the regulator's political and financial independence, the transparency of accounts and regulatory decisions, the clarity of the allocation of responsibilities across institutions, the nature of the legal environment, and the degree of social participation in regulatory decisions.[8] Global accountability is captured in variables reflecting the quality of the institutional framework (government integrity, efficiency of bureaucracy, strength of courts and enforcement

8. The study thus contributes to the literature on the impact of infrastructure industry reforms by extending the set of variables capturing regulatory governance. In that respect, it stands with Gutierrez (2003a), who has constructed detailed indices of regulator characteristics for the telecommunications sector in Latin American countries, and Holder and Stern (1999), who have done the same for the electricity sector in Asian countries. Estache and Martimort (1999) emphasize the importance of these dimensions to the sustainability of regulatory agencies. In the samples for this study, the regulator became independent at some point during the period under study in 26 of the 29 developing countries and 21 of the 23 developed countries.

capacity, government commitment capacity, and currency risk) and the quality of the political process (strength of checks and balances).[9]

The variables in the group of other variables control for some effects deemed important when estimating the relationship between political accountability and regulatory performance. Because the telecommunications sector has undergone considerable market structure changes during the period under study, some reform variables are included to reflect these changes, such as privatization of the incumbent and the introduction of competition in fixed and cellular service, as liberalization of these segments has arguably had different market implications (Gasmi and Recuero Virto forthcoming). In the data set on developing countries, 18 of 29 countries partially privatized their telecommnications operator, 14 introduced competition in the local fixed-line segment, and 24 introduced competition in the cellular segment. In the data set on developed countries, 20 of 23 countries partially privatized their telecommnications operator, 10 introduced competition in the local fixed-line segment, and 15 introduced competition in the cellular segment. In both groups, the reforms have coincided with the introduction of new technologies that have substantially reduced costs and increased demand. This group of other control variables thus includes some country-specific demand features that provide information on population density and distribution (urban or rural).

Correlation coefficients between the variables of political accountability and those of regulatory performance show that the correlation is generally stronger for developing countries than for developed countries (table 2). (The appendix provides some summary statistics on the data for developing and developed countries.) Correlation is particularly strong when regulatory performance is measured by mainline coverage, cellular subscription, and mainlines per employee and when political accountability is captured by the strength of checks and balances. The same is true when regulatory performance is measured by mainlines per employee and political accountability by the regulatory governance index, when regulatory performance is measured by cellular subscription or price of cellular service, and when political accountability is captured by the quality of the institutional environment. In both samples, the regulatory performance variables tend to be correlated relatively more strongly with the variables that reflect the quality of the broad institutional environment than with those that reflect the quality of regulatory governance in the sector.

It is instructive to examine the evolution of these variables over the sample period (tables S1 and S2 in the supplemental appendix). When measured by mainline coverage, cellular subscription, or mainlines per employee, regulatory performance has increased twice as much on average in developing countries as

9. Both the empirical and the theoretical literature suggest that it is less the extent of democracy that is relevant to investors and more the ability of the government to credibly commit to a policy regime. The level of policy stability is captured here through an index indicating whether there are an "effective" number of checks and balances.

TABLE 2. Correlation Coefficients for Developing and Developed Countries

	Regulatory performance									
	ml		cel		eff		p_res		p_cel	
Political accountability	Developing countries	Developed countries	Developing countries	Developed countries	Developing countries	Developed countries	Developing countries	Developed countries	Developing countries	Developed countries
Global accountability										
institutional	0.41	0.63	0.65	0.24	0.42	0.22	0.23	0.28	0.60	0.01
checks	0.34	0.07	0.39	0.04	0.36	0.01	−0.01	0.12	0.30	0.24
Local accountability										
reg	0.19	0.43	0.57	0.55	0.30	0.05	−0.06	0.01	0.61	−0.07

Source: Authors' analysis based on data described in text and in Gasmi, Noumba, and Recuero Virto (2006).

in developed countries, most likely reflecting the much higher level of unmet demand in developing countries in the early part of the study period. In contrast, when measured by the price of monthly subscription to fixed-line service, which has increased in both developing and developed countries, or the price of cellular service, which has decreased, regulatory performance has improved more noticeably in developed countries. This conclusion should be moderated, however. The significantly greater increase in the price of monthly subscription to fixed-line service in developing countries might be due to the more intense tariff rebalancing in these countries. Furthermore, the significantly lower decline in the price of cellular service in developing countries might reflect a relatively less mature segment of the market—and hence with less effective competition—than in developed countries.

This brief review of the data also reveals greater improvement in the quality of the institutional environment and the political process in developing countries than in developed countries during the period under study. However, again, caution is required in interpreting this observation as it might reflect only the fact that these countries lagged considerably behind on these two dimensions.

III. Empirical Analysis of the Relationship between Political Accountability and Regulatory Performance

This section briefly reviews the methodology, summarizes some results on data stationarity and Granger causality, and discusses the results of the regressions of regulatory performance and measures of political accountability. (For more details on the methodology, see Gasmi, Noumba, and Recuero Virto 2006.)

Econometric Methodology

As the data sets include time-series and cross-sectional data, differenced generalized method of moments (DIF-GMM) was used for estimating. Preliminary statistical tests supported the presence of dynamic and fixed effects and suggested the use of this method, developed by Arellano and Bond (1991) for analyzing panel data and applied by Beck and Katz (2004) to time-series and cross-sectional data.

A typical relationship is specified as a dynamic equation given by

$$\log(y_{it}) = \alpha_0 + \alpha_1 \log(y_{it-1}) + \mathbf{x}'_{it}\beta + \mu_i + \varepsilon_{it}, \tag{1}$$

where $i = 1, 2, \ldots, N$; $t = 1, 2, \ldots, T$; y_{it} is a one-dimensional dependent variable representing regulatory performance; α_0 and α_1 are scalar parameters; $\mathbf{x}_{it}$ is a vector of regressors representing, among other things, political accountability, in country i at time t; β is the associated vector of parameters; μ_i captures a country-specific fixed effect; and ε_{it} is a disturbance term. For both data sets,

$T = 15$. For the developing country data set, $N = 29$, and for the developed country data set, $N = 23$. Standard assumptions $E(\mu_i) = 0$, $E(\varepsilon_{it}) = 0$, $E(\varepsilon_{it}\mu_i) = 0$, and $E(y_{i1}\varepsilon_{it}) = 0$ are made on the fixed-effect and disturbance terms.

In this setting, estimation can potentially be plagued by endogeneity coming from a correlation between the regressors and the fixed-effect term and a correlation between the regressors and the disturbance term.[10] The endogeneity problem stemming from the correlation of the first type is taken care of by expressing equation (1) in first differences. However, this transformation brings with it another endogeneity problem due to the contemporaneous correlation between log (y_{it-1}) and the error term ε_{it-1}. This correlation is of the same nature as the correlation of the second type. Thus, the endogeneity problem basically comes down to finding instruments to use in estimating this equation in first differences. The standard approach, followed here, is to select instruments from lagged values of the potentially endogenous regressors.

Before estimating the equation, the technical issue of stationarity of the dependent variable must be addressed because lack of stationarity can have two undesirable consequences in this context. One is that any estimation method applied to a nonstationary dynamic system is likely to yield inaccurate estimates. Another consequence, related to the application of DIF-GMM, is that the available instruments for the equation in first differences are likely to be weak, which would impoverish the finite-sample properties of the estimator. A method suggested by Blundell and Bond (1998, 1999) is used to address stationarity (see also Arellano and Bover 1995).

As indicated, this investigation of the effect of political accountability relies on a set of regressions. While estimation of the coefficients enables assessment of the (quantitative) impact of the political accountability variables on the regulatory performance variables, first asking whether there is a causal relationship between these variables permits meaningful interpretation of this impact. For this purpose, the DIF-GMM estimation technique is combined with a Granger-causality testing procedure developed by Holtz-Eakin, Newey, and

10. In this context, a correlation might be expected between the extent of reforms, captured by some regressors, and some country characteristics, such as country size and wealth, which are embodied in the fixed-effect term. Moreover, the regressors used to capture the degree of privatization and competition are likely to be endogenous, especially in the early stages of reform (Ros 1999). For example, licenses are typically granted conditional on the fulfillment of specified performance targets based on coverage, quality, or some other dimensions of the industry and are often associated with exclusivity periods. Endogeneity might also be a concern when using variables to capture some aspects of the structure of regulatory institutions. An example is the variable on the existence of an independent regulator, since the decision to create an independent regulator and its timing can be influenced by pre-regulatory performance. For an empirical account of the endogeneity of regulatory policies, see Gasmi and Recuero Virto (forthcoming), Gutierrez (2003b), and Ros (2003), among others.

Rosen (1988) for panel data. The following equation is estimated:

$$\Delta \log(y_{it}) = \sum_{k=1}^{m} \alpha_k \Delta \log(y_{it-k}) + \sum_{k=1}^{m} \delta_k \Delta x_{it-k} + \Delta \mathbf{x}'_{it}\beta + \Delta\varepsilon_{it}, \qquad (2)$$

where Δ is the first difference operator. This equation tests whether a variable used to capture political accountability, x, Granger-causes the variable to measure regulatory performance y.

Results

On stationarity, tables S3 (for developing countries) and S4 (for developed countries) in the supplemental appendix show the results of the estimation of a first-order autoregressive process, AR(1), with both the DIF-GMM and system (SYS)-GMM methods applied to the variables in levels and a time trend. The tables also show the results for the DIF-GMM method applied to the variables that capture regulatory performance in first differences where they are found to be nonstationary in levels. The tables give the DIF-GMM and SYS-GMM (one-step robust) estimates of the AR(1) coefficient; the estimate of the time trend coefficient, *Time*; the first- and *n*th-order autocorrelation coefficients of the residuals in first differences, *m*1 and *mn*, respectively; the value of the *J*-statistic for testing the validity of the instruments; the value of the Dif-Sargan statistic for testing the validity of the additional SYS-GMM conditions; the value of the starting lag of the instruments, *L*; and the number of observations. Based on the analysis of the results, the series in first differences was used for both the developing and developed countries data sets (see Gasmi, Noumba, and Recuero Virto 2006 for details).

On the existence of causal relationships, tables S5–S10 show the DIF-GMM estimation results on which the testing procedures are built, asking whether the variables of local accountability (the regulatory governance index *reg*) and global accountability (the institutional environment index, *institutional*; and the index of checks and balances, *checks*) Granger-cause the variables of regulatory performance (mainline coverage, *ml*; cellular subscription, *cel*; mainlines per employee, *eff*; price of monthly subscription to fixed-line service, *p_res*; and price of cellular service, *p_cel*).[11] In addition to showing the estimated values of the parameters associated with the explanatory variables, tables S5–S10 include three Wald statistics. Goodness of fit tests the joint significance of the coefficients associated with the explanatory variables. Lag length tests the joint significance of the coefficients associated with the dependent variable and the political accountability variable with the greatest lag length. Causality tests the joint significance of the coefficients associated with the lagged political accountability variables when the lag length test accepts the significance of the

11. Some additional control variables are also included, as needed, and account for any possible endogeneity problem. The estimates shown in these tables are those of the parameters of equation (2).

TABLE 3. Granger-causality relationships for developing and developed countries

	Local accountability		Global accountability			
	reg		institutional		checks	
Variable	Developing countries	Developed countries	Developing countries	Developed countries	Developing countries	Developed countries
ml	Yes	No	Yes	Yes	Yes	No
cel	No	Yes	Yes	Yes	Yes	Yes
eff	No	No	Yes	No	No	No
p_res	Yes	Yes	Yes	Yes	Yes	No
p_res	Yes	No	Yes	No	No	No

Note: Yes indicates evidence of a causal relationship running from the accountability variable to the regulatory performance variable; No indicates no evidence of a causal relationship.

Source: Authors' analysis based on data described in text and in Gasmi, Noumba, and Recuero Virto (2006).

coefficients. The results reported in these tables inform the choice of valid instruments.

For developing countries, the results in all estimations indicate the existence of an acceptable lag length. The Granger-causality test shows that causality runs from regulatory governance to regulatory performance, except when cellular subscription or mainlines per employee variables are used to measure regulatory performance (see table S5). The institutional environment has a causal effect on regulatory performance independently of which of the five variables is used to measure regulatory performance (see table S6). Finally, the political process has a causal effect on regulatory performance except when performance is measured by the variable mainlines per employee or price of cellular service (see table S7).

While some causal relationships are also found in the data on developed countries, the empirical evidence is somewhat weaker (tables S8–S10). In some estimations there is no lag length that is statistically significant and thus no Granger-causality relationship is accepted. For example, when mainline coverage or price of cellular service is used to test whether regulatory governance has a causal relationship with regulatory performance, no lag length is accepted (see table S8). Similarly, the estimations for developed countries do not show causal relationships between the institutional environment and regulatory performance when performance is measured by mainlines per employee or price of cellular service (see table S9) or between the political process and regulatory performance when performance is measured by price of cellular service (see table S10). In cases where a certain lag length is accepted, Granger-causality tests show that regulatory governance is causally related to regulatory performance when performance is measured by cellular subscription or price of monthly subscription to fixed-line service (see table S8) and that institutional

environment is causally related to regulatory performance when performace is measured by mainline coverage, cellular subscription, or the price of monthly subscription to fixed-line service (see table S9). Finally, the political process is causally related to regulatory performance only when performance is measured by cellular subscription.

Table 3 summarizes the findings on the existence of causal relationships in the two data sets. Overall, the results support the proposition of a causal relationship between political accountability and regulatory performance in both developing and developed countries. This is especially so when political accountability is examined through the quality of the institutional environment. The causal relationships with regulatory performance are stronger for the global accountability variables than for the local accountability variables, especially in developing countries. Even though the empirical evidence of such relationships is stronger for developing countries, the policy implications of the issue warrant careful analysis of the quantitative aspects of these relationships.

The analyses conducted thus far set the ground for closer inspection of the relationship between political accountability and regulatory performance in the two data sets. The Granger-causality tests provided both empirical evidence on the causal relationships and information on the dynamic structure of the relationships and resulted in a list of potential variables for use as regressors when estimating the quantitative impact of political accountability on regulatory performance. To minimize the risk of estimation inaccuracy, a serious concern in dynamic data analysis, the variables used to measure regulatory performance were transformed to make them stationary when needed.

Tables 4 and 5 report DIF-GMM regressions in which some of the main political accountability regressors are drawn from the set of variables that passed the Granger-causality test. The tables, similar to tables S5–S10 of the supplemental appendix, contain three additional items. First, they include two country-specific variables, population density (*density*) and extent of rural population (*rural*). Second, they indicate whether the regressors' privatization (*priva*), competition in fixed line (*comp_fix*), competition in cellular (*comp_cel*), and regulatory governance index (*reg*) were found to be endogenous. Valid instruments are chosen according to the procedure described above.[12] Third, they give the value of a Wald statistic for testing the joint significance of time-specific effects captured in time dummy variables.

For developing countries, at least one variable for political accountability significantly affects each of the five variables for regulatory performance (see table 4). Except when regulatory performance is measured by the price of monthly subscription to fixed-line service, the sign of the impact is as expected—the greater the political accountability, the better the regulatory

12. In the estimations reported in these tables, the disturbance term in levels does not exhibit any serial correlation except for the series with mainline coverage, where it follows an MA(3), and the series for mainline per employee and cellular subscription, where it follows an MA(1).

TABLE 4. Differenced generalized method-of-moments parameter estimates, developing countries

y_{it}	ml_{it}	cel_{it}	eff_{it}	y_{it}	p_res_{it}	p_cel_{it}
y_{it-1}	0.248*	0.329**	−0.136*	y_{it-1}	−0.241**	−0.221***
reg_{it-1}	0.003**			reg_{it-1}	−0.010*	
$corruption_{it-1}$		0.080***	0.007	reg_{it-2}		−0.008**
$bureau_{it-1}$		−0.010	0.021	$corruption_{it-1}$	−0.003	0.030
law_{it-1}		0.017	0.019*	$bureau_{it-1}$	0.001	0.040
$expropri_{it-1}$		0.002	−0.001	law_{it-1}	0.035	0.003
$currency_{it-1}$		0.002	−0.004	$expropri_{it-1}$	0.218***	−0.029
$corruption_{it-3}$	−0.012			$currency_{it-1}$	−0.003	−0.034***
$bureau_{it-3}$	0.004			$checks_{it-1}$	0.035**	
law_{it-3}	0.006			$priva_{it}$	0.185*	0.373*
$expropri_{it-3}$	0.011			$comp\_fix\_it$	−0.216	0.026
$currency_{it-3}$	−0.004			$comp\_cel\_it$	0.001	0.072
$checks_{it-1}$		0.007*		$rural_{it}$	0.108*	0.036**
$checks_{it-2}$	0.003**			$density_{it}$	−0.001	−0.008
$checks_{it-3}$	0.001			$m1$	−2.00**	−2.00**
$priva_{it}$	0.067**	0.174*	0.249***	$m2$	−0.81	−0.88
$comp\_fix_{it}$	−0.004	0.033	−0.137***	J	5.12	5.86
$comp\_cel_{it}$	0.021*	0.108**	0.046	Time dummy variables	15.36***	4.68***
$rural_{it}$	−0.002	0.007	−0.003	Endogenous reforms	No	Yes
$density_{it}$	0.001	−0.003*	0.005	L	2	2
$m1$	−3.20***	−2.74***	−3.06***	Number of observations	150	162
$m2$		0.73		Goodness of fit	116.81***	15.50***
$m3$			1.32			
$m5$	0.94					
J	1.33	3.91	7.62			
Time dummy variables	3.01***	8.20***	2.11*			

(Continued)

TABLE 4. Continued

y_{it}	ml_{it}	cel_{it}	eff_{it}	y_{it}	p_res_{it}	p_cel_{it}
Endogenous reforms	Yes	No	Yes			
L	5	2	3			
Number of observations	295	318	316			
Goodness of fit	32.83***	130.40***	10.10***			

*Significant at the 10 percent level; **significant at the 5 percent level; ***significant at the 1 percent level.

Note: For notational simplicity, the log and Δ transformations are implicit; *mn* refers to the autocorrelation coefficient of order *n* of the residuals in first differences found in the procedure used to search for (lag) instruments. The entry in the table corresponds to the value of its *t*-statistic. *J* is the chi-square statistic used to test for the validity of the instruments. A high value indicates that the instruments are valid. Time dummy variables are a Wald (chi-square) statistic used to test the significance of time-specific effects. A yes or no for endogenous reforms indicates that the variables *priva*, *comp_fix*, *comp_cel*, and *reg* did or did not improve the overall goodness-of-fit statistic by being considered as endogenous and hence were or were not instrumented. L ($L - 1$) is the starting lag for the instruments used in the equation in first differences (levels).

Source: Authors' analysis based on data described in text and in Gasmi, Noumba, and Recuero Virto (2006).

TABLE 5. Differenced generalized method-of-moments parameter estimates, developed countries

y_{it}	ml_{it}	cel_{it}	p_res_{it}
y_{it-1}	0.039	0.503***	−0.159
reg_{it-2}		0.009*	0.006
$corruption_{it-1}$		−0.043	0.038
$bureau_{it-1}$		0.012	0.073
law_{it-1}		0.046	0.026
$expropri_{it-1}$		0.072	0.269
$currency_{it-1}$		0.007	−0.037**
$corruption_{it-2}$	0.007		
$bureau_{it-2}$	−0.005		
law_{it-2}	−0.006		
$expropri_{it-2}$	−0.006*		
$currency_{it-2}$	0.001		
$checks_{it-1}$		0.021*	
$priva_{it}$	−0.014*	0.031	0.073
$comp\_fix_{it}$	0.012	−0.082**	0.109*
$comp\_cel_{it}$	−0.006	−0.025	−0.007
$rural_{it}$	0.003	0.023	−0.032
$density_{it}$	0.001	0.008	0.020***
$m1$	−2.87***	−3.52***	−2.49**
$m2$	0.09		−0.87
$m3$		0.11	
J	4.41	0.00	3.55
Time dummy variables	3.74***	3.57***	75.60***
Endogenous reforms	Yes	Yes	No
L	2	3	2
Number of observations	276	275	182
Goodness of fit	62.31***	53.35***	11,625.52***

*Significant at the 10 percent level; **significant at the 5 percent level; ***significant at the 1 percent level.

Note: For notational simplicity, the log and Δ transformations are implicit. *mn* refers to the autocorrelation coefficient of order *n* of the residuals in first differences found in the procedure used to search for (lag) instruments. The entry in the table corresponds to the value of its *t*-statistic. *J* is the chi-square statistic used to test for the validity of the instruments. A high value indicates that the instruments are valid. Time dummy variables are a Wald (chi-square) statistic used to test the significance of time-specific effects. A yes or no for endogenous reforms indicates that the variables *priva*, *comp_fix*, *comp_cel*, and *reg* did or did not improve the overall goodness-of-fit statistic by being considered as endogenous and hence were or were not instrumented. L ($L - 1$) is the starting lag for the instruments used in the equation in first differences (levels).

Source: Authors' analysis based on data described in text and in Gasmi, Noumba, and Recuero Virto (2006).

performance as reflected in higher output (increase in mainline coverage and cellular subscription), higher efficiency (increase in mainlines per employee), and lower prices (decrease in price of cellular service).

The apparently counterintuitive case, where greater political accountability (less risk of expropriation for operators and stronger checks and balances)

TABLE 6. Impact of Political Accountability on Regulatory Performance in Developing and Developed Countries

	Local accountability		Global accountability			
	reg		institutional		checks	
Variable	Developing countries	Developed countries	Developing countries	Developed countries	Developing countries	Developed countries
ml	+	NA	NS	−	+	NA
cel	NA	+	+	NA	+	+
eff	NA	NA	+	NA	NA	NA
p_res	−	NA	+	−	+	NA
p_cel	−	NA	−	NA	NA	NA

NA is not applicable; NS indicates that no significant effect was found; + indicates a positive and significant effect; − indicates a negative and significant effect.

Source: Authors' analysis based on data described in text and in Gasmi, Noumba, and Recuero Virto (2006).

leads to a higher price of monthly subscription to fixed-line service, might in fact only reflect the extent of tariff rebalancing that typically takes place in developing countries during the early stages of reform. When local accountability (regulatory governance) is distinguished from global accountability, global accountability is more often found to have a significant impact on regulatory performance. Nevertheless, in the cases when regulatory governance is found to be significant, its effect on regulatory performance has the expected sign.

The results for the developed country data sets do not convey the same messages, with results that are generally poor compared with those for developing countries (see table 5). Indeed, the only expected results were found when using mainline coverage, cellular subscription, or price of monthly subscription to fixed-line service to measure regulatory performance. For the impact of political accountability on regulatory performance, the only expected result was a positive effect of regulatory governance and checks and balances on cellular subscription and the price of monthly subscription to fixed-line service, which decreases when the currency risk to operators diminishes.

The dummy variables used to capture time-specific effects are highly significant, suggesting that attention should be given to important political and economic events in a country when examining the performance of regulation. The reform variables were found to be endogenous in all the regressions except when regulatory performance was measured by cellular subscription and by the price of monthly subscription to fixed-line sevice in the developing country regressions and by the price of monthly subscription to fixed-line service in developed country regressions. These results are consistent with the idea that reforms are increasingly performance based.

Overall, then, there are reasons to believe that local accountability (here synonymous with regulatory governance) generally affects regulatory

performance in a significant way in developing as well as developed countries. The story is not so clear for global accountability. For developing countries, the quality of the political process and the institutional environment have a favorable impact on regulatory performance in the telecommunications industry when performance is evaluated by the level of output, price, or efficiency. For developed countries, however, while the quality of the political process has a positive impact on regulatory performance when performance is measured by output, the impact was ambiguous when regulatory performance is measured by output and prices.

Table 6 summarizes the findings. The variable *institutional*, an index reflecting the quality of the institutional environment, is constructed by aggregating five indices on the extent of corruption in the country (*corruption*), the burden of the bureaucracy (*bureau*), the strength of the judicial system and the degree of observance of the law (*law*), the risk of expropriation through asset confiscation or nationalization (*expropri*), and the risk of losses to operators due to exchange rate fluctuations (*currency*). The effect of this aggregate index is consistent with the effects of the individual indices that compose it (shown in tables 4 and 5).

IV. Conclusion

Two data sets, one on developing countries and one on developed countries, were used to estimate the impact of political accountability variables on outcome variables measuring regulatory performance. The empirical analysis showed a relatively weak effect of political accountability on the performance of regulation in developed countries and a much more clear-cut effect in developing countries, where greater political accountability yields higher regulatory performance.

What implications can be derived from these findings for the telecommunications industry and, to some extent, for other infrastructure industries?[13]

During the last two decades, many developing countries have created regulatory agencies mostly on the model recommended by international financial institutions and international lawyers. However, these new regulatory institutions were not adequately adapted to local cultural, political, and social endowments. This article again stresses this as an important requirement for success in developing new institutions. Furthermore, the article goes beyond most recent analyses by extending the focus to issues of global (economywide) accountability, which reflect the quality of political institutions.

13. Although the results found for the telecommunications industry might be expected to hold for other infrastructure industries in their general spirit, this study calls for a careful examination of alternative sectors such as energy or water that, among other things, accounts for the industry technological specificities.

Recent contributions to the literature have deepened understanding of regulatory effectiveness along two dimensions: regulatory governance, a concept that is somewhat broader than encompassed by the concept of local accountability here, and regulatory substance, a concept that captures how regulation is actually implemented. Brown and others (2006) have proposed a comprehensive process for evaluating the effectiveness of regulatory institutions that highlights not only structural weaknesses but also deficiencies stemming from the surrounding environment of regulation, particularly the political environment.

Thus, it is important to devise policy mitigation instruments that incorporate both of these dimensions. Instead, donor interventions over the last decade or so have centered on structural issues. The empirical analysis reported in this article strongly supports defining a set of instruments of effective intervention, with the objective of improving political accountability in the implementation of regulation. Building sound regulatory institutions in developing countries should be viewed as part of a broader strategy of good governance rather than merely as a sectoral matter, as has been the case.

Building regulatory institutions in developing countries has proven to be more complex than expected. The analysis here shows that regulatory governance is a necessary condition for regulatory performance but it is not sufficient to ensure good performance. Political accountability matters for the way regulatory institutions operate and make decisions. Deepening the understanding of these relationships calls for a better assessment of the political economy of infrastructure reforms as well as better analysis and understanding of how political systems work, and that calls for greater integration of the work of economists and political scientists in the design of regulatory institutions. Finally, building new institutions requires time. Regulatory institutions in developing countries need support. For development partners, this means a greater concentration on countries where preconditions for success are relatively tangible.

Supplementary Material

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

Funding

We gratefully acknowledge the financial support of the Public-Private Partnerships in Infrastructure Program of the World Bank Institute.

Appendix: Data Summary Statistics for Developing and Developed Countries

Variable	Designation	Number of observations		Median		Standard deviation		Minimum		Maximum	
		Developing countries	Developed countries	Developing countries	Developed countries	Developing countries	Developed countries	Developing countries	Developed countries	Developing countries	Developed countries
Regulatory performance											
ml	Mainline coverage	435	345	3.76	47.49	4.96	10.87	0.11	14.52	22.36	73.56
cel	Cellular subscription	431	344	0.01	2.55	2.09	13.51	0.00	0.00	15.96	63.37
eff	Mainlines per employee	424	345	53.06	166.08	58.85	57.53	7.78	43.48	371.16	358.76
p_res	Price of monthly subscription to fixed-line service	256	252	4.44	4.70	4.23	4.70	0.00	5.60	21.29	26.27
p_cell	Price of cellular service	324	192	0.00	1.40	0.53	0.86	0.00	0.00	2.24	4.95
Local accountability											
reg	Regulatory governance index	435	345	0.00	0	4.60	3.11	0.00	0.00	13.50	8.00
Global accountability											
corruption	Corruption	435	345	5.00	8.33	1.43	1.37	1.66	3.33	10.00	10.00
bureau	Bureaucracy	420	345	5.00	10.00	1.86	1.33	1.66	4.50	10.00	10.00
law	Law and order	435	345	5.00	10.00	2.06	1.11	0.00	5.00	10.00	10.00
expropri	Expropriation	420	345	7.35	10.00	2.00	0.66	2.00	4.60	10.00	10.00
currency	Currency risk	435	345	6.00	9.00	1.98	1.16	1.00	4.00	10.00	10.00
institutional	Institutional environment index	435	345	28.66	47.00	7.10	3.99	8.00	25.26	41.16	50 .00
checks	Checks and balances	423	345	3.00	4.00	2.06	1.62	1.00	2.00	18.00	16.00
Other variables											
priva	Privatization	435	345	0.00	0.00	0.32	0.48	0.00	0.00	1.00	1.00
comp_fix	Competition in fixed	435	345	0.00	0.00	0.29	0.42	0.00	0.00	1.00	1.00
comp_cell	Competition in cellular	435	345	1.00	0.00	1.10	0.47	0.00	0.00	3.00	1.00
rural	Rural population	435	345	49.82	24.70	20.95	12.73	10.95	2.95	90.31	62.84
density	Population density	435	345	48.07	94.59	79.39	119.50	5.38	2.01	330.34	466.49

Source: Authors' analysis based on data described in text and in Gasmi, Noumba, and Recuero Virto (2006).

References

Ackerman, J.M. 2005. "Social Accountability in the Public Sector: A Conceptual Discussion." Social Development Paper 82. World Bank, Washington, D.C.

Arellano, M., and S.R. Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies* 58(2):277–97.

Arellano, M., and O. Bover. 1995. "Another Look at the Instrumental Variable Estimation of Error-Component Models." *Journal of Econometrics* 68(1):29–51.

Beck, N., and J. Katz. 2004. "Time-Series-Cross-Section Issues: Dynamics." Paper presented at the 2004 Annual Meeting of the Society for Political Methodology. Stanford University, Palo Alto, Calif., July 29–31.

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

———. 1999. *GMM Estimation with Persistent Panel Data: An Application to Production Functions.* Institute for Fiscal Studies Working Paper Series W99/4. London: Institute for Fiscal Studies.

Brown, A.C., J. Stern, B. Tenenbaum, and D. Gencer. 2006. *Handbook for Evaluating Infrastructure Regulatory Systems.* Washington, D.C.: World Bank.

Correa, P., M. Melo, B. Mueller, and C. Pereira. 2008. "Regulatory Governance in Brazilian Infrastructure Industry." *The Quarterly Review of Economics and Finance* 48(2):202–16.

Cubbin, J., and J. Stern. 2005a. "Regulatory Effectiveness and the Empirical Impact of Variations in Regulatory Governance: Electricity Industry Capacity and Efficiency in Developing Countries." Policy Research Working Paper 3535. World Bank, Washington, D.C.

———. 2005b. "Regulatory Effectiveness: The Impact of Regulation and Regulatory Governance Arrangements on Electricity Industry Outcomes." Policy Research Working Paper 3536. World Bank, Washington, D.C.

Estache, A., and D. Martimort. 1999. "Politics, Transactions Costs, and the Design of Regulatory Institutions." Policy Research Working Paper 2073. World Bank, Washington, D.C.

Gasmi, F., P. Noumba, and L. Recuero Virto. 2006. "Political Accountability and Regulatory Performance in Infrastructure Industries: An Empirical Analysis." Policy Research Working Paper 4101. World Bank, Washington, D.C.

Gasmi, F., and L. Recuero Virto. Forthcoming. "The Determinants of Reforms and Their Impact on Telecommunications Deployment in Developing Countries." *Journal of Development Economics.*

Gutierrez, L. H. 2003a. "The Effect of Endogenous Regulation on Telecommunications Expansion and Efficiency in Latin America." *Journal of Regulatory Economics* 23(3):257–86.

———. 2003b. "Regulatory Governance in the Latin American Telecommunications Sector." *Utilities Policies* 11(4):225–40.

Heller, W.B., and M.D. McCubbins. 1996. "Politics, Institutions, and Outcomes: Electricity Regulation in Argentina and Chile." *Journal of Policy Reform* 1(4):357–87.

Holder, S., and J. Stern. 1999. "Regulatory Governance: Criteria for Assessing the Performance of Regulatory Systems." *Utilities Policies* 8(1):35–50.

Holtz-Eakin, D., W. Newey, and H.S. Rosen. 1988. "Estimating Vector Autoregressions with Panel Data." *Econometrica* 56(6):1371–95.

Laffont, J.J. 2005. *Regulation and Development.* Cambridge: Cambridge University Press.

Levy, B., and P.T. Spiller. 1994. "The Institutional Foundations of Regulatory Commitment: A Comparative Analysis of Telecommunications Regulation." *Journal of Law, Economics and Organization* 10(2):201–46.

McCubbins, M.D., R.G. Noll, and B.R. Weingast. 1987. "Administrative Procedures as Instruments of Political Control." *Journal of Law, Economics and Organization* 3(2):243–77.

McCubbins, M.D., and T. Schwartz. 1984. "Congressional Oversight Overlooked: Police Patrol vs Fire Alarms." *American Journal of Political Science* 28(2):165–79.

North, D.C. 2000. "Institutions and the Performance of Economies over Time." *Paper presented at the Second Annual Global Development Conference*, Tokyo December 10–13.

Rodrik, D., A. Subramanian, and F. Trebbi. 2004. "Institution Rule: The Primacy of Institutions Over Geography and Integration in Economic Development." *Journal of Economic Growth* 9(2):131–65.

Ros, A.J. 1999. "Does Ownership and Competition Matter?: The Effects of Telecommunications Reform on Network Expansion and Efficiency." *Journal of Regulatory Economics* 15(1):65–92.

———. 2003. "The Impact of the Regulatory Process and Price Cap Regulation in Latin American Telecommunications Markets." *Review of Network Economics* 2:270–86.

Spiller, P.T., and M. Tommasi. 2003. "The Institutions of Regulation: An Application to Public Utilities." In S. Majumdar, I. Vogelsang, and M. Cave eds., *Handbook of Telecommunications Economics: Technology Evolution and the Internet*, Vol. 2., Amsterdam: North-Holland.

THE WORLD BANK
1818 H Street, NW
Washington, DC 20433, USA
World Wide Web: http://www.worldbank.org/
E-mail: wber@worldbank.org

ISBN-13: 978-0-19-957453-7
9 780199 574537