THE WORLD BANK ECONOMIC REVIEW

Volume 15 • 2001 • Number 2

OXFORD

ISSN 0258-6770

THE WORLD BANK ECONOMIC REVIEW

The World Bank Economic Review is a professional journal for the dissemination of World Bank–sponsored and outside research that may inform policy analyses and choices. It is directed to an international readership among economists and social scientists in government, business, and international agencies, as well as in universities and development research institutions. The *Review* emphasizes policy relevance and operational aspects of economics, rather than primarily theoretical and methodological issues. It is intended for readers familiar with economic theory and analysis but not necessarily proficient in advanced mathematical or econometric techniques. Articles will illustrate how professional research can shed light on policy choices. Inconsistency with Bank policy will not be grounds for rejection of an article.

Articles will be drawn from work conducted by World Bank staff and consultants and from papers submitted by outside researchers. Before being accepted for publication, all articles will be reviewed by two referees who are not members of the Bank's staff and one World Bank staff member. Articles must also be recommended by a member of the Editorial Board. Non-Bank contributors are requested to submit a proposal of not more than two pages in length to the Editor or a member of the Editorial Board before sending in their paper.

Comments or brief notes responding to *Review* articles are welcome and will be considered for publication to the extent that space permits. Please direct all editorial correspondence to the Editor, *The World Bank Economic Review*, The World Bank, 1818 H Street, Washington, DC 20433, USA, or wber@worldbank.org.

For more information, please visit the Web sites of the *Economic Review* at www.wber.oupjournals.org, the World Bank at www.worldbank.org, and Oxford University Press at www.oup-usa.org.

THE WORLD BANK ECONOMIC REVIEW

Volume 15 • 2001 • Number 2

WHAT HAVE WE LEARNED FROM A DECADE OF EMPIRICAL RESEARCH ON GROWTH?

The World Bank Economic Review (ISSN 0258-6770) is published three times a year by Oxford University Press, 2001 Evans Road, Cary, NC 27513-2009 for The International Bank for Reconstruction and Development / THE WORLD BANK. Communications regarding original articles and editorial management should be addressed to The Editor, *The World Bank Economic Review*, 66, avenue d'Iéna, 75116 Paris, France. E-mail: wber@worldbank.org.

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide.

SUBSCRIPTIONS: Subscription is on a yearly basis. The annual rates are US$40 (£30 in UK and Europe) for individuals; US$90 (£63 in UK and Europe) for academic libraries; US$110 (£75 in UK and Europe) for corporations. Single issues are available for US$17 (£13 in UK and Europe) for individuals; US$38 (£26 in UK and Europe) for academic libraries; US$46 (£31 in UK and Europe) for corporations. All prices include postage. Individual rates are applicable only when a subscription is for individual use and are not available if delivery is made to a corporate address. *Subscriptions are provided free of charge to non-OECD countries.* All subscription requests, single issue and back issue orders, changes of address, and claims for missing issues should be sent to:

North America: Oxford University Press, Journals Customer Service, 2001 Evans Road, Cary, NC 27513-2009, USA. Toll-free in the USA and Canada: 800-852-7323, or 919-677-0977. Fax: 919-677-1714. E-mail: jnlorders@oup-usa.org.

Elsewhere: Oxford University Press, Journals Customer Service, Great Clarendon Street, Oxford OX2 6DP, UK. Tel: +44 1865 267907. Fax: +44 1865 267485. E-mail: jnl.orders@oup.co.uk.

ADVERTISING: Helen Pearson, Oxford Journals Advertising, P.O. Box 347, Abingdon SO, OX14 1GJ, UK. Tel/Fax: +44 1235 201904. E-mail: helen@oxfordads.com.

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 Sciences Citation Index*. The microform edition is available through UMI, 300 North Zeeb Road, Ann Arbor, MI 48106, USA.

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).

POSTAL INFORMATION: *The World Bank Economic Review* (ISSN 0258-6770) is published three times a year by Oxford University Press, 2001 Evans Road, Cary, NC 27513-2009. Send address changes to *The World Bank Economic Review*, Journals Customer Service Department, Oxford University Press, 2001 Evans Road, Cary, NC 27513-2009.

THE WORLD BANK ECONOMIC REVIEW, VOL. 15, NO. 2 177–219

What have we learned from a decade of empirical research on growth?

It's Not Factor Accumulation: Stylized Facts and Growth Models

William Easterly and Ross Levine

The article documents five stylized facts of economic growth. (1) The "residual" (total factor productivity, TFP) rather than factor accumulation accounts for most of the income and growth differences across countries. (2) Income diverges over the long run. (3) Factor accumulation is persistent while growth is not, and the growth path of countries exhibits remarkable variation. (4) Economic activity is highly concentrated, with all factors of production flowing to the richest areas. (5) National policies are closely associated with long-run economic growth rates. These facts do not support models with diminishing returns, constant returns to scale, some fixed factor of production, or an emphasis on factor accumulation. However, empirical work does not yet decisively distinguish among the different theoretical conceptions of TFP growth. Economists should devote more effort toward modeling and quantifying TFP.

The central problem in understanding economic development and growth is not understanding the process by which an economy raises its savings rate and increases the rate of physical capital accumulation.[1] Although many development practitioners and researchers continue to target capital accumulation as the driv-

William Easterly is senior advisor, Development Research Group, at the World Bank. His e-mail address is *weasterly@worldbank.org*. Ross Levine is with the University of Minnesota. His e-mail address is *rlevine@csom.umn.edu*. The authors are grateful to Lant Pritchett, who shaped the paper, gave comments, and provided many of the "stylized facts." They also thank François Bourguignon, Ashok Dhareshwar, Robert G. King, Michael Kremer, Peter Klenow, Paul Romer, Xavier Sala-i-Martin, Robert Solow, Albert Zeufack, two anonymous referees, and students and faculty at the Economics Education Research Consortium program in Kiev, Ukraine, Harvard University's Kennedy School of Government, and Johns Hopkins School of Advanced International Studies for useful comments. An earlier version of this article was presented at the World Bank conference "What Have We Learned from a Decade of Empirical Research on Growth?" held on 26 February 2001.

1. This is a reversal and slight rewording of Arthur Lewis's (1954, p. 155) famous quote, "The central problem in the theory of economic development is to understand the process by which a community which was previously saving and investing 4 or 5 percent of its national income or less, converts itself into an economy where voluntary saving is running at about 12 to 15 percent of national income or more. This is the central problem because the central fact of development is rapid capital accumulation (including knowledge and skills with capital)." Though Lewis recognizes the importance of knowledge and skills and later in his book highlights the importance of institutions, many development economists who followed Lewis adopted the more limited focus on savings and physical capital accumulation.

ing force in economic growth,[2] "something else" besides capital accumulation is critical for understanding differences in economic growth and income across countries. This conclusion is based on evidence on the sources of economic growth, the patterns of economic growth, the patterns of factor flows, and the impact of national policies on economic growth.

This study does not argue that factor accumulation is unimportant in general or deny that it is critically important for some countries at specific junctures. As Robert Solow noted in 1956, economists construct models to reproduce crucial empirical regularities and then use these models to interpret economic events and make policy recommendations. This article documents important empirical regularities about economic growth with the hope of highlighting productive directions for future research and improving public policy.

I. Something Else

A growing body of research suggests that, even after physical and human capital accumulation are accounted for, something else accounts for the bulk of cross-country differences in the level and growth rate of gross domestic product (GDP) per capita. Economists typically refer to the something else as *total factor productivity* (TFP). This article follows that convention.

Different theories offer very different conceptions of TFP. These range from changes in technology (the instructions for producing goods and services) to the role of externalities, changes in the sector composition of production, and the adoption of lower-cost production methods. Evidence that confidently assesses how well these conceptions of TFP explain economic growth is lacking. Economists need to provide much more shape and substance to the amorphous term TFP, distinguishing empirically among these different theories.

This article examines five stylized facts that illuminate TFP and its determinants to enable more precise modeling of long-run economic growth and the design of appropriate policies.

2. Academic researchers in the 1990s started a "neoclassical revival" (in the words of Klenow and Rodriguez-Clare 1997b). The classic works in the academic literature's stress on factor accumulation were Mankiw, Romer, and Weil (1992); Barro and others (1995); Mankiw (1995); and Young (1995). The summary of the Global Development Network conference in Prague in June 2000, representing many international organizations and development research institutes, says "physical capital accumulation was found to be the dominant source of growth both within and across regions. Total factor productivity growth (TFPG) was not as important as was previously believed" (*www.gdnet.org/pdfs/GRPPragueMtgReport.pdf*). A leading development textbook (Todaro 2000) says that an increase in investment is "a necessary condition" for economic takeoff. The development textbook of Ray (1998, p. 54) refers to investment and saving as "the foundations of *all* models of economic growth." Many development practitioners also stress investment. For example, the International Monetary Fund (Hadjimichael and others 1996, p. 1) argues, "The adjustment experience of sub-Saharan Africa has demonstrated that to achieve gains in real per capita GDP an expansion in private saving and investment is key." The Bank for International Settlements (1996, p. 50) concludes, "Recent experience has underlined the central importance of national saving and investment rates in promoting growth." And the International Labor Organization (1995, p. 12) argues that "policies to raise the rate of investment . . .

- *Stylized fact 1. Factor accumulation does not account for the bulk of cross-country differences in the level or growth rate of* GDP *per capital*; something else—TFP—does. In the search for the secrets of long-run economic growth, a high priority should go to rigorously defining TFP, empirically dissecting it, and identifying the policies and institutions most conducive to its growth.
- *Stylized fact 2. There are huge and growing differences in* GDP *per capita; divergence—not conditional convergence—is the big story.* An emphasis on TFP growth with increasing returns to technology is more consistent with divergence than are models of factor accumulation with decreasing returns, no scale economies, and some fixed factor of production. Over the past two centuries, the big story has been the widening difference between the richest and the poorest countries. Moreover, the growth rates of the rich are not slowing, and returns to capital are not falling. Just as business cycles look like little wiggles around the big story when viewed over a long horizon, understanding slow, intermittent conditional convergence seems less intriguing than uncovering why the United States has enjoyed steady growth for 200 years while much of world still lives in poverty.
- *Stylized fact 3. Growth is not persistent over time, but capital accumulation is.* Some countries take off, others experience peaks and valleys, a few grow steadily, and some have never grown. Changes in factor accumulation do not closely track changes in economic growth. This finding is consistent across very different frequencies of data. Tangentially, but critically, this stylized fact also suggests that models of steady-state growth, whether based on capital externalities or technological spillovers, will not capture the experiences of many countries. While steady-state growth models may fit U.S. experience over the past 200 years, these models will not fit the experiences of Argentina, the Republic of Korea, Thailand, or Venezuela very well. In contrast, models of multiple equilibria do not fit the U.S. data very well. Thus models tend to be country-specific rather than general theories. Meanwhile, empirical work is still going on to explain why the United States is different, how

are critical for raising the rate of growth and employment in an economy." Finally "additional investment is the answer—or part of the answer—to most policy problems in the economic and social arena" (United Nations 1996, p. 8). Similarly, the World Bank (1993, p. 191) states that in East Asia, "accumulation of productive assets is the foundation of economic growth." World Bank (1995, p. 10, 23) promises that in Latin America "enhancing saving and investment by 8 percentage points of GDP would raise the annual growth figure by around 2 percentage points." The World Bank (2000a, p. 10) says the saving rate of the typical African country "is far below what is needed to sustain a long-term boost in economic performance." The World Bank (2000c, p. 1) says that southeastern Europe can seize trade opportunities only if "domestic and foreign entrepreneurs increase their investment dramatically." For more citations, see Easterly (1999a) and King and Levine (1994). Although common, the stress on capital accumulation is far from universal among development practitioners and researchers. For example, the World Bank (2000b, p. 4) report on East Asia's recovery suggests that "future growth hinges less on increasing physical capital accumulation and more on raising the productivity growth of all factors." Collier, Dollar, and Stern (2000) stress policies, incentives, institutions, and exogenous factors as the main drivers in growth with little mention of investment, as does *World Development Report 2000/2001* (World Bank 2000/2001, pp. 49–52).

Argentina can go from being like the United States early in this century to the struggling middle-income country it is today, and how Korea or Thailand can go from being like Somalia to having thriving economies.

- *Stylized fact 4. All factors of production flow to the same places, suggesting important externalities.* Although this has been noted and modeled by Lucas (1988), Kremer (1993), and others, this article further demonstrates the pervasive tendency for all factors of production, including physical and human capital, to bunch together. As a consequence, economic activity is highly concentrated. This tendency holds whether considering the world, countries, regions, states, ethnic groups, or cities. Thus the something else that accounts for the bulk of differences in growth across these units needs to be fleshed out and given a prominent position in theories and policy recommendations.
- *Stylized fact 5. National policies influence long-run growth.* In models with zero productivity growth, diminishing returns to factors of production, and some fixed factor, national policies that boost physical or human capital accumulation have only a transitional effect on growth. In models that emphasize TFP growth, national policies that enhance the efficiency of capital and labor or alter the endogenous rate of technological change can boost productivity growth and accelerate long-run economic growth. Thus the finding that policy influences growth is consistent with theories that emphasize productivity growth and technological externalities and cast increasing doubt on theories that focus excessively on factor accumulation.

Although many economists have examined TFP growth and assessed growth models, this article makes several new contributions. Besides conducting traditional growth accounting with new Penn-World Table 5.6 capital stock data, this article fully exploits the panel nature of the data. Using an international cross-section of countries, it addresses two questions:

- What accounts for cross-country growth differences?
- What accounts for growth differences over time?

Overwhelmingly the answer is TFP, not factor accumulation.

The article also examines differences in the level of GDP per worker across countries. It updates Denison's (1962) original level accounting study and extends Mankiw, Romer, and Weil's (1992) study by allowing technology to differ across countries and by assessing the importance of country-specific effects. Unlike Mankiw, Romer, and Weil (1992), it finds that large differences in TFP account for the bulk of cross-country differences in income per capita, even controlling for country-specific effects.

The article also compiles new information documenting massive divergence in the level of income per capita across countries. Although many studies base their modeling strategies on the U.S. experience of steady long-run growth (see, for example, Jones 1995a, 1995b; and Rebelo and Stokey 1995), the U.S. experience is the exception. In much of the world miracles and disasters and changing long-run growth rates are the rule, not stable long-run growth rates. Finally, the article presents abundant new evidence on the concentration of economic activity, drawing on cross-country information, county-level data for the United States, developing country studies, and information on the international flow of capital, labor, and human capital to demonstrate the geographic concentration of activity and relate this to models of economic growth. The overwhelming concentration of economic activity is consistent with some theories of economic growth and inconsistent with others. Though individual countries at specific points in their development fit different models of growth, the big picture emerging from cross-country growth comparisons is that creating the incentives for productive factor accumulation is more important for growth than factor accumulation itself.

II. Stylized Fact 1. It's Not Factor Accumulation, It's TFP

Although physical and human capital accumulation may play key roles in igniting and accounting for economic progress in some countries, something else—TFP—accounts for the bulk of cross-country differences in the level and growth of GDP per capita in a broad cross-section of countries. The empirical importance of TFP has motivated economists to develop models of TFP. These focus variously on technological change (Aghion and Howitt 1998; Grossman and Helpman 1991; Romer 1990); impediments to adopting new technologies (Parente and Prescott 1996); externalities (Romer 1986; Lucas 1988); sectoral development (Kongsamut, Rebelo, and Xie 1997); or cost reductions (Harberger 1998).

This section briefly presents evidence on factor accumulation and growth and discusses the implications for models and policy. It considers three questions. First, what part of a country's growth rate is accounted for by factor accumulation and TFP growth? Looking at the sources of growth in individual countries over time helps answer this question. Second, what part of cross-country differences in economic growth rates is accounted for by cross-country differences in growth rates of factor accumulation and TFP? Third, what part of the intertemporal difference in economic growth rates is accounted for by time-series differences in growth rates of factor accumulation and TFP? Traditional growth accounting forms the basis for answering these questions.

Growth Accounting

The organizing principle of growth accounting is the Cobb-Douglas aggregate production function:

$$y = Ak^{\alpha}(n^{1-\alpha}), \tag{1}$$

where y is national output per person,[3] A is technological progress, k is the physical capital stock per person, n is the number of units of labor input per person

3. We switch between output per worker and output per person depending on data availability and what's appropriate for each usage.

(reflecting work patterns, human capital, and the like), and α is a production function parameter (equal to the share of capital income in national output under perfect competition).

Output growth is then divided into components attributable to changes in the factors of production. Rewriting equation 1 in growth rates:

$$(\Delta y / y) = (\Delta A / A) + \alpha(\Delta k / k) + (1 - \alpha)(\Delta n / n). \tag{2}$$

Consider a hypothetical country with a growth rate of output per person of 2 percent, growth in capital per capita of 3 percent, growth in human capital of 0, and capital's share of national income of 40 percent ($\alpha = 0.4$). In this example, TFP growth is 0.8 percent, and therefore, TFP-growth accounts for 40 percent (0.8/2) of output growth in this country.

DETAILED GROWTH ACCOUNTING. Many researchers conduct detailed growth accounting exercises of one or a few countries, using disaggregated data on capital, labor, human capital, and capital shares of income. Early, detailed growth accounting exercises of a few countries by Solow (1957) and Denison (1962, 1967) found that the rate of capital accumulation per person accounted for between one-eighth and one-fourth of GDP growth rates in the United States and other industrialized countries, whereas TFP-growth accounted for more than half of GDP growth in many countries.

Subsequent studies showed that it is important to account for changes in the quality of labor and capital (see papers in Jorgenson 1995). For example, if growth accountants fail to consider improvements in the quality of labor inputs due to improved education and health, they would assign these improvements to TFP growth. Unmeasured improvements in physical capital would similarly be inappropriately assigned to TFP. Nonetheless, to the extent that TFP includes quality improvements in capital, a finding that TFP explains a substantial amount of economic growth will properly focus attention on productivity rather than on factor accumulation itself.

Later detailed growth accounting exercises for a few countries incorporated estimates of such changes in the quality of human and physical capital (table 1).[4] These studies also find that TFP growth tends to account for a large component of the growth of output. Christenson, Cummings, and Jorgenson (1980) do this for a few Organisation for Economic Co-operation and Development (OECD) countries, albeit prior to the productivity growth slowdown. Dougherty (1991) does the exercise for some OECD countries including the slow productivity growth period. Elias (1990) conducts a rigorous growth accounting study for seven Latin American countries. Young (1995) focuses on fast growing East Asian countries. Although there are large cross-country variations in the fraction of growth accounted for by TFP growth, some general patterns emerge. TFP growth accounts

4. We use the summary in Barro and Sala-i-Martin (1995, pp. 380–81).

for about half of output growth in OECD countries. Variation is greater among Latin American countries, with an average of 30 percent. Young (1995) argues that factor accumulation was a key component of the growth miracle in some East Asian economies.

These detailed growth accounting exercises may seriously underestimate the influence of TFP growth on growth in output per worker as emphasized by Klenow and Rodriguez-Clare (1997a). The studies summarized in table 1 examine output growth. If the analysis is adjusted to focus on output per worker, TFP growth accounts for a much larger share of output per worker growth than for the output growth figures in table 1. In an extension of Young (1995), Klenow and Rodriguez-Clare (1997a) show that factor accumulation plays the crucial role

TABLE 1. Selected Growth Accounting Results for Individual Countries (percent)

Economy	Share of capital in national output	GDP growth	Share contributed by: Capital	Labor	TFP
OECD 1947–73					
France	.40	5.40	41	4	55
Germany	.39	6.61	41	3	56
Italy	.39	5.30	34	2	64
Japan	.39	9.50	35	23	42
United Kingdom	.38	3.70	47	1	52
United States	.40	4.00	43	24	33
OECD 1960–90					
France	.42	3.50	58	1	41
Germany	.40	3.20	59	–8	49
Italy	.38	4.10	49	3	48
Japan	.42	6.81	57	14	29
United Kingdom	.39	2.49	52	–4	52
United States	.41	3.10	45	42	13
Latin America 1940–80					
Argentina	.54	3.60	43	26	31
Brazil	.45	6.40	51	20	29
Chile	.52	3.80	34	26	40
Mexico	.69	6.30	40	23	37
Venezuela	.55	5.20	57	34	9
East Asia 1966–90					
Hong Kong, China	.37	7.30	42	28	30
Singapore	.53	8.50	73	32	–5
Korea, Rep. of	.32	10.32	46	42	12
Taiwan, China	0.29	9.10	40	40	20

Source: For OECD, Christenson, Cummings, and Jorgenson (1980) and Dougherty (1991); for Latin America, Elias (1990); for East Asia, Young (1995).

only in Singapore (a small city-state) but in none of the other East Asian miracle economies. In addition, the share attributed to capital accumulation may be exaggerated because it does not take into account how much TFP growth induces capital accumulation (Barro and Sala-i-Martin 1995, p. 352.)

In sum, although there are cases in which factor accumulation is closely tied to economic success, detailed growth accounting examinations suggest that TFP growth frequently accounts for the bulk of growth in output per worker.

AGGREGATE GROWTH ACCOUNTING. There are also aggregate growth accounting exercises of a large cross-section of countries that use a conglomerate measure of capital and an average value of the capital share parameter from microeconomic studies. King and Levine (1994) and Nehru and Dhareshwar (1994) make some initial estimates of the capital stocks of countries in 1950. They then use aggregate investment data and assumptions about depreciation rates to compute capital stocks in later years for over 100 countries. The importance of the estimate of the *initial* capital stock diminishes over time due to depreciation.

This study uses the new Penn-World Table (PWT) 5.6 capital stock data, based on disaggregated investment and depreciation statistics for 64 countries. Though these data exist for a smaller number of countries, they suffer from fewer aggregation and measurement problems than the aggregate growth accounting exercises using less precise data.[5]

5. The Penn World Tables document the construction of these data. Capital stock figures were also constructed for more countries using aggregate investment figures. For some countries, the data start in 1951. These data use real investment in 1985 prices and real GDP per capita (chain index) in constant 1985 prices. A perpetual inventory method was used to compute capital stocks. Specifically, let $K(t)$ equal the real capital stock in period t. Let $I(t)$ equal the real investment rate in period t. Let d equal the depreciation rate, assumed to be .07. Thus, the capital accumulation equation states that $K(t+1) = (1 - d)\, K(t) + I(t)$. To compute the capital per worker ratio, divide $K(t)$ by $L(t)$, where $L(t)$ is the working age population in period t as defined in the Penn World Tables. To compute the capital-output ratio, divide $K(t)$ by $Y(t)$, where $Y(t)$ is real GDP per capita in period t. To make an initial estimate of the capital stock, we make the assume that the country is at its steady-state capital-output ratio. Thus in terms of steady-state value, let $k = K/Y$, let g = the growth rate of real output, let $i = I/Y$. Then, from the capital accumulation equation plus the assumption that the country is at its steady-state, $k = i/(g + d)$. Thus, with reasonable estimates of the steady-state values of i, g, and d, a reasonable estimate of k can be computed. The Penn World Tables have data on output back to 1950. Thus, the initial capital stock estimate can be computed as kY(initial). To make the initial estimate of k, the steady state capital output ratio, set $d = .07$. The steady-state growth rate g is computed as a weighted average of the countries average growth rate during the first ten years for which we have output and investment data and the world growth rate , computed as 0.0423. Based on Easterly and others (1993), the world growth rate is given a weight of 0.75 and the country growth rate 0.25 in computing an estimate of the steady-state growth rate for each country. Then i can be computed as the average investment rate during the first ten years for which there are data. Thus, with values for d, g, and i for each country, k can be estimated for each country. Average real output value between 1950–52 is used as an estimate of initial output, Y(initial), to reduce the influence of business cycles in estimating Y(initial). Thus the capital stock in 1951 is given as Y(initial)k. If output and investment data do not start until 1960, everything is moved up one decade for that country. Given depreciation, the guess at the initial capital stock becomes relatively unimportant decades later.

The aggregate growth accounting results for a broad selection of countries also emphasize TFP's role in economic growth. There is enormous cross-country variation in the fraction of growth accounted for by capital and TFP growth. In the average country, considering only physical capital accumulation, TFP growth accounts for about 60 percent of growth in output per worker using the PWT 5.6 capital data and setting the share of capital in national output (α) at .4, which is consistent with individual country studies. Other measures of the capital stock from King and Levine (1994) and Nehru and Dhareshwar (1993) yield similar results.

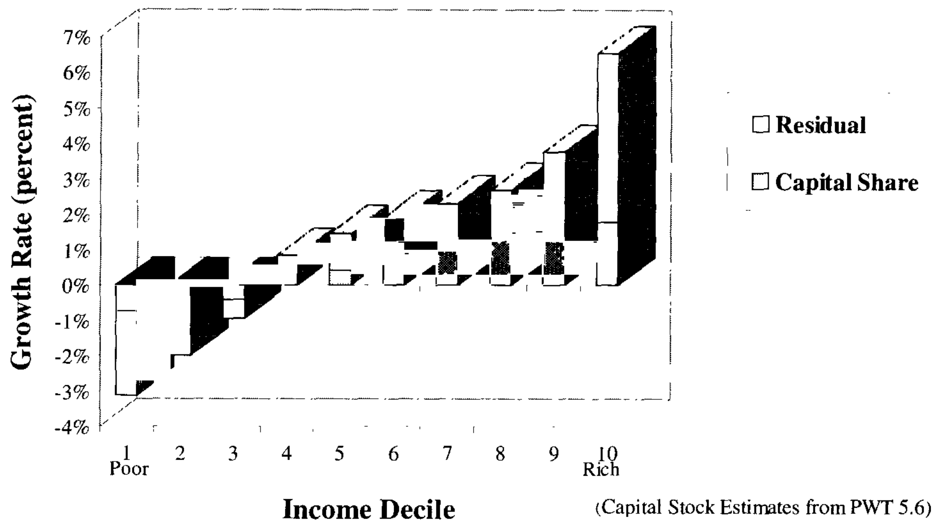
Aggregate growth accounting results are illustrated in figure 1 using data from PWT 5.6 for 1980–92. Countries are grouped by decile based on output per capita growth, from the slowest growing (group 1) to the fastest. Capital growth generally accounts for less than half of output growth, and the share of growth accounted for by TFP growth is frequently larger in the faster growing countries. There are large differences across countries in the relationship between capital accumulation and growth. For example, Costa Rica, Ecuador, Peru, and Syria all saw real per capita GDP fall by more than 1 percent a year, while real per capita capital stocks grew by more than 1 percent a year and educational attainment was rising. Clearly, these factor injections were not being used productivity. Albeit unrepresentative, these cases illustrate the shortcoming of focusing too heavily on factor accumulation.[6]

Incorporating estimates of human capital accumulation into these aggregate growth accounting exercises does not materially alter the findings. In the average country, TFP growth still accounts for more than half of growth in output per worker. Moreover, the data suggest a weak—and sometimes inverse—relationship between improvements in educational attainment of the labor force and growth of output per worker growth. Benhabib and Spiegel (1994) and Pritchett (2001), using cross-country data on economic growth rates, show that increases in human capital resulting from improvements in educational attainment have not positively affected the growth in output per worker (perhaps because of a mismatch between education and the skills needed for activities that generate social returns).

There is disagreement, however. Krueger and Lindahl (1999) argue that measurement error accounts for the failure to find a relationship between growth per capita and human capital accumulation. Hanushek and Kimko (2000) find that the quality of education is strongly linked with economic growth. Klenow (1998) demonstrates that models that highlight the role of ideas and productivity growth do a much better job of matching the data than models that focus on the accumulation of human capital. More work is needed on the relationship between education and economic development.

6. It may be that the conventional measure of investment effort is a cost-based measure that does not translate necessarily into increasing the value of the capital stock. Pritchett (1999) makes this point, especially—but not only—with regard to public investment.

FIGURE 1. Growth Accounting: Growth Rates by Decile



Variance Decomposition

Although traditional growth accounting measures that part of a country's growth rate that may be attributable to factor accumulation, this study uses variance decomposition to construct indicators of that part of *cross-country* differences in economic growth rates accounted for by *cross-country* differences in TFP and factor growth (Jones 1997). Assuming that $\alpha = .4$, the following holds for the cross-section of countries:

$$\text{(3)} \qquad VAR(\Delta y/y) = VAR(\Delta TFP/TFP) + (0.4)^2\{VAR(\Delta k/k)\} + 2(0.4)\{COV(\Delta TFP/TFP, \Delta k/k)\}.$$

Decomposing the sources of growth across countries using different data sets shows that cross-country variations in TFP growth account for more that 60 percent of output growth using alternative data sets (table 2). The cross-country variation in physical capital alone—excluding the covariance with TFP growth—never accounts for more than 25 percent of the cross-country variation in per capita GDP growth.

Researchers also incorporate human capital accumulation into decomposition exercises, rewriting the variance-decomposition equations as:

$$\text{(4)} \qquad VAR(\Delta y/y) = VAR(\Delta TFP/TFP) + (0.7)^2\{VAR(\Delta f/f)\} + 2(0.7)\{COV(\Delta TFP/TFP, \Delta f/f)\},$$

where $\Delta f/f$ is factor accumulation per worker, defined as the average growth rate of physical capital per worker and educational attainment per worker. Specifically, $\Delta f/f = (\Delta k/k + \Delta h/h)/2$, where h is educational attainment per worker.[7]

7. Again, different authors use different weights, though this tends not to change the basic findings.

TABLE 2. Variance Decomposition

	Contribution of		
	TFP growth	Capital growth	Covariance of capital growth and TFP growth
Without human capital[a]			
1960–92	0.58	0.41	0.01
1980–92	0.65	0.21	0.13
With human capital			
1960–92[b]	0.94	0.52	−0.45
1980–87[c]	0.68	0.20	0.12

[a]Sixty non–oil-exporting countries.
[b]Forty-four countries.
[c]Fifty countries.
Source: Authors's calculations based on the PWT 5.6 capital stock series and Benhabib and Spiegel's (1994) estimates of human capital growth.

Incorporating human capital does not alter the basic result. TFP growth differentials account for the bulk of cross-country growth differences. Klenow and Rodriguez-Clare (1997b) estimate that differences in TFP growth account for about 90 percent of the variation in growth rates of output per worker for a sample of 98 countries during 1960–95 after accounting for human capital accumulation (based on schooling and job experience). The use of the PWT 5.6 capital stock series and estimates of the growth rate of human capital from Benhabib and Spiegel (1994) also shows that differences in TFP growth account for about 90 percent of cross-country differences in real per capita GDP growth during 1960–92. Thus differences in TFP growth—rather than in factor accumulation rates—seem like the natural place to start in explaining cross-country differences in long-run growth rates.

Growth accounting has several limitations. It is a mechanical procedure, and using it to elucidate a causal story is dangerous. For example, in Solow's (1956) model, if technological progress (A) grows at the exogenously given steady-state rate x, then y and k grow at the steady-state rate x, too. Growth accounting will, therefore, attribute αx of output growth to capital growth, yielding the conclusion that α times 100 percent of growth is due to physical capital accumulation. Also, growth accounting does not test the statistical significance of the relationship between output growth and capital accumulation. (The temporal—Granger-causal—relationships between growth and savings, investment, and education are discussed later.)

LEVEL ACCOUNTING AND THE K/Y RATIO. Hall and Jones (1999) recently reexamined the level accounting question, asking what part of cross-country differences in income per capita is accounted for by differences in physical capital per capita. They find that productivity differences across countries account for the bulk of cross-country differences in output per worker. This study addresses this question using the traditional Denison (1962) approach and a modified Mankiw, Romer, and Weil (1992) approach.

To conduct *Denison-level accounting*, take the ratio of two national incomes of output per person from equation 1:

$$[y_i / y_i] = [A_i / A_j][k_i / k_j]^{\alpha}[n_i / n_j]^{1-\alpha}. \quad (5)$$

Given data on the factors of production, cross-country differences in TFP can be measured by:

$$[A_i / A_j] = [y_i / y_j]/\{[k_i / k_j]^{\alpha}[n_i / n_j]^{1-\alpha}\}. \quad (6)$$

The fraction of differences in national output levels due to capital equals the ratio, ϕ_{ki}.

$$\phi_{ki} = \alpha \log(k_i / k_j) / \log(y_i / y_j). \quad (7)$$

Equation 7 can be rewritten as:

$$\phi_{ki} = \alpha + \alpha \log(k_i / k_j) / \log(y_i / y_j), \quad (8)$$

because $\log(k_i/k_j) = \log(\kappa_i/\kappa_j) - \log(y_{yi}/y_j)$, letting $\kappa = k/y$. This allows measurement of the contribution of capital due to capital share (α) and that due to differences in the capital-output ratios. If capital-output ratios are constant across countries i and j, then the contribution of capital due to differences in output per capita in countries i and j simply equals α.

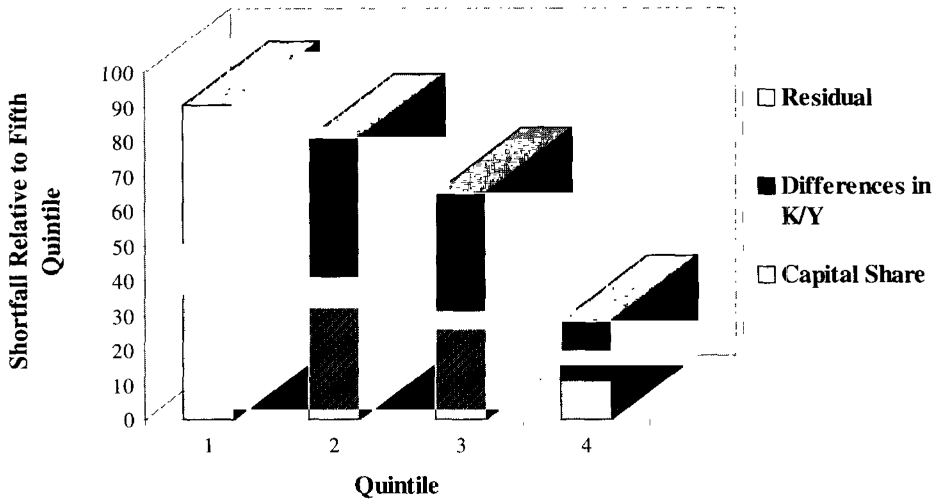
To conduct level accounting, first calculate the percentage shortfall in output of country i relative to the reference country j. $P_i = 100(y_j - y_i)/y_j$. Then construct the contribution of capital due to the output difference as, $P_i\, \phi_{ki}$. As in King and Levine (1994), the level accounting uses figures on aggregate capital stocks (but from PWT 5.6). Countries are classified into five groups, from poorest to richest. The richest group is the reference group.

Figure 2 summarizes the level accounting results. Group 1, the poorest, has more than a 90 percent shortfall in GDP per capita relative to the reference group. TFP accounts for the bulk of cross-country differences in levels of income per capita. The rest is due to capital share of output, assuming constant capital-output ratios, and to the tendency for capital-output ratios to rise with income per capita. Even accounting for systematic cross-country differences in capital-output ratios, the data indicate that capital differences account for less than 40 percent of the cross-group differences in income per capita.[8]

8. Though not directly related to growth accounting, note that the *K/Y* ratio systematically varies with income per capita. Capital-output ratios are systematically larger in richer countries; and, capital-output ratios tend to rise as countries grow, which are inconsistent with Kaldor's stylized fact on capital-output ratios. Consider the regression of the capital-output ratio (κi) on a measure of income per capita relative to that in the United States in the 1980s (yi/y_{USA}). The regression yields the following result:

$$\kappa i = \underset{(0.10)}{0.76} + \underset{(0.18)}{0.59}[yi/y_{USA}],$$

FIGURE 2. Development Accounting by Income Quintiles
(57 Non-Oil-Exporting Countries)



Note: Data cover 57 non-oil-exporting countries.
Source: Authors' calculations based on Penn World Table 5.6 for capital stock estimates.

MANKIW, ROMER, AND WEIL (MRW) LEVEL ACCOUNTING. A second approach to level accounting is suggested by Mankiw, Romer, and Weil (1992), who argue that the Solow model does a good job of accounting for cross-country differences in the level of income per capita. In the steady-state of the Solow model, output per person is given by:

$$Y/L = A\ [s/(x+\delta+n)]^{\alpha/(1-\alpha)}, \tag{9}$$

where Y/L is output per person, A is the level of labor-augmenting productivity, s is the ratio of investment to GDP, x is the rate of labor-augmenting productivity growth, δ is depreciation, n is population growth, and α is the share of capital income in GDP. A 2 percent productivity growth rate and a 7 percent depreciation rate are assumed. Logs are taken of both sides, and the log of output per person is regressed on a constant (ln A) and on the log of the second multiplicative term in equation 9:

---

where κi is the capital-output ratio in country i, standard errors are in parentheses, and the regression includes 57 non-oil-exporting countries. There is a strong positive relationship between output per person relative to the United States and the K/Y ratio. Also, figure 3 shows that the K/Y ratio tends to rise in fast growing countries. Here, the average value of K/Y ratios are plotted year by year for countries with per capita growth rates higher than 3.5 percent a year over the period 1960–92. The K/Y ratio rises rapidly over this fast growth period. Though these differences might be due to transitional dynamics, past works suggests that physical capital accumulation along the transition path is unlikely to explain fully level and growth differences (King and Rebelo 1993).

$$\ln(Y/L) = \ln A + \alpha/(1\text{-}\alpha)\ [\ln s - \ln(x+\delta+n)]. \tag{10}$$

This second term will be called MRW.

The MRW approach is then extended by allowing A to differ across regions, oil-producing and non-oil-producing countries, and OECD and non-OECD countries. (The regions are all-inclusive; the *OECD* and *OIL* dummy variables measure shifts relative to their respective regions.)

Though there is a significant correlation of income with the MRW investment term (consistent with the Solow model), the results in table 3 refute the original MRW idea that productivity levels are the same across countries. South Asia and Sub-Saharan Africa have significantly lower productivity than other regions (income differences that are not explained by the MRW term). The OECD has higher productivity than the rest of the world by a factor of 3 ($e^{1.087}$). Once the productivity level is allowed to vary, the coefficient on MRW implies a capital share of .31—which is in line with most estimates from national income accounting.

Mankiw, Romer, and Weil report that they are even more successful at explaining cross-country income differences when they include a measure of human capital investment, which they define as [$\ln s_h - \ln(x + \delta + n)$]. They define the flow of investment in human capital s_h as the secondary enrollment ratio times the proportion of the labor force of secondary school age. Klenow and Rodriguez-Clare (1997b) and Romer (1995) criticize this measure as overestimating the cross-country variation in human capital by ignoring primary enrollment, which varies much less across countries than secondary enrollment. The results for this

TABLE 3. MRW Least Squares Regression with Regional, Oil, and OECD Dummy Variables

Variable	Coefficient	Standard error	t-statistic	Probability
OECD	1.087817	0.107084	10.15857	0.0000
East Asia	7.559995	0.176696	42.78525	0.0000
South Asia	7.065895	0.139239	50.74634	0.0000
Sub-Saharan Africa	6.946945	0.090968	76.36658	0.0000
Western Hemisphere	7.838313	0.102363	76.57349	0.0000
Middle East and North Africa	7.777138	0.143632	54.14642	0.0000
Europe	7.717543	0.133190	57.94384	0.0000
OIL	0.691058	0.157605	4.384760	0.0000
MRW	0.442301	0.096847	4.567031	0.0000
R^2	0.752210	Mean dependent variable		7.79
Adjusted R^2	0.738969	Standard error of dependent variable		0.994
Standard error of regression	0.508076	Akaike information criterion		1.539
Sum of squared residual	33.81651	Schwarz criterion		1.708
Log likelihood	-98.99247	F-statistic		56.810
		Probability (F-statistic)		0.000

Note: Average log income per capita in 1960–95 is the dependent variable. Number of observations = 139. Standard errors and covariance are White heteroskedasticity-consistent.

Source: Authors' calculations based on World Bank data.

new regression show that although the human capital investment term is highly significant, the original physical capital investment term is only marginally significant (table 4). The OECD productivity advantage and the regional differences in productivity are of the same magnitude as before.

When equation 10 is estimated in first differences from the first half of the period to the second to eliminate country fixed effects, the MRW variable is not statistically significant while TFP growth—the constant in the equation in first differences—varies significantly across regions. This is consistent with the earlier finding that most of the cross-country variation in growth rates per capita is do to differences in TFP growth and not to transitional dynamics between steady states.

Causality

Growth accounting is different from causality. Factor accumulation could ignite productivity growth and overall economic growth. Thus factor accumulation could cause growth even though it does not account for much the cross-country differences in growth rates in the level of GDP per capita. If this were the case, it would be both analytically appropriate and policy wise to focus on factor accumulation. There is also the well-known cross-section correlation between investment share and growth (see Levine and Renelt 1992).

Evidence suggests, however, that physical and human capital accumulation do not cause faster growth. For instance, Blomstrom, Lipsey, and Zejan (1996) show that output growth Granger-causes investment. Injections of capital do not seem to be the driving force of future growth. Similarly, Carroll and Weil (1993)

TABLE 4. MRW Least Squares Regression Including Human Capital, with Regional, Oil, and OECD Dummy Variables

Variable	Coefficient	Standard error	*t*-statistic	Probability
OECD	0.999172	0.126361	7.907255	0.0000
East Asia	8.040507	0.212161	37.89818	0.0000
South Asia	7.593671	0.184937	41.06093	0.0000
Sub-Saharan Africa	7.636055	0.207923	36.72545	0.0000
Western Hemisphere	8.285468	0.136361	60.76117	0.0000
Middle East and North Africa	8.345100	0.192838	43.27516	0.0000
Europe	8.222288	0.161656	50.86290	0.0000
OIL	0.618785	0.179383	3.449517	0.0008
MRW	0.168531	0.095305	1.768343	0.0796
MRWH	0.433868	0.089235	4.862086	0.0000
R^2	0.812286	Mean dependent variable		7.779659
Adjusted R^2	0.797722	Standard error of dependent variable		1.024315
Standard error of regression	0.460689	Akaike information criterion		1.363849
Sum of squared residual	24.61913	Schwarz criterion		1.588951
Log likelihood	–75.92250	*F*-statistic		55.77363
		Probability (*F*-statistic)		0.000000

Note: Average log income per capita in 1960–95 is the dependent variable. Number of observations = 126. Standard errors and covariance are White heteroskedasticity-consistent.

Source: Authors' calculations based on World Bank data.

show that causality tends to run from output growth to savings, not the other way around. Evidence on human capital tells a similar story. Bils and Klenow (1996) argue that the direction of causality runs from growth to human capital, not from human capital to growth. Thus in terms of both physical and human capital, the data do not provide strong support for the contention that factor accumulation ignites faster growth in output per worker.

Summary

Although there are important exceptions, as Young (1995) makes clear, "something else" besides factor inputs accounts for the bulk of cross-country differences in income per capita and growth rates. Furthermore, although growth accounting does not show causality, research suggests that increases in factor accumulation do not ignite faster output growth in the future. While more work is needed, the evidence does not suggest that causality runs from physical or human capital accumulation to economic growth in the broad cross-section of countries. Finally, measurement error may reduce confidence in growth and level accounting. However, the residual is large in both level and growth accounting. Also, level and growth accounting for the 1950s and 1960s produce similar estimates as those for the 1990s. This implies that measurement error would have to have two systematic components. Both the growth rate of measurement error and the level component of measurement error would have to be positive and large in rich, fast-growing countries. Measurement problems may play a role, but a considerable body of evidence suggests that something else—TFP—besides factor accumulation is critical for understanding cross-country differences in the level and growth of GDP per capita.

In giving theoretical content to this residual, Grossman and Helpman (1991), Romer (1990) and Aghion and Howitt (1998) focus on technology, on better instructions for combining raw materials into useful products and services. Romer (1986), Lucas (1988), and others focus on externalities, including spillovers, economies of scale, and various complementarities in explaining the large role played by TFP.[9] Harberger (1998) views TFP as real cost reductions and urges economists not to focus on one underlying cause of TFP because several factors may produce real costs reductions in different sectors of the economy at different times.[10] This is consistent with industry studies that reveal considerable cross-sector variation in TFP growth (Kendrick and Grossman 1980). Prescott (1998) also focuses on technology. He suggests that cross-country differences in resistance to the adoption of better technologies—arising from politics and policies—help explain cross-country differences in TFP (see Holmes and Schmitz 1995; Parente 1994; Parente and Prescott 1996; and Shleifer and Vishny 1993). It would

9. Yet, Burnside (1996) presents evidence suggesting that physical capital externalities are relatively unimportant. Klenow (1998) presents evidence that is consistent with technological change-based model of growth.

10. Costello (1993) shows that TFP has a strong country component and is not specific to particular industries.

be useful in designing models and policies to determine empirically the relative importance of each of these conceptions of TFP.

III. Stylized Fact 2. Divergence, Not Convergence, Is the Big Story

Over the very long run, there has been "divergence, big time," in the words of Pritchett (1997). The richest countries in 1820 subsequently grew faster than the poorest countries in 1820. The ratio of richest to poorest went from 6 to 1 in 1820 to 70 to 1 in 1992 (figure 3). Prior to the Industrial Revolution (1700–50), the difference between the richest and poorest countries was probably only about 2 to 1 (Bairoch 1993, pp. 102–6). Thus, the big story over the past 200–300 years is one of massive divergence in the levels of income per capita between the rich and the poor.[11]

The poor are not getting poorer, but the rich are getting richer a lot faster than the poor. Absolute divergence has continued over the past 30 years, though not as dramatically as in earlier periods (see table 5). And while China and India—countries with very large populations—have performed well recently, growth has diverged significantly even using recent data.[12]

Moreover, the data presented in table 5 understate absolute divergence over 1960–92 because data were lacking for many low- and middle-income countries for the 1990s but not for any high-income countries. This imparts a bias toward convergence in the data similar to that pointed out by De Long (1988) regarding Baumol's (1986) finding of convergence among industrial countries. When the countries that are rich at the end are overrepresented in the sample, the sample is biased toward convergence. The growth rates of the lower three-fifths of the sample would be even lower if data were available for some of the poorly performing low- and middle-income economies in the 1990s.

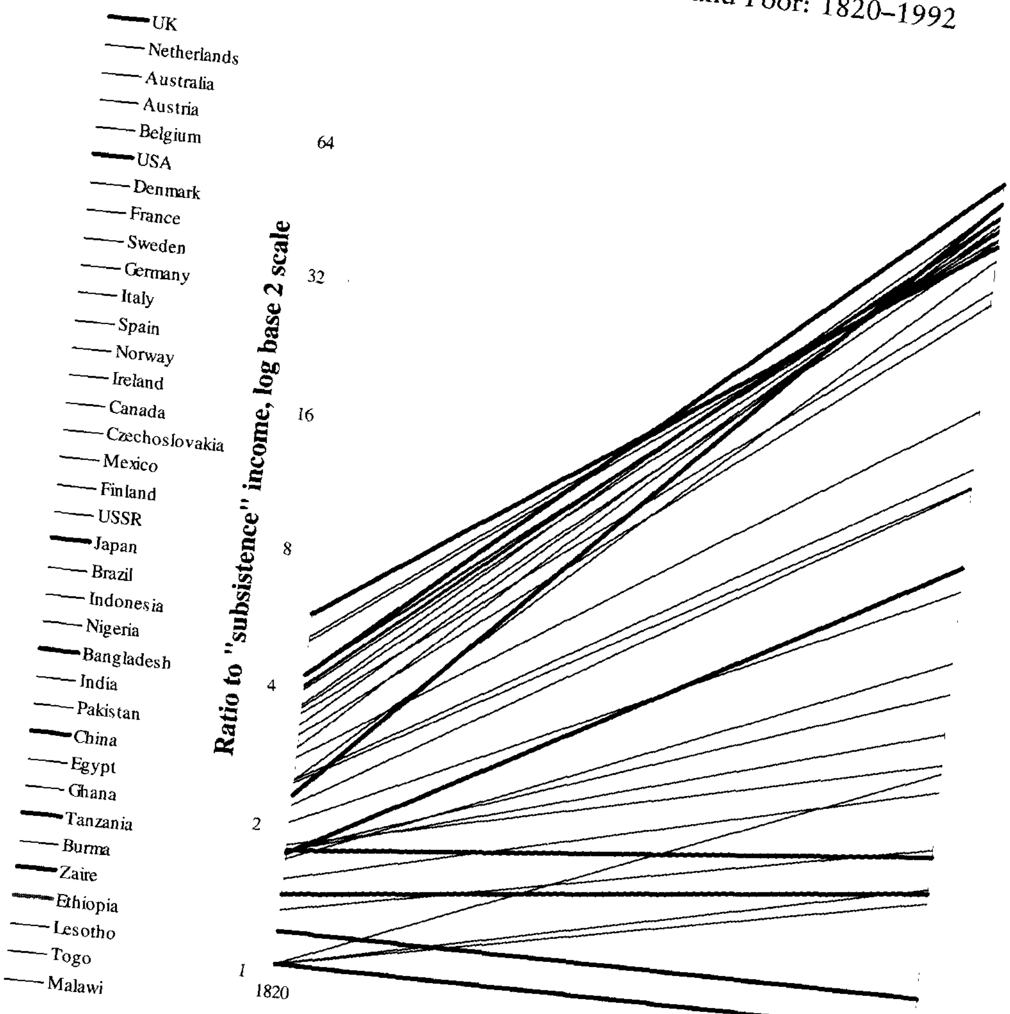
Within the postwar period, this tendency toward divergence has become more pronounced with time. Easterly (2001) found that the bottom half of countries ordered by per capita income in 1980 registered zero per capita growth over 1980–98, while the top half continued to register positive growth. The reason was not a divergence in policies; policies in poor countries were converging toward those of rich countries over 1980–98.

Although many cross-economy data sets exhibit conditional convergence (Barro and Sala-i-Martin 1992), it is difficult to look at the growing differences between the rich and poor and not focus on divergence. Conditional convergence findings hold only after conditioning on an important mechanism for divergence—spillovers

11. See Lucas (1998) for an extensive discussion of this divergence, which he interprets as reflecting different takeoff times for various economies, and which he predicts will decrease as new countries take off.

12. The usual finding that initial income and growth are uncorrelated relied on data that went through 1981 or 1985, using a linear regression of growth on initial income. The use of more recent data (through 1992) and the analysis of quintiles account for this finding of absolute divergence.

FIGURE 3. Growth Rates Diverge between Rich and Poor: 1820–1992



Note: Order in 1820 from richest (top) to poorest (bottom).
Source: Maddison 1995.

from the initial level of knowledge (for which conditional convergence regressions may be controlling with initial level of schooling). Conditional convergence also could follow mechanically from mean reversion (Quah 1993). Because most growth models are closed economy models, it is worth looking at what happens to convergence in closed economies. Kremer (1993) and Ades and Glaeser (1999) have found absolute divergence in the majority of closed developing economies, suggesting an "extent of the market" effect on growth in closed economies.

These findings on divergence should be seen within the context of other stylized facts. Romer (1986) shows that the growth rates of the riches countries have

TABLE 5. Rich Countries Grew Rapidly, Poor Countries Slowly in 1960–92

Income quintile	Average growth of income per person, 1960–92 (%)
Poorest fifth of countries	1.4
Second poorest fifth of countries	1.2
Middle fifth of countries	1.8
Second richest fifth of countries	2.6
Richest fifth of countries	2.2

Note: Countries are classified by income per person in 1960.
Source: Authors' calculations based on Summers-Heston 1991 data with subsequent on-line updates.

not slowed over the last century. King and Rebelo (1993) show that returns to capital in the United States have not been falling over the last century. Together, these observations do not naturally suggest a model that emphasizes capital accumulation and that has diminishing returns to factors, some fixed factor of production, and constant returns to scale. Neither do they provide unequivocal support for any particular conception of what best explains the something else behind these stylized facts.

IV. Stylized Fact 3. Growth Is Not Persistent, but Factor Accumulation Is

Growth is remarkably unstable over time. The correlation of per capita growth in 1977–92 with per capita growth in 1960–76 across 135 countries is only .08.[13] This low persistence is not just a characteristic of the postwar era. For the 25 countries for which there are data (Maddison 1995), the correlation between 1820–70 and 1870–1929 is only .097.

In contrast, the cross-period correlation of growth in capital per capita is 0.41. For models that postulate a linear relationship between growth and the share of investment in GDP (using investment share in GDP as an alternative measure of capital accumulation), the mismatch in persistence is even worse.[14] The correlation of investment share in GDP in 1977–92 with investment share in 1960–76 is .85. Nor do models that postulate growth per capita as a function of human capital accumulation do better. The correlation across 1960–76 and 1977–92 for primary enrollment is .82, while the cross-period correlation for secondary enrollment is .91. This suggests that much of the large variation of growth over

13. Data on per capita growth are from the PWT 5.6. The low persistence of growth rates, and the high persistence of investment and education, was previously noted in Easterly and others (1993).

14. Models supposing a linear relationship between growth and investment have a long history in economics. See Easterly (1999b) for a review of the Harrod-Domar tradition that continues down to the present. For a new growth theory justification of this relationship, see McGrattan (1998).

time is not explained by the much smaller variation in physical and human capital accumulation.

Takeoff into Steady-State Growth Is Rare

The typical model of growth, in both the old and new growth literatures, features a steady-state growth rate. Historically, this was probably inspired by the U.S. experience of remarkably steady growth of about 2 percent per capita over nearly two centuries (Jones 1995a, 1995b; Rebelo and Stokey 1995).

Because all countries must have had prior histories of stagnation, another characterization of the typical growth path is the "takeoff into self-sustained growth" (the phrase is from Rostow 1960; more recent theoretical modeling of takeoff includes Baldwin 1998, Krugman and Venables 1995, Jones 1999, Lucas 1998, and Hansen and Prescott 1998). The prevailing image is a smooth acceleration from stagnation into steady-state growth. Developing countries are supposed to have taken off beginning in the 1960s, when their growth was rapid and exceeded expectations.

Experience did not bear out the idea of steady growth beginning in the 1960s. Many countries experienced booms and crashes (Pritchett 2000, Rodrik 1998). Even when ten-year average growth rates are used, which should be long enough to iron out cyclical swings, the cross-section standard deviation is about 2.5 percentage points and the variation over time swamps the cross-section variation. In 48 of 119 countries with 20 years or more of data over 1960–97, a breakpoint can be found in which the subsequent decade's per capita growth is more than 5 percentage points—two cross-section standard deviations—above or below the previous decade's growth.[15] All of the countries with growth booms or crashes were developing countries, except for Greece and Portugal. Figure 4 illustrates the rollercoaster ride of Côte d'Ivoire, Guyana, Jamaica, and Nigeria.

Stable growth may be a better description of industrial than developing countries. Of 88 industrial and developing countries with complete data for 1960–97, only 12 had growth above 2 percent per capita in every decade. Half were in East Asia.

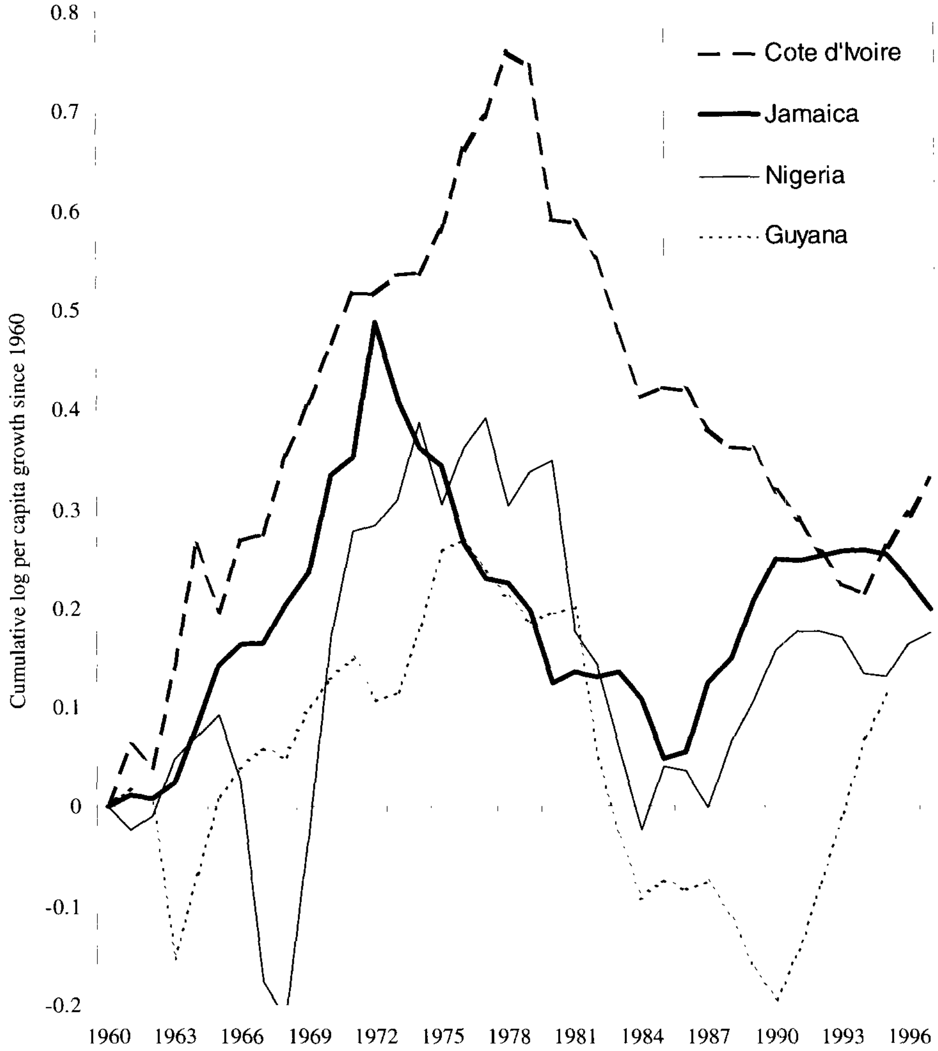
Variance Decomposition over Time

This supposition of unstable growth is further confirmed by variance decomposition exercises, with decomposition over time rather than across countries. In conjunction with the cross-country variance decomposition presented above, this analysis represents a full exploration of the panel data on growth and its factors.

A panel of seven five-year time periods was constructed for each country for per capita growth and growth in physical capital per capita. Country means are

15. Thirty-seven countries had a growth drop of 5 percentage points or more, 19 countries of 5 percentage points or more, and 8 countries were included in both groups.

FIGURE 4. Examples of Variable Per Capita Income over Time: 1960–96



Source: World Bank data.

then subtracted and the variance is analyzed using the same formula as before (see equation 3). For the same sample of countries, TFP accounts for 86 percent of the intertemporal variation in overall growth and 61 percent of the cross-sectional variation. Thus, growth is much more unstable over time than physical capital growth.

Besides emphasizing the importance of TFP in explaining long-run development patterns, the findings that growth is not persistent and that growth patterns are very different across countries complicate the challenge for economic theorists.

Existing models miss important development experiences. Some countries grow steadily (the United States). Some grow steadily and then stop for long periods (Argentina). Some do not grow for long periods and then suddenly take off (Republic of Korea, Thailand). Others have never grown (Somalia). Accounting for these very different growth experiences will be very difficult with sole reliance on either steady-state models or standard multiple-equilibria models. Different models may be needed for different patterns of growth across countries. Steady-state models fit the U.S. type experience. Multiple equilibria models are a better fit for unstable growth cases because countries' long-run fundamentals are stable.[16]

V. Stylized Fact 4. When It Rains, It Pours: All Factors Flow in the Same Direction

This section presents new information on the concentration of economic activity, using cross-country data, data from counties in the United States, information on developing countries, and data on international flows of capital, labor, and human capital. This concentration has a fractal-like quality. It recurs at all levels of analysis, from the global to the urban. It suggests that some regions have "something" that attracts all factors of production, whereas others do not.

Better policies (legal systems, property rights, political stability, public education, infrastructure, taxes, regulations, macroeconomic stability) in one area than in another could explain these factor flows. But such policies are national; they cannot explain findings of within-country concentration (discussed below). Externalities may lead to factor congregation. Critically, policy differences, or externalities, or differences in something else do not have to be large. Small differences can have dramatic long-run implications. So, although no specific explanation is offered, the results of this analysis suggest a need for more work on economic geography as a vehicle for understanding economic growth.

Concentration

An obvious observation at the global level is that high income is concentrated among a small number of countries (see map 1). The top 20 countries have only 15 percent of world population but produce 50 percent of world GDP. The poorest

16. The nonpersistence of growth rates does not inherently contradict the stylized fact of divergence or the stylized fact that national policies influence long-run growth rates. While policies are persistent and significantly associated with long-run growth (which is not persistent), the R^2 of the growth regression is generally smaller than 0.50. Thus, something else (besides national policies) is very important for explaining cross-country differences in long-run growth rates. In terms of divergence, the stylized fact of the nonpersistence of growth rates emphasizes that growth follows very different paths across countries and that there is a high degree of volatility. Nevertheless, there are countries that have achieved comparatively greater success over the long run. While France, Germany, and the United Kingdom have experienced growth fluctuations, they have enjoyed a steeper—and less volatile—growth path than Argentina and Venezuela, for example, whose growth paths have not only been more volatile but exhibited dramatic changes in trends.

MAP 1. The Rich and the Poor

The countries in black contain 15 percent of world population but produce 50 percent of world GDP. The countries in gray contain 50 percent of world population but produce 14 percent of world GDP.

half of the world's population accounts for only 14 percent of its GDP.[17] These concentrations of wealth and poverty have an ethnic and geographic dimension: 18 of the top 20 countries are in Western Europe or were settled primarily by Western Europeans; 17 of the poorest 20 countries are in tropical Africa. The richest country in 1985 (the United States) had an income 55 times that of the poorest country (Ethiopia). When inequality within countries is considered, international income differences are even starker. The income of the richest quintile in the United States was 528 times that of the poorest quintile in Guinea-Bissau.

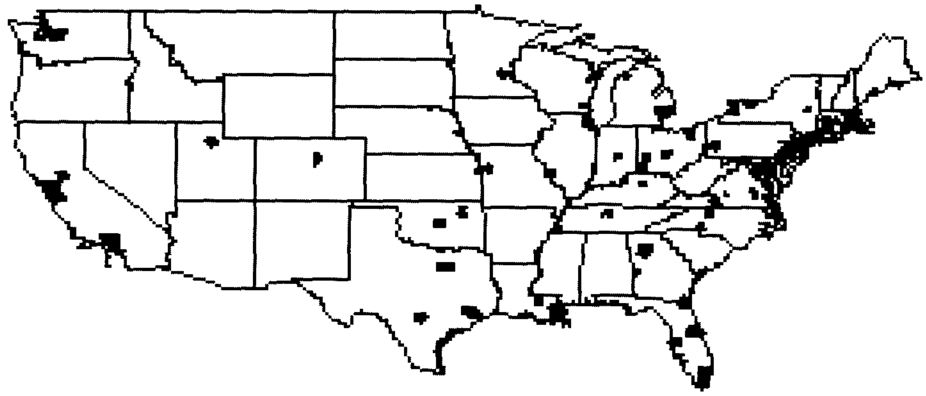
Income is highly concentrated by area as well, as shown by data on GDP per square kilometer. The densest 10 percent of world land area accounts for 54 percent of global GDP; the least dense for only 11 percent.[18]

These calculations understate the degree of concentration because they assume that income is evenly spread among people and land area within countries. A more detailed look within countries also shows high concentrations of wealth and poverty.

17. These calculations omit the oil-exporting countries, in which GDP is not properly measured because all of oil extraction is treated as current income rather than asset depletion.

18. An alternative explanation would be that some land areas, accounting for a small share of the earth's surface, have a large productivity advantage. Mellinger, Sachs, and Gallup (1999) argue that temperate coastal zones have a large productivity advantage. If this were true, economic activity would be distributed fairly evenly along temperate coastal zones (adjusting for any small intrinsic differences among such zones). However, even along temperate coastal zones, casual observations would suggest high bunching of activity.

MAP 2. Densely Populated U.S. Counties



Counties shown in black take up 2 percent of U.S. land area but account for half of U.S. GDP.

Consider the United States. Data on GDP per square mile for 3,141 counties show that counties accounting for only 2 percent of the land produce 50 percent of GDP, while the least dense counties that account for 50 percent of the land produce only 2 percent of GDP (map 2). Nor is this finding a consequence merely of the large unsettled areas of the West and Alaska. The same calculation for land east of the Mississippi River yields similarly extreme concentration: 50 percent of GDP is produced on 4 percent of the land. The densest county is New York, New York, with a GDP per square mile of $1.5 billion. This is about 55,000 times more than the least dense county east of the Mississippi ($27,000 per square mile in Keweenaw, Michigan). Even this understates the degree of concentration because even the most casual empiricism will detect rich and poor areas within a given county (New York county contains Harlem as well as Wall Street).

The concentration of counties accounting for half of U.S. GDP is explained by the fact that these are metropolitan counties and most economic activity takes place in densely populated metropolitan areas. Metropolitan counties are $3,300 richer per person than rural counties (the difference is statistically significant, with a *t*-statistic of 29). More generally, there is a strong correlation between per capita income of U.S. counties and their population density (correlation coefficient of .48 for the log of both concepts, with a *t*-statistic of 30 on the bivariate association). But concentration is high even within metropolitan counties: 50 percent of metropolitan GDP is produced in counties accounting for only 6 percent of metropolitan land area.[19]

There are also regional income differences between metropolitan areas. Metropolitan areas in the Boston–Washington corridor have a per capita in-

19. Metropolitan counties are those that belong to a PMSA or MSA in the census classification of counties.

come $5,874 higher on average than other metropolitan areas. This is a huge difference: It is equal to 2.4 standard deviations in the metropolitan area sample. Although there may be differences in the cost of living, they are unlikely to be so large as to explain this difference. (The rent component of the cost of living may reflect the productivity or the amenity advantages of the area—it seems unlikely that amenities are different enough among areas to explain these differences.)

There are other possible explanations of geographic concentration, such as inherent geographic advantages. Like Mellinger, Sachs, and Gallup (1999), Rappaport and Sachs (1999) argue that spatial concentration of activity in the United States has much to do with access to the coast. However, casual observation suggests high concentration even within coastal areas (there are sections along the Boston–Washington corridor with no radio reception). Some studies suggest that high transport costs and low congestion costs could also play a role (Krugman 1991, 1995, 1998; Fujita, Krugman, and Venables 1999). However, these studies also point to locations of particular industries (the Silicon Valley phenomenon) as evidence of other types of geographic spillover, including technology spillovers and specialized producer services with high fixed costs. And the high rents in downtown metropolitan areas suggest that congestion costs are significant. As Lucas (1988, p. 39) says, "What can people be paying Manhattan or downtown Chicago rents *for*, if not for being near other people?"

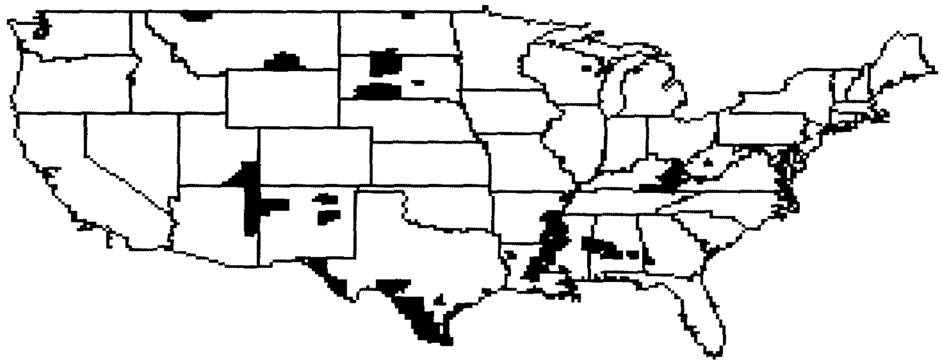
Poor Areas

Like wealth, poverty is also concentrated. In the United States, poverty is regionally concentrated. These concentrations have an ethnic dimension as well (see map 3). Four ethnogeographic clusters of counties have poverty rates above 35 percent:

- Counties in the West with large proportions (> 35 percent) of Native Americans.
- Counties along the Mexican border with large proportions (> 35 percent) of Hispanics.
- Counties along the lower Mississippi River in Arkansas, Mississippi, and Louisiana and in the "black belt" of Alabama, all of which have large proportions of blacks (> 35 percent).
- Virtually all-white counties in the mountains of eastern Kentucky.

The county data did not pick up the well-known phenomenon of inner-city poverty, mainly among blacks, because counties that include inner cities also include rich suburbs. (An isolated example of an all-black city is East St. Louis, Illinois, which is 98 percent black and has a poverty rate of 44 percent.) Of course, poverty is concentrated in the inner city as well. An inner city ZIP code in Washington, D.C., College Heights in Anacostia, has only one-fifth of the income of a rich ZIP code in Bethesda, Maryland. This has an ethnic dimension again because College Heights is 96 percent black and the rich ZIP code in Bethesda is 96 percent white. The Washington, D.C., metropolitan area as a whole shows

MAP 3. Poverty Traps in the U.S. County Data



Counties in black have more than 35 percent poverty rate.

the striking East–West divide between poor and rich ZIP codes, which again roughly corresponds to the black–white ethnic divide (see map 4).[20]

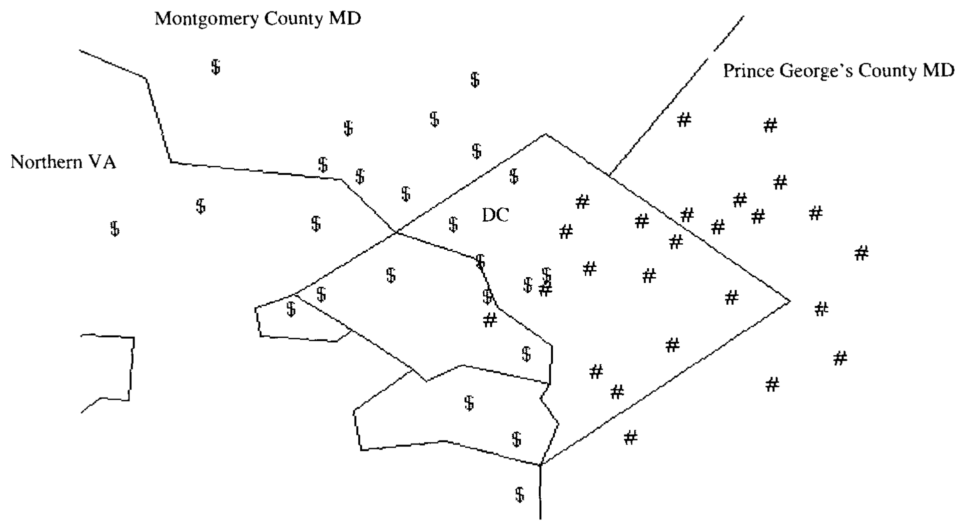
Borjas (1995, 1999) suggests that strong neighborhood and ethnic externalities may help explain poverty and ethnic clusters within cities. The 1990 census tracts with the highest shares of blacks have 50 percent of the black population but only 1 percent of the white population.[21] Although this segregation by race and class could simply reflect the preferences of rich white people to live next to each other, economists usually prefer to offer economic motivations rather than exogenous preferences as explanations of economic phenomena. Benabou (1993, 1996) stresses the endogenous sorting between rich and poor, so the rich can take advantage of such externalities as locally funded schools.

Poverty areas exist within many countries: northeast Brazil, southern Italy, Chiapas in Mexico, Balochistan in Pakistan, and the Atlantic Provinces in Canada. Researchers have found that externalities explain part of these poverty clusters. Bouillon, Legovini, and Lustig (1999) find a negative Chiapas effect in Mexican household income data, an effect that has worsened over time. Households in the poor region of Tangail-Jamalpur in Bangladesh earned less than households with similar characteristics in the better-off region of Dhaka (Ravallion and Wodon 1998). Ravallion and Jalan (1996) and Jalan and Ravallion (1997) likewise found that households in poor counties in southwest China earned less than households with identical human capital and other characteristics in rich Guangdong Province. Rauch (1993) found that individuals with identical characteristics earn less in low human capital cities in the United States than in high human capital cities.

20. Brookings Institution (1999) notes that this East–West geographic divide of the Washington, D.C., area shows up in many socioeconomic variables (poverty rates, free and reduced price school lunches, road spending).

21. From the Urban Institute's Underclass Database, which contains data on white, black, and "other" population numbers for 43,052 census tracts in the United States.

MAP 4. Rich and Poor ZIP Codes in the Washington, D.C., Metropolitan Area



$ indicates richest fourth of zip codes in metropolitan area; # indicates poorest fourth.

Ethnic and Related Differentials

Some theories stress in-group externalities (Borjas 1992, 1995, 1999; Benabou 1993, 1996). Poverty and riches are also concentrated in certain ethnic groups. Exogenous savings preferences are not an appealing explanation. Discrimination and intergenerational transmission could explain ethnic differences, but for growth models the differences seem more consistent with in-group spillovers than with individual factor accumulation.

The purely ethnic differentials in the United States are well known. Asians earn 16 percent more than whites, and blacks earn 41 percent less, Native Americans 36 percent less, and Hispanics 31 percent less.[22] There are also more subtle ethnic earnings differentials. Third-generation immigrants with Austrian grandparents had 20 percent higher wages in 1980 than third-generation immigrants with Belgian grandparents (Borjas 1992). Among Native Americans, the Iroquois earn almost twice the median household income of the Sioux.

Other ethnic differentials appear by religion. Episcopalians earn 31 percent more than Methodists (Kosmin and Lachman 1993, p. 260). Of the Forbes 400 richest Americans, 23 percent are Jewish, although only 2 percent of the U.S. population is Jewish (Lipset 1997).[23]

22. Tables 52 and 724, *1995 Statistical Abstract of the United States* (United States Government 1996).

23. Ethnic differentials are also common in other countries. The ethnic dimension of rich trading elites is well known—the Lebanese in West Africa, the Indians in East Africa, and the overseas Chinese in Southeast Asia. Virtually every country has its own ethnographic group noted for their success. For

In Latin America, the main ethnic divide is between indigenous and nonindigenous populations (table 6). But even within indigenous groups, there are ethnic differentials. For example, there are four main language groups among Guatemala's indigenous population. Patrinos (1997) shows that the Quiche-speaking indigenous groups earn 22 percent less on average than Kekchi-speaking groups.

For Africa, there are numerous anecdotes about income differentials between ethnic groups, but little hard data. An exception is South Africa, where whites have 9.5 times the income of blacks. More surprisingly, among all-black traditional authorities (an administrative unit something like a village) in the state of KwaZulu-Natal, the richest traditional authority has 54 times the income of the poorest (Klitgaard and Fitschen 1997).

Factor Movement

Factor movement toward the richest areas reinforces the concentration of economic activity. Each factor of production flows to where it is already abundant.

Labor migration is overwhelmingly toward the richest countries. The three richest countries alone (the United States, Canada, and Switzerland) receive half the net immigration of all countries reporting net immigration. Countries in the richest quintile are all net recipients of migrants. Only 8 of the 90 countries in the bottom four-fifths of the sample are net recipients of migrants. Barro and Sala-i-Martin (1995, pp. 403–10) find that migration goes from poorer to richer regions in samples of U.S. states, Japanese prefectures, and European regions.

Migration also goes from sparsely populated to densely populated areas. There is a statistically significant correlation of .20 between the immigration rate of U.S. counties from 1980 to 1990 and population density in 1980. Labor flowed to areas where it was already abundant. Migration goes from poor to rich counties, with a statistically significant correlation of .21 between initial income and immigration rate (confirming the Barro and Sala-i-Martin 1995 finding for U.S. states). These two findings are related, as there is a significant positive correlation between population density and per capita income across counties.[24] A regression of the immigration rate for 1980–90 by county on population density in 1980 and income per capita in 1980 finds both to be highly significant.[25]

Embodied in this flow of labor are flows of human capital toward rich countries, the famous "brain drain." In the poorest fifth of countries, the probability of emigrating to the United States is 3.4 times higher for an educated person than

example, in The Gambia a tiny indigenous ethnic group called the Serahule is reported to dominate business out of all proportion to their numbers. In Zaire, Kasaians have been dominant in managerial and technical jobs since the days of colonial rule (*New York Times*, 9/18/1996).

24. Ciccone and Hall (1996) have a related finding for U.S. states.

25. The t-statistics are 8.2 for the log of population density in 1980 and 8.9 for the log of per capita income in 1979. The equation has an R^2 of .065 and has 3,133 observations. The county data are from Alesina, Baqir, and Easterly (1999).

TABLE 6. Poverty Rate Differential among Indigenous and Nonindigenous Groups in Selected Latin American Countries

Country	Indigenous groups	Nonindigenous groups
Bolivia	64.3	48.1
Guatemala	86.6	53.9
Mexico	80.6	17.9
Peru	79.0	49.7

Source: Psacharopoulos and Patrinos 1994 (p. 6).

for an uneducated person (based on data from Grubel and Scott 1977). Because education and income are strongly and positively correlated, human capital is flowing to where it is already abundant—the rich countries.

Carrington and Detragiache (1998) found that in 51 of 61 developing countries in their sample, people with a university education were more likely to emigrate to the United States than people with a secondary education. In all 61 countries, migration rates to the United States were lower for people with a primary education or less than for people with a secondary or university education. Lower bound estimates for the highest rates of emigration by those with university education range as high as 77 percent (Guyana), with rates of 59 percent for The Gambia, 67 percent for Jamaica, and 57 percent for Trinidad and Tobago.[26] None of the emigration rates for the primary or less educated exceeds 2 percent.

The disproportionate weight of the skilled population in U.S. immigration may reflect U.S. policy. However, Borjas (1999) notes that U.S. immigration policy has tended to favor unskilled labor with family connections in the United States rather than skilled labor. In the richest fifth of countries, moreover, the probability is roughly the same that educated and uneducated will emigrate to the United States. Borjas, Bronars, and Trejo (1992) also find that the more highly educated are more likely to migrate within the United States than the less educated.[27]

Capital also flows mainly to areas that are already rich, as Lucas (1990) famously pointed out. In 1990, the richest 20 percent of world population received 92 percent of gross portfolio capital inflows, whereas the poorest 20 percent received 0.1 percent. The richest 20 percent of the world population received 79 percent of foreign direct investment, and the poorest 20 percent received 0.7 percent. Altogether, the richest 20 percent of the world population received 88 percent of gross private capital gross inflows, and the poorest 20 percent received 1 percent.

26. Note these are all small countries. Carrington and Detragiache (1998) point out that U.S. immigration quotas are less binding for small countries, because with some exceptions the legal immigration quota is 20,000 per country regardless of population size.

27. Casual observation suggests "brain drain" within countries. The best lawyers and doctors congregate within a few metropolitan areas like New York, where skilled doctors and lawyers are abundant, while poorer areas have difficulty attracting the top-drawer professionals.

Skill Premia and Human Capital

Skilled workers earn less, rather than more, in poor countries. This seems inconsistent with the open economy version of the neoclassical factor accumulation model by Barro and others (1995). In the Barro and others model, capital flows equalize the rate of return to physical capital across countries, while human capital is immobile. Immobile human capital explains the difference in per worker income across countries. As Romer (1995) points out, this implies that both the skilled wage and the skill premium should be much higher in poor countries than in rich countries. To illustrate, specify a standard production function for country i:

$$Y_i = AK_i^{\alpha} L_i^{\beta} H_i^{1-\alpha-\beta}. \tag{11}$$

Assuming that technology (A) is the same across countries and that rates of return to physical capital are equated across countries, the ratio of the skilled wage in country i to that in country j is a function of their per capita incomes:

$$\frac{\dfrac{\partial Y_i}{\partial H_i}}{\dfrac{\partial Y_j}{\partial H_j}} = \left[\frac{Y_i/L_i}{Y_j/L_j}\right]^{\frac{-\beta}{1-\alpha-\beta}}. \tag{12}$$

On the basis of the physical (.3) and human capital shares (.5) suggested by Mankiw (1995), skilled wages should be five times greater in India than in the United States (to correspond to a 14-fold difference in per capita income). In general, equation 12 shows that skilled wages differences across countries should be inversely related to per capita income if human capital abundance explains income differences across countries, as in the Barro and others model. The skill premium should be 70 times higher in India than in the United States. If the ratio of skilled to unskilled wages is about 2 in the United States, the ratio should be 140 in India. This would imply an astonishing rate of return to education—70 times larger than in the United States.

The facts do not support these predictions. Skilled workers earn more in rich countries. Fragmentary data from wage surveys show that engineers average \$55,000 in New York and \$2,300 in Bombay (Union Bank of Switzerland 1994). Far from being 5 times higher in India than in the United States, skilled wages are 24 times higher in the United States than in India. The higher wages across all occupational groups are consistent with greater technological progress (A) in the United States than in India. The skilled wage (proxied by salaries of engineers, adjusted for purchasing power) is positively associated with per capita income across countries, as a productivity explanation of income differences would imply (figure 5), and not negatively correlated, as a Barro and others model of human capital explanation would imply. The correlation between skilled wages and per capita income across 44 countries is .81.

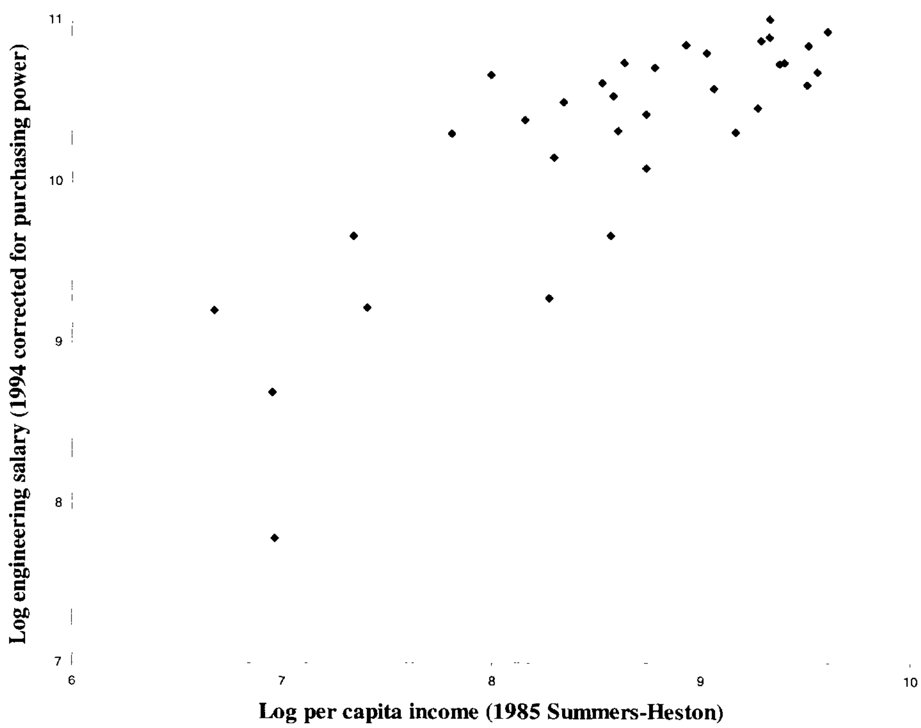
Within India, engineers earn only about three times what building laborers earn. Rates of return to education are also only about twice as high in low-income countries (11 percent) as in high-income countries (6 percent; Psacharopolous 1994, p. 1332)—not 42 times higher. Consistent with this evidence, the flow of human capital is toward rich countries, despite barriers to immigration.

Evaluating Growth Models in Light of Income Concentration

The high concentration of income, reinforced by the flow of all factors toward the richest areas, is inconsistent with the neoclassical growth model. The distribution of income across space and across people at all levels is highly skewed to the right (skewness coefficient of 2.58 across countries in 1980, 2.2 across U.S. cities, and 1.6 across U.S. counties in 1990, where 0 is symmetry). There is no reason to think that the determinants of income in the neoclassical model (saving, population growth) are skewed to the right, but models of technological complementarities (see, for example, Kremer 1993) can explain the skewness.

Moreover, the concentration of factors in rich, densely populated areas even within countries is incompatible with a version of the neoclassical model that

FIGURE 5. Skilled Real Wage and Per Capita Income across Countries



Source: Authors' calculations based on Summers and Heston for per capita income and Union Bank of Switzerland for engineering salaries.

includes land as a factor of production. With land in fixed supply, physical and human capital and labor should all flow to areas abundant in land (adjusting for land quality) but scarce in other factors.

Furthermore, in the neoclassical model of Mankiw, Romer, and Weil (1992), physical and human capital should also flow from rich to poor areas, and unskilled labor from poor to rich. But as this study shows, physical and human capital flow toward rich areas, as does unskilled labor, though it is less mobile.

Stylized fact 4 is in harmony with Klenow and Rodriguez-Clare (1997b), who complain that the "neoclassical revival in growth economics" has "gone too far." The neoclassical model does not explain why wealth and poverty are concentrated in certain regions within countries or why there are such pronounced income differences between ethnic groups. Stylized fact 4 is consistent with poverty trap models (Azariadis and Drazen 1990, Becker, Murphy, and Tamura 1990, Kremer 1993, Murphy, Shleifer, and Vishny 1989); with models of in-group ethnic and neighborhood externalities (Borjas 1992, 1995, 1999, Benabou 1993, 1996), and with models of geographic externalities (Krugman 1991, 1995, 1998, Fujita, Krugman, and Venables 1999).

Stylized fact 4 also seems to be more consistent with a productivity explanation of income differences than with a factor accumulation story. If a rich area is rich because technology *(A)* is more advanced, then all factors of production will tend to flow toward this rich area, reinforcing the concentration. Spillovers between agents also seem more natural with technological models of growth, as technological knowledge is inherently more nonrival and nonexcludable than factor accumulation. Technological spillovers between agents will lead to endogenous matching of rich agents with each other, and those matches will reinforce the matching of poor people with other poor people (as in the O-ring story of Kremer [1993] or the inequality model of Benabou [1996]). A better understanding of economic geography and externalities would help shape more realistic models of economic growth.

VI. Stylized Fact 5. Policy Matters

The empirical literature on national policies and economic growth is huge. There is considerable disagreement about which policies are most strongly linked with economic growth. Some analysts focus on openness to international trade (Frankel and Romer 1999), some on fiscal policy (Easterly and Rebelo 1993), some on financial development (Levine, Loayza, and Beck 2000), and some on macroeconomic policies (Fischer 1993). All these studies have at least one feature in common: They all find that some indicator of national policy is strongly linked with economic growth, confirming the argument made by Levine and Renelt (1992).

Most empirical assessments of the growth-policy relationship are plagued by three shortcomings. First, most do not confront endogeneity. Even when instrumental variables are used, studies frequently assume that many of the regressors are exogenous and focus only on the potential endogeneity of one variable of interest. This failure to fully confront causality may produce biased assessments. Second, traditional cross-country regressions may suffer from omitted variable bias. That is, cross-country growth regressions may omit an important country-specific effect and thereby produce biased coefficient estimates. Third, almost all cross-country regressions included lagged real per capita GDP as a regressor. Because the dependent variable is the growth rate of real per capita GDP, this specification may produce biased coefficient estimates.

This study uses recent econometric techniques to examine the links between economic growth and a range of national policies. These new techniques ameliorate these potential biases so that more accurate inferences can be drawn about the impact of national policies on economic growth. The goal is not to identify the most important policies influencing growth; it is to compile key stylized facts associated with long-run growth.

Use of the latest econometric techniques (see appendix) confirms earlier findings that national policies are strongly linked with economic growth. The regression results are consistent with policies having significant long-run effects on national growth rates or on steady-state levels of national output. The regression results also show that national policies are strongly linked with TFP growth (Beck, Levine, and Loayza 2000).

The relationship between the exogenous component of national policies and economic growth is assessed using a set of conditioning information and policy indicators suggested by theory and past empirical work. Specifically, the initial level of real income per capita is included to control for convergence. The standard neoclassical growth model predicts convergence to the steady-state output per person ratio (Barro and Sala-i-Martin 1995). The coefficient on initial income does not necessarily capture only neoclassical transitional dynamics. In technology diffusion models, initial income may proxy for the initial gap in TFP between economies. In these models, therefore, catch-up can be in TFP as well as in traditional factors of production. Average years of schooling was included as an indicator of the human capital stock in the economy. Its inclusion can help in controlling for differences in steady-state levels of human capital (Barro and Sala-i-Martin 1992). Also, schooling may directly influence economic growth (Lucas 1988).

Five policy indicators were used. The inflation rate and the ratio of government expenditures to GDP were included as indicators of macroeconomic stability (Easterly and Rebelo 1993, Fischer 1993). Exports plus imports as a share of GDP and the black market exchange rate premium were included to capture degree of openness (Frankel and Romer 1999). Financial intermediary credit to the private sector as a share of GDP was included as a measure of financial intermediary development (Levine, Loayza, and Beck 2000). There is no attempt to suggest that these are the most important policy indicators. They are used only to assess whether economic growth is strongly linked with these national policy indicators after controlling for endogeneity and other biases in existing empirical work.

As in much of the cross-country literature, the regression results show evidence of conditional convergence (table 7). Specifically, contingent of the level of human capital, poorer countries tend to grow faster than richer countries as each country converges toward its steady-state. This finding is consistent with a major implication of the textbook neoclassical growth model. The regression also shows that greater human capital—as measured by average years of schooling of the working age population—is associated with faster economic growth. Moreover, since the GMM panel estimator controls for endogeneity, this finding suggests that the exogenous component of schooling exerts a positive impact on economic growth. These results are consistent with models that focus on factor accumulation or on TFP growth.

The results are consistent with—but not proof of—long-run growth effects of national policies, which is consistent with an endogenous productivity growth model. In contrast, models that feature only transitional factor accumulation dynamics usually predict weaker policy effects on growth than endogenous productivity growth models. Furthermore, complementary work in Beck, Levine, and Loayza (2000) suggests a powerful connection between national policies and TFP growth. The exogenous components of international openness—as measured

TABLE 7. Economic Growth and National Policies

Variable	Result
Constant	0.082
	(0.875)
Initial income per capita[a]	-0.496
	(0.001)
Average years of schooling[b]	0.950
	(0.001)
Openness to trade[a]	1.311
	(0.001)
Inflation[b]	0.181
	(0.475)
Government size[a]	-1.445
	(0.001)
Black market premium[b]	-1.192
	(0.001)
Private credit[a]	1.443
	(0.001)
Sargan test[c] (*p*-value)	0.506
Serial correlation test[d] (*p*-value)	0.803

Note: Numbers in parentheses are *p*-values. The dependent variable is real per capita GDP growth.

[a]Included as log(variable).

[b]Included as log(1 + variable).

[c]The null hypothesis is that the instruments used are not correlated with the residuals.

[d]The null hypothesis is that the errors in the first-difference regression exhibit no second-order serial correlation.

Source: Authors's calculations based on analyses in Beck, Levine, and Loayza (2000).

by the ratio of trade to GDP and by black market exchange rate premia—are significantly correlated with economic growth.

Macroeconomic policy is also important. Large government tends to hurt economic growth, although inflation does not enter significantly. A higher black market exchange rate premium exerts a negative impact on growth. More international trade tends to boost economic growth. While considerable research suggests a negative link between inflation and economic performance (Bruno and Easterly 1998), recent research suggests that inflation is strongly linked with financial development (Boyd, Levine, and Smith 2001). Thus, it may not enjoy an independent link with growth when financial development is controlled for. Finally, a higher level of financial development boosts economic growth. In sum, national policies are strongly linked with economic growth.

VII. Conclusion

The major empirical regularities of economic growth emphasize the role of something else besides factor accumulation. The TFP residual accounts for most of the cross-country and cross-time variation in growth. Income across countries diverges over the long run, while the growth rates of the rich are not slowing and returns to capital are not falling. This observation is less consistent with simple models that feature diminishing returns, factor accumulation, some fixed factor of production, and constant returns to scale and more consistent with the observation that something else is important for explaining long-run economic success. Growth is highly unstable over time, whereas factor accumulation is more stable, which certainly emphasizes the role of something else in explaining variations in economic growth. All factors of production flow to the richest areas, suggesting that they are rich because of high A rather than high K. Divergence of per capita incomes and the concentration of economic activity suggest that technology has increasing returns. Finally, national policies are strongly linked with long-run economic growth rates.

Nothing in this study argues that factor accumulation is unimportant in general or denies that it is critically important for some countries at specific junctures. TFP does not explain everything, everywhere, always. Rather, the study shows that something else—besides factor accumulation—plays a prominent role in explaining differences in economic performance across countries.

More research is needed on the "residual" determinants of growth and income, such as technology and externalities. There is little doubt that technology is a formidable force. Nordhaus (1994) estimates that 1 Btu of fuel consumption today buys 900 times more lighting (measured in lumen hours) than it did in 1800. Over the past two decades, computing power per dollar invested has risen by a factor of 10,000, and the cost of sending information over optical fiber has fallen by a factor of 1,000 (World Bank 1999, pp. 5 and 57). Just from 1991 to 1998, the price of a megabyte of hard disk storage fell from \$5 to \$0.03.[28]

28. *www.duke.edu/~mccann/q-tech.htm#Death of Distance.*

Not every technology has improved at this speed of course. But Mokyr (1992) was right to call technology "the lever of riches."

APPENDIX: ECONOMETRIC METHODOLOGY

A generalized method of moments (GMM) dynamic panel estimator was used to assess the relationship between policy and economic growth. Panel data for 73 countries over the period 1960–95 were averaged over seven nonoverlapping five-year periods.

Consider the following equation:

$$y_{i,t} - y_{i,t-1} = (\alpha - 1)y_{i,t-1} + \beta' X_{i,t} + \eta_i + \varepsilon_{i,t} \tag{13}$$

where y is the logarithm of real per capita GDP, X is the set of explanatory variables (other than lagged per capita GDP), η is an unobserved country-specific effect, ε is the error term, and the subscripts i and t represent country and time period. Time dummy variables were also included to account for time-specific effects.

Equation 13 can be rewritten as:

$$y_{i,t} = \alpha y_{i,t-1} + \beta' X_{i,t} + \eta_i + \varepsilon_{i,t}. \tag{14}$$

First differences of equation 14 are taken to eliminate the country-specific effect:

$$y_{i,t} - y_{i,t-1} = \alpha(y_{i,t-1} - y_{i,t-2}) + \beta'(X_{i,t} + X_{i,t-1}) + (\varepsilon_{i,t} - \varepsilon_{i,t-1}). \tag{15}$$

The use of instruments is required to deal with the correlation, by construction, of the new error term $\varepsilon_{i,t} - \varepsilon_{i,t-1}$ with the lagged dependent variable $y_{i,t-1} - y_{i,t-2}$ and with the likely endogeneity of the explanatory variables. Under the tested assumptions that the error term (ε) is not serially correlated and the explanatory variables (X) are weakly exogenous (the explanatory variables are assumed to be uncorrelated with future realizations of the error term), appropriately lagged values of the regressors can be used as instruments, as specified in the following moment conditions:

$$E[y_{i,t-s} \cdot (\varepsilon_{i,t} - \varepsilon_{i,t-1})] = 0 \quad \text{for } s \geq 2;\ t = 3, \ldots, T \tag{16}$$

$$E[X_{i,t-s} \cdot (\varepsilon_{i,t} - \varepsilon_{i,t-1})] = 0 \quad \text{for } s \geq 2;\ t = 3, \ldots, T. \tag{17}$$

The GMM estimator based on these conditions is referred to as the *difference estimator*.

There are, however, conceptual and statistical shortcomings with this difference estimator. It eliminates the cross-country relationship between national policies and per capita GDP growth, which is of conceptual interest. Statistically, when the regressors in equation 15 are persistent, lagged levels of X and y are weak instruments. Instrument weakness influences the asymptotic and small-sample performance of the difference estimator. Asymptotically, the variance of the coefficients rises. In small samples, weak instruments can produce biased coefficients.

To reduce the potential biases and imprecision associated with the usual difference estimator, Arellano and Bover (1995) and Blundell and Bond (1997) develop a system of regressions in differences and levels. The instruments for the regression in differences are the same as those above. The instruments for the regression in levels are the lagged differences of the corresponding variables. These are appropriate instruments under the following additional assumption: although there may be correlation between the levels of the right-hand-side variables and the country-specific effect in equation 14, there is no correlation between the differences of these variables and the country-specific effect. This assumption results from the following stationarity property:

$$E[y_{i,t+p} \cdot \eta_i] = E[y_{i,t+q} \cdot \eta_i] \text{ and } E[X_{i,t+p} \cdot \eta_i] = E[X_{i,t+q} \cdot \eta_i] \quad \text{for all } p \text{ and } q. \tag{18}$$

The additional moment conditions are.

$$E[(y_{i,t-s} - y_{i,t-s-1})\cdot(\eta_i + \varepsilon_{i,t})] = 0 \quad \text{for } s = 1 \tag{19}$$

$$E[(X_{i,t-s} - X_{i,t-s-1})\cdot(\eta_i + \varepsilon_{i,t})] = 0 \quad \text{for } s = 1. \tag{20}$$

Thus the moment conditions presented in equations 16, 17, 19, and 20 are used with a GMM estimator to generate consistent and efficient parameter estimates.

Consistency of the GMM estimator depends on the validity of the instruments. To address this issue two specification tests were considered, as suggested by Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1997). The first is a Sargan test of overidentifying restrictions to test the overall validity of the instruments by analyzing the sample analog of the moment conditions used in the estimation. The second test examines the hypothesis that the error term $\varepsilon_{i,t}$ is not serially correlated. In both the difference regression and the system regression the differenced error term is tested for second-order serial correlation (by construction, the differenced error term is probably first-order serially correlated even if the original error term is not). This system estimator is used to assess the impact of policies on economic growth. In addition to the system estimator, the analyses use purely cross-section, ordinary least squares regressions with one observation per country, the pure different estimator described above, and the panel estimator with only the level component of the system estimator. All yield similar results and parameter values (Levine, Loayza, and Beck 2000).

References

Ades, Alberto, and Edward Glaeser. 1999. "Evidence on Growth, Increasing Returns, and the Extent of the Market." *Quarterly Journal of Economics* 114(3):1025–46.

Aghion, P., and P. Howitt. 1998. *Endogenous Growth Theory*. Cambridge, Mass.: MIT Press.

Alesina, Alberto, Reza Baqir, and William Easterly. 1999. "Public Goods and Ethnic Divisions." *Quarterly Journal of Economics* 114(4):1243–84.

Arellano, Manuel, and Bond, Stephen. "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, Manuel, and Bover, Olympia. 1995. "Another Look at the Instrumental-Variable Estimation of Error-Components Models." *Journal of Econometrics* 68(1):29–52.

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

Bairoch, Paul. 1993. *Economics and World History: Myths and Paradoxes.* Chicago: University of Chicago Press.

Baldwin, Richard E. 1998. "Global Income Divergence, Trade and Industrialization. The Geography of Growth Take-offs." NBER Working Paper Series No. 6458.1. NBER, Washington, D.C.

Bank for International Settlements. 1996. *66th Annual Report.* BIS, Basel.

Barro, Robert, and Xavier Sala-i-Martin. 1992. "Convergence." *Journal of Political Economy* 100(2):223–51.

———. 1995. *Economic Growth.* New York: McGraw-Hill.

Barro, Robert, J. Mankiw, N. Gregory, and X. Sala-i-Martin. 1995. "Capital Mobility in Neoclassical Models of Growth." *American Economic Review* 85:103–15.

Baumol, William J. 1986. "Productivity Growth, Convergence, and Welfare. What the Long Run Data Show." *American Economic Review* 76(5):1072–85.

Beck, Thorsten. Ross Levine, and Norman Loayza. 2000. "Finance and the Sources of Growth." *Journal of Financial Economics* 58(1–2):261–300.

Becker, Gary S., Kevin M. Murphy, and Robert Tamura. 1990. "Human Capital, Fertility, and Economic Growth." *Journal of Political Economy* 98(5):S12–S37.

Benabou, Roland. 1993. "Workings of a City. Location, Education, and Production." *Quarterly Journal of Economics* 108:619–52.

———. 1996. "Heterogeneity, Stratification, and Growth. Macroeconomic Implications of Community Structure and School Finance." *American Economic Review* 86(3):584–609.

Benhabib. J., and M. Spiegel. 1994. "Role of Human Capital in Economic Development. Evidence from Aggregate Cross-Country Data." *Journal of Monetary Economics* 34:143–73.

Bils, Mark, and P. Klenow. 1996. "Does Schooling Cause Growth?" *American Economic Review* 90(5.):1160–83.

Blomstrom, Magnus, R. Lipsey, and M. Zejan. 1996. "Is Fixed Investment the Key to Economic Growth?" *Quarterly Journal of Economics* 111(1):269–76.

Blundell, Richard, and Bond, Stephen. 1997. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." University College London Discussion Paper 97-07.

Borjas, George J. 1992. "Ethnic Capital and Intergenerational Mobility." *Quarterly Journal of Economics* 107:123–50.

———. 1995. "Ethnicity, Neighborhoods, and Human Capital Externalities." *American Economic Review* 85(3):365–90.

———. 1999. *Heaven's Door. Immigration Policy and the American Economy.* Princeton, N.J.: Princeton University Press.

Borjas, George J., Stephen G. Bronars, and Stephen J. Trejo. 1992. "Self Selection and Internal Migration in the United States." *Journal of Urban Economics* 32:159–85.

Bouillon, César, Arianna Legovini, and Nora Lustig. 1999. "Rising Inequality in Mexico, Returns to Household Characteristics and the 'Chiappas' Effect." Inter-American Development Bank, mimeo.

Boyd, John H., Ross Levine, Bruce D. Smith. 2001. "The Impact of Inflation on Financial Sector Performance." *Journal of Monetary Economics*. Forthcoming.

Brookings Institution Center on Urban and Metropolitan Policy. 1999. *A Region Divided. The State of Growth in Greater Washington.* Washington, D.C.: Brookings Institution.

Bruno, Michael, and Easterly, William. 1998. "Inflation Crises and Long-run Growth." *Journal of Monetary Economics* 41:3–26.

Burnside, Craig. 1996. "Production Function Regressions, Returns to Scale and Externalities." *Journal of Monetary Economics* 37:177–200.

Carrington, William J., and Enrica Detragiache. 1998. "How Big Is the Brain Drain?" International Monetary Fund Working Paper 98/102.

Carroll, C. D., and D. N. Weil. 1993. "Saving and Growth. A Reinterpretation." Carnegie-Rochester Series on Public Policy.

Ciccone, Antonio, and Robert E. Hall. 1996. "Productivity and the Density of Economic Activity." *American Economic Review* 86(1):54–70.

Collier, Paul, David Dollar, and Nicholas Stern. 2000. "Fifty Years of Development." World Bank, mimeo.

Costello, Donna M. 1993. "Cross-Country, Cross-Industry Comparison of Productivity Growth." *Journal of Political Economy* 101:207–22.

Christiansen, L. R. D. Cummings, and D. Jorgenson. 1980. "Economic Growth, 1947—1973: An International Comparison." In J. W. Kendrick and B. Vaccara (eds.), *New Developments in Productivity Measurement and Analysis*. Studies in Income and Wealth, Vol. 41, Chicago: University of Chicago Press.

De Long, J. Bradford. 1988. "Productivity Growth, Convergence, and Welfare. Comment." *American Economic Review* 78(5):1138–54.

Denison, Edward F. 1962. *Sources of Economic Growth in the United States and the Alternatives Before Us*. New York. Committee for Economic Development.

———. 1967. *Why Growth Rates Differ*. Washington, D.C. Brookings Institution.

Dougherty, Christopher. 1991. *A Comparison of Productivity and Economic Growth in the G-7 Countries*. Ph.D. dissertation, Harvard University.

Easterly, William. 1999a. "The Ghost of Financing Gap. Evaluating the Growth Model of the International Financial Institutions." *Journal of Development Economics*, December.

———. 1999b. "Life During Growth." *Journal of Economic Growth* 4(3):239–76.

———. 2001. "The Lost Decades. Developing Countries' Stagnation in Spite of Policy Reform 1980–98." *Journal of Economic Growth* 6(2):135–57.

Easterly, William, and Sergio Rebelo. 1993. "Fiscal Policy and Economic Growth. An Empirical Investigation." *Journal of Monetary Economics* 32:417–58.

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* 32:459–83.

Elias, Victor J. 1990. *Sources of Growth. A Study of Seven Latin American Economies*. International Center for Economic Growth.

Fischer, Stanley. 1993. "The Role of Macroeconomic Factors in Growth." *Journal of Monetary Economics* 32:485–512.

Frarikel, J. A., and David Romer. 1999. "Does Trade Cause Growth?" *American Economic Review* 89:379–99.

Fujita, Masahisa, Paul Krugman, and Anthony Venables. 1999. *The Spatial Economy: Cities, Regions, and International Trade.* Cambridge, Mass.: MIT Press.

Grossman, G., and E. Helpman. 1991. "Quality Ladders in the Theory of Economic Growth." *Review of Economic Studies* 58:43–61.

Grubel, Herbert G., and Anthony Scott. 1977. *The Brain Drain. Determinants, Measurement and Welfare Effects.* Waterloo, Ontario: Wilfrid Laurier University Press.

Hadjmichael, Michael T., Michael Nowak, Robert Sharer, and Amor Tahari. 1996. *Adjustment for Growth. The African Experience.* Washington, D.C.: IMF.

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

Hansen, Gary D., and Edward Prescott. 1998. "Malthus to Solow." NBER Working Paper Series No. 6858.1–24.

Hanushek, Eric A., and Dennis D. Kimko. 2000. "Schooling, Labor-Force Quality, and the Growth of Nations." *American Economic Review* 90(5):1184–1208.

Harberger, Arnold C. 1998. "A Vision of the Growth Process." *American Economic Review* 88(1):1–32.

Holmes, T., and J. A. Schmitz Jr. 1995. "Resistance to New Technologies and Trade between Areas." *Federal Reserve Bank of Minneapolis Quarterly Review* 19:2–18.

International Labor Organization. 1995. *World Employment.* Geneva.

Jalan, Jyotsna, and Martin Ravallion. 1997. "Spatial Poverty Traps?" World Bank Policy Research Working Paper No. 1862.

Jones, Charles. 1995a. "R&D-Based Models of Economic Growth." *Journal of Political Economy* 103:759–84.

———. 1995b. "Time Series Tests of Endogenous Growth Models." *Quarterly Journal of Economics* 105(2):495–526.

———. 1997. "Comment on Peter Kienow and Andres Rodriguez-Clare, 'The Neoclassical Revival in Growth Economics. Has It Gone Too Far?'" *NBER Macroeconomics Annual* 12:73–103.

———. 1999. "Was an Industrial Revolution Inevitable? Economic Growth Over the Very Long Run." Stanford University, mimeo.

Jorgenson, Dale W. 1995. *Productivity.* Cambridge, Mass.: MIT Press.

Kendrick, John W., and Elliot S. Grossman. 1980. *Productivity in the United States: Trends and Cycles.* Baltimore, Md.: Johns Hopkins University Press.

King, Robert G., and Ross Levine. 1994. "Capital Fundamentalism, Economic Development and Economic Growth." *Carnegie-Rochester Conference Series on Public Policy* 40:259–92.

King, Robert G., and Sergio Rebelo. 1993. "Transitional Dynamics and Economic Growth in the Neoclassical Model." *American Economic Review* 83:908–31.

Klenow, Peter. 1998. "Ideas versus Rival Human Capital. Industry Evidence on Growth Models." *Journal of Monetary Economics* 42:2–23.

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

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

Klitgaard, Robert, and Amand Fitschen. 1997. "Exploring Income Variations across Traditional Authorities in KwaZulu-Natal, South Africa." *Development Southern Africa* 14(3).

Kongsamut, Piyabha, Sergio Rebelo, and Danyang Xie. 1997. "Beyond Balanced Growth." University of Rochester Working Paper.

Kosmin, Barry A., and Seymour P. Lachman. 1993. *One Nation under God. Religion in Contemporary American Society*. New York: Harmony Books.

Kremer, Michael. 1993. "O-ring Theory of Economic Development." *Quarterly Journal of Economics* 108:551–75.

Krueger, Alan B., and Mikael Lindahl. 1999. "Education for Growth in Sweden and the World." NBER Working Paper Series No. 7190.1–54.

Krugman, Paul R. 1991. *Geography and Trade*. Cambridge, Mass.: MIT Press.

———. 1995. *Development, Geography, and Economic Theory*. Cambridge, Mass.: MIT Press.

———. 1998. "Space. The Final Frontier." *Journal of Economic Perspectives* 12:161–74.

Krugman, Paul R., and Anthony J. Venables. 1995. "Globalization and the Inequality of Nations." *Quarterly Journal of Economics* 110(4):857–80.

Lipset, Seymour Martin. 1997. *American Exceptionalism. A Double Edged Sword*. New York: Norton.

Levine, Ross, and David Renelt. 1992. "A Sensitivity Analysis of Cross-Country Growth Regressions." *American Economic Review* 82(4):942–63.

Levine, Ross, Norman Loayza, and Thorsten Beck. 2000. "Financial Intermediation and Growth. Causality and Causes." *Journal of Monetary Economics* 46:31–77.

Lewis, W. A. 1954. "Economic Development with Unlimited Supplies of Labor." *Manchester School of Economic and Social Studies* 22(2):139–91.

Lucas, Robert E., Jr. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22:3–42.

———. 1990. "Why Doesn't Capital Flow from Rich to Poor Countries?" *American Economic Review, Papers and Proceedings* 80:92–96.

———. 1998. "The Industrial Revolution. Past and Future." Mimeo, University of Chicago.

Maddison. Angus. 1995. *Monitoring the World Economy, 1820–1992*. Paris: Development Centre of the Organisation for Economic Co-operation and Development.

Mankiw, N. Gregory. 1995. "The Growth of Nations." *Brookings Papers on Economic Activity* 1:275–326.

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

McGrattan, Ellen R. 1998. "A Defense of AK Growth Models." *Quarterly Review*, Federal Reserve Bank of Minneapolis 22:13–27.

Mellinger, Andrew D., Jeffrey D. Sachs, and John L. Gallup. 1999. "Climate, Water Navigability, and Economic Development." Harvard Center for International Development Working Paper No. 24.

Mokyr, Joel. 1992. *The Lever of Riches. Technological Creativity and Economic Progress.* New York: Oxford University Press.

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

Nehru, Vikram, and A. Dhareshwar. 1993. "A New Database on Physical Capital Stocks. Sources, Methodology, and Results." *Revista de Análisis Económico* 8(1):37–60.

Nordhaus, William. 1994. "Do Real Output and Real Wage Measures Capture Reality? The History of Lighting Suggests Not." Yale Cowles Foundation Discussion Paper 1078.

Parente, Stephen. 1994. "Technology Adoption, Learning-by-Doing, and Economic Growth." *Journal of Economic Theory* 63:346–69.

Parente, S. L., and E. C. Prescott. 1996. "Barriers to Technology Adoption and Development." *Journal of Political Economy* l02(2):298–321.

Patrinos, Harry Anthony. 1997. "Differences in Education and Earnings across Ethnic Groups in Guatemala." *Quarterly Review of Economics and Finance* 37(4):807–21.

Prescott, Edward. 1998. "Needed. A Theory of Total Factor Productivity." *International Economic Review* 39(3):525–32.

Pritchett, Lant. 1997. "Divergence, Big Time." *Journal of Economic Perspectives* 11(3): 3–17.

———. 1999. "The Tyranny of Concepts. CUDIE (Cumulated, Depreciated, Investment Effort) Is Not Capital." World Bank, mimeo.

———. 2000. "Patterns of Economic Growth. Hills, Plateaus, and Mountains." *World Bank Economic Review* 14(2):221–49.

———. 2001. "Where Has All the Education Gone?" *World Bank Economic Review.* Forthcoming.

Psacharopoulos, George. 1994. "Returns to Investment in Education. A Global Update." *World Development* 22:1325–1343.

Quah, Danny. 1993. "Galton's Fallacy and Tests of the Convergence Hypothesis." *Scandinavian Journal of Economics* 95(4):427–43.

Rappaport, Jordan, and Jeffrey Sachs. 1999. "The United States as a Coastal Nation." Harvard Center for International Development, mimeo.

Rauch, James E. 1993. "Productivity Gains from Geographic Concentration of Human Capital. Evidence from the Cities." *Journal of Urban Economics* 34:380–400.

Ravallion, Martin, and Jyotsna Jalan. 1996. "Growth Divergence Due to Spatial Externalities." *Economics Letters* 53:227–32.

Ravallion, Martin, and Quentin Wodon. 1998. "Poor Areas or Only Poor People?" World Bank, mimeo.

Ray, Debraj. 1998. *Development Economics.* Princeton, N.J.: Princeton University Press.

Rebelo, Sergio, and Nancy L. Stokey. 1995. "Growth Effects of Flat-Rate Taxes." *Journal of Political Economy* 103:519–50.

Rodrik, Dani. 1998. "Where Did all the Growth Go? External Shocks, Social Conflict and Growth Collapses." NBER Working Paper Series No. 6350.

Romer, Paul. 1986. "Increasing Returns and Long-Run Growth." *Journal* of *Political Economy* 94:1002–37.

———. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98:S71–S102.

———. 1995. "Comment on N. Gregory Mankiw, 'The Growth of Nations'" *Brookings Papers on Economic Activity* 1:313–20.

Rostow, W. W. 1960. *The Stages of Economic Growth. A Non-Communist Manifesto.* Cambridge: Cambridge University Press.

Shleifer, Andrei, and Robert Vishny. 1993. "Corruption." *Quarterly Journal of Economics* 108(3):599–667.

Solow, R. 1956. "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics* 70:65–94.

———. 1957. "Technical Change and the Aggregate Production Function." *Review of Economics and Statistics* 39:312–20.

Todaro, Michael P. 2000. *Economic Development,* 7th ed. Reading, Mass.: Addison-Wesley.

Union Bank of Switzerland. 1994. *Prices and Earnings around the Globe.* Zurich: United Bank of Switzerland.

United Nations. 1996. *World Economic and Social Survey.*

United States Government. 1996. *1995 Statistical Abstract of US.* Washington, D.C.

World Bank. 1993. *The East Asian Miracle.* Washington, D.C.: World Bank.

———. 1995. *Latin America after Mexico. Quickening the Pace.* Washington, D.C.: World Bank.

———. 1999. *World Development Report.* Oxford: Oxford University Press.

———. 2000a. *Can Africa Claim the 21st Century?* Washington, D.C.: World Bank.

———. 2000b. *East Asia. Recovery and Beyond.* Washington, D.C.: World Bank.

———. 2000c. *The Road to Stability and Prosperity in South Eastern Europe.* Europe and Central Asia Region, Washington, D.C.: World Bank.

———. 2000/2001. *World Development Report. Attacking Poverty.* Washington, D.C.: World Bank.

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

THE WORLD BANK ECONOMIC REVIEW, VOL. 15, NO. 2 221–224

What have we learned from a decade of empirical research on growth?

Comment on "It's Not Factor Accumulation: Stylized Facts and Growth Models," by William Easterly and Ross Levine

Pete Klenow

William Easterly and Ross Levine document five stylized facts about growth and argue that they imply a bigger role for total factor productivity (TFP) and technology than for physical and human capital. I agree with the first four of their facts and believe facts 1 and 3 provide strong support for their conclusion that TFP should be the focus of growth research.

Fact 1: TFP Accounts for Most Income and Growth Differences

I would add another piece of evidence pointing to the same conclusion. Jasso, Rosenzweig, and Smith (2000) compare earnings of U.S. immigrants with their earnings in their country of origin. With adjustments for local purchasing power, the average immigrant earns 2.2 times as much in the United States as in their country of origin. That is 75 percent as big as the earnings gap between the average U.S. worker and the average worker in source countries, suggesting that 75 percent of the gap between U.S. and source country earnings cannot be explained by general human capital. Easterly and Levine attribute about 25 percent of the gap to physical capital per worker. That leaves about 50 percent accounted for by TFP.

TFP differences could reflect disembodied technology, human capital externalities, access to specialized or high-quality capital or intermediate goods, the degree of competition, or measurement error. Research has barely begun to quantify the contributions of each of these.

As important as TFP is for country differences, it seems less important for the overall upward trend in GDP per capita. Averaging across 98 countries, Klenow and Rodriguez-Clare (1997) attributed 70 percent of growth to physical and human capital.

Pete Klenow is a senior economist at the Federal Reserve Bank of Minneapolis.

Fact 2: Income Diverges over the Long Run

Whether income diverges over the long run is not so clear for more recent periods. Parente and Prescott (2000) find that the "East" grew much faster than the "West" during 1960–95. Using Summers and Heston data for 102 countries over 1960–90, I find a 2 percent *decline* in the population-weighted standard deviation of log per capita income.

Easterly and Levine argue that divergence is inconsistent with growth being driven by factor accumulation; because of diminishing returns, factors should tend to equalize rather than diverge. But differences in institutions can work against factor convergence. Certain policies (high tax rates, protectionism, weak property rights) can reduce accumulation of physical and human capital. Divergence in policies can therefore lead to divergence of factors. The West's institutions may have improved in the 19th century, producing a century of divergence. The East may have adopted better policies since 1960, leading to factor convergence since then.

This is not to say that divergence—even if traced to policy divergence—requires a factor interpretation. Policies can affect TFP, too. Similarly, one can offer TFP explanations for convergence episodes. Parente and Prescott (2000) note that all 20th-century growth miracles have occurred in countries starting far behind the richest countries, consistent with technology catch-up. And the later they take off, the faster they catch up.

Factors can diverge or converge, and so can TFP. Episodes of divergence and convergence need to be examined more closely to determine the roles of physical capital, human capital, and TFP. For the five Asian miracle economies, Young (1995, 2000) points to factor accumulation.

Fact 3: Factor Accumulation Is Persistent, Growth Is Not

The corollary to fact 3 is that TFP growth is not very persistent. Related, Easterly and others (1993) show that country policy differences are much more persistent than country growth rate differences. Why would policies be more persistent than TFP growth?

The answer may be that policies affect the *level* of a country's TFP, not its growth rate. And changes in policies may only temporarily affect a country's growth rate. Picture countries trending together, linked by technology diffusion. Countries with higher levels of income are the ones with better policies. Growth miracles are produced by dramatic improvements in policies, and growth disasters by deteriorating policies. China is a fast grower not because its institutions are among the best but because it has improved its institutions so much in the last two decades. If it does not reform further, its per capita income might level out at, say, 30th percentile in the world distribution.

In this view, TFP growth is not persistent because changes in policies are not persistent. Some countries follow up reforms with more reforms, but others go back on their reforms. This conjecture could be investigated by looking at the persistence of policy changes and how policy changes correlate with TFP growth.

If policy changes are not persistent, however, why is factor accumulation persistent? Schooling yields returns for decades. Schooling may respond not to current policies but to the much smoother "average" policy expected over future working life. Physical capital investments are shorter lived, but still last a decade or more on average.

In contrast, private returns to raising TFP may be shorter-lived and more sensitive to current policy. A firm improving its efficiency may gain for only a few years, until competitors imitate it. Changes in market share between efficient and less efficient firms may be reversed when policies are reversed.

Fact 4: Economic Activity Is Highly Concentrated Geographically

As Easterly and Levine document, economic activity is highly concentrated even within countries. They argue that, without differences in TFP, factors should spread out evenly because of diminishing returns. That factors do not spread out evenly suggests differences in TFP, perhaps owing to technology and externalities.

But just as policy divergence could lead to factor divergence, policy differences can lead to geographic concentration within countries. Take the coastal areas of China. These areas were opened to foreign trade much earlier and more extensively than the rest of China, attracting huge inflows of labor and capital. For another example, Holmes (1998) documents heavy concentration of manufacturing in right-to-work states within the United States.

If land is not a very important factor, then lots of concentration can result from modest differences in policy or geography or technology. Suppose land's share is 5 percent, as in Lucas (forthcoming). If TFP differs across locations by a factor of 1.4 (compared to say 5 or 6 across countries), then output per unit of land will differ by a factor of 1,000 to equate returns to physical and human capital. If TFP differs by a factor of 1.6, output per unit of land will differ by a factor of 10,000.

Fact 5: National Policies Affect Long-Run National Growth Rates

Easterly and Levine show that country growth rates are correlated with country policy variables such as schooling, openness to trade, the size of government, the black market premium, and private credit. They are keenly aware that some of these variables may result from growth rather than cause it, so they are careful to instrument with lags. But even lagged variables can be subject to reverse causality. For example, increases in private credit could result from higher expected future growth.

If establishing any causal effect is difficult, establishing a *long-run* effect on the growth rate is harder still. The data could be revealing temporary rather than permanent growth effects. Consistent with this, many policies are correlated with growth only when simultaneously controlling for initial income. Openness to trade, for example, might have a positive coefficient because it facilitates technology diffusion. But once a country closes in on the technology frontier, openness will not continue to keep the growth rate high, only the level of income high. The fact that policies are more persistent than growth rates is consistent with policies affecting long-run levels more than long run growth rates.

References

Easterly, William, Michael Kremer, Lant Pritchett, and Lawrence H. Summers. 1993. "Good Policy or Good Luck? Country Growth Performance and Temporary Shocks." *Journal of Monetary Economics* 32(3):459–83.

Holmes, Thomas J. 1998. "The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders." *Journal of Political Economy* 106(4):667–705.

Jasso, Guillermina, Mark R. Rosenzweig, and James P. Smith. 2000. "The Earnings of U.S. Immigrants: Skill Transferability and Selectivity." Department of Economics, New York University, New York, N.Y.

Klenow, Peter J., and Andres Rodriguez-Clare. 1997. "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?" *Macroeconomics Annual 1*. Cambridge, Mass.: National Bureau of Economic Research.

Lucas, Robert E. Forthcoming.. "Externalities and Cities." *Review of Economic Dynamics*.

Parente, Stephen L., and Edward C. Prescott. 2000. *Barriers to Riches*. Cambridge, Mass.: MIT Presss.

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

———. 2000. "Gold into Base Metals: Productivity Growth in the People's Republic of China during the Reform Period." NBER Working Paper 7856. National Bureau of Economic Research, Cambridge, Mass.

THE WORLD BANK ECONOMIC REVIEW, VOL. 15, NO. 2 225–227

What have we learned from a decade of empirical research on growth?

Comment on "It's Not Factor Accumulation: Stylized Facts and Growth Models," by William Easterly and Ross Levine

Paul Romer

When economists in the 1950s and 1960s used growth models to understand the experience of developing countries, they allowed for the possibility of technology differences between developing countries and the United States. But because they did not have a good theory for talking about the forces that determined the level of the technology—in the United States any more than in developing countries—technology factors tended to be pushed into the background in policy discussions.

Modeling the Technology Factor

In the 1980s, several economists began to develop formal models of this technology factor, which has conventionally been designated A. We used those models to think about the behavior of A in the United States and in other countries. These "new growth" models had striking implications. Output per capita could diverge. Capital and skilled workers might flow from poor to rich countries. Trade in goods or investment decisions by firms could influence the diffusion of A between developed and developing countries. All this meant that government policies that affect incentives for firms could have big effects on economic outcomes.

The new growth theory of the 1980s generated a counterreaction in the 1990s that Pete Klenow and Andres Rodriguez-Clare have called the "neoclassical revival." Proponents of the neoclassical view argued that for purposes of explaining cross-country variation in levels of income or rates of growth, economists could return to the framework of the 1960s and append the strong assumption that A is the same in all countries. In this approach, the cross-country variation in the level and rate of economic development could be understood entirely in terms of differences in the level and rate of accumulation of traditional inputs: physical capital, human capital, and unskilled labor.

Paul Romer is professor of economics at Stanford University.

Attacking Business as Usual in Empirical Growth Theory

The article by William Easterly and Ross Levine is part of the next swing in the scholarly pendulum. It moves away from the critical assumption in the neoclassical revival that the level of technology is the same in all countries. Equally important (if not more so), it makes an implicit case for moving away from a narrow focus on testing models. In its place, the authors challenge economists to understand what happens in the countries that they study.

At a substantive level, they suggest that there is abundant evidence that something such as the level of the technology does vary across countries. That much is clear. But at a methodological level, they are doing more. They attack business as usual in empirical growth theory. To avoid the threat that a wide variety of evidence would pose to the neoclassical revival, economists who supported this program advocated a narrow methodology based on model testing. Pick a few summary statistics generated from a specially selected data source. (Typically they are partial correlations generated by running a regression on a cross-country data set.) Use strong theoretical priors to restrict attention to a very small subset of all possible models. Then show that one of the models from within this narrow set fits the data and, if possible, show that there are other models that do not. Having tested and rejected some models so that the exercise looks like it has some statistical power, accept the model that fits the data as a "good model."

The obvious problem with this approach is that many models are consistent with the few correlations that emerge from a single data set. Many alternative models are never considered in the standard model testing exercise, so it has no power against these alternatives, no ability to weigh their plausibility relative to that of the "good model." Suppose an economist runs a regression conditioned on other variables and finds that countries with lower initial income grow more rapidly. Within a neoclassical model with an exogenously determined level of A, this finding can be interpreted as evidence of diminishing returns to physical capital or human capital because of exogenous variation in rates of investment in physical or human capital. But it is also possible that the technology is lower in the country that starts at a lower level of development and grows faster as better technology diffuses there. If the economist looks only at the cross-country regression evidence, there is nothing that would raise questions about the initial assumption of identical technologies in all countries.

But as Easterly and Levine point out, if you bring other evidence, such as the pattern of flows of people between countries, the identical technologies model no longer fits. For someone who wants to maintain an unreasonable prior assumption, the advantage of a narrow focus on one piece of data is that it does not threaten the convenient theoretical framework built on this prior. It is possible to go through the motions of doing science, testing various theories and rejecting some in favor of others. But far from advancing the science, this approach is a dead end. It does not allow for rejecting or modifying prior beliefs that simply turn out to be wrong.

Of course, economists who would like to stay in the business of doing these narrow, model-testing exercises have a response. They say that one model should be used to explain cross-country regressions and another to explain factor flows between countries. If they were defense attorneys in a criminal case, they could provide a different theory to explain every specific piece of evidence at the crime scene. But they would not be able to tell a consistent story about what actually happened.

No amount of methodological obfuscation about "as-if" modeling, "parsimony," "pushing the limits of benchmark models," and the like should be allowed to hide the fact that, like judges and jurors, policymakers and economists have to make judgments about what actually happened. Otherwise, they have no basis for making well-informed decisions about what to do next. Easterly and Levine show economists how they can best contribute in this effort—assemble as much evidence as possible and search for a consistent theoretical explanation that fits it all.

At a substantive level, the authors suggest that there is abundant evidence that something like the level of technology does vary between countries. Admitting this possibility could have important implications for policy. If economists take the classical model seriously as a description of the development process—that is, if they accept that an economy is well described as a competitive equilibrium with an exogenously given level of technology—then Harberger's pioneering analysis of the welfare costs of distortions tells us that bad government policies imply tiny welfare losses. Economists would then be left with a striking conclusion, one that few practitioners would take seriously: Countries are poor not because of bad policies but because of exogenous preference differences that cause them to accumulate less physical and human capital. The risk is not that practitioners will take this conclusion seriously, but that they will dismiss formal economic theory and empirical work as flawed beyond repair.

THE WORLD BANK ECONOMIC REVIEW, VOL. 15, NO. 2 229–272

What have we learned from a decade of empirical research on growth?

Growth Empirics and Reality

William A. Brock and Steven N. Durlauf

This article questions current empirical practice in the study of growth. It argues that much of the modern empirical growth literature is based on assumptions about regressors, residuals, and parameters that are implausible from the perspective of both economic theory and the historical experiences of the countries under study. Many of these problems, it argues, are forms of violations of an exchangeability assumption that implicitly underlies standard growth exercises. The article shows that these implausible assumptions can be relaxed by allowing for uncertainty in model specification. Model uncertainty consists of two types: theory uncertainty, which relates to which growth determinants should be included in a model; and heterogeneity uncertainty, which relates to which observations in a data set constitute draw from the same statistical model. The article proposes ways to account for both theory and heterogeneity uncertainty. Finally, using an explicit decision-theoretic framework, the authors describe how one can engage in policy-relevant empirical analysis.

There are more things in heaven and earth, Horatio,
Than are dreamt of in your philosophy. —William Shakespeare
Hamlet, act 1, scene 5

The objective of this article is ambitious—to outline a perspective on empirical growth research that will both address some of the major criticisms to which this research has been subjected and facilitate policy-relevant empirics. It is no exaggeration to say that the endogenous growth models pioneered in Romer (1986, 1990) and Lucas (1988) have produced a sea change in the sorts of questions around which macroeconomic research is focused. In empirical macroeconomics, efforts to explain cross-country differences in growth behavior since World War II have become a predominant area of research. The implications of this work for policymakers are immense. For example, strong links exist between national growth performance and international poverty and inequality. Differ-

William A. Brock and Steven N. Durlauf are with the Department of Economics, University of Wisconsin. Their e-mail addresses are *wbrock@ssc.wisc.edu* and *sdurlauf@ssc.wisc.edu*. The authors thank the National Science Foundation, John D. and Catherine T. MacArthur Foundation, Vilas Trust, and Romnes Trust for financial support. They thank François Bourguignon both for initiating this research and for helpful suggestions, as well as Gernot Doppelhofer, Paul Evans, Cullen Goenner, Andros Kourtellos, Artur Minkin, Eldar Nigmatullin, Xavier Sala-i-Martin, Robert Solow, seminar participants at Carnegie-Mellon and Pittsburgh, and three anonymous referees for helpful comments. Chih Ming Tan has provided superb research assistance. Special thanks go to William Easterly and Ross Levine for sharing data and helping with replication of their results. An earlier version of this article was presented at the World Bank conference "What Have We Learned from a Decade of Empirical Research on Growth?" held on 26 February 2001.

ences in per capita income across countries are substantially larger than those within countries; Schultz (1998) concludes that two-thirds of (conventionally measured) inequality across individuals internationally is due to intercountry differences, so that efforts to reduce international inequality naturally focus on cross-country growth differences. In turn, the academic community has used this new empirical work as the basis for strong policy recommendations. A good example is Barro (1996). Based on a linear cross-country growth regression of the type so standard in this literature, Barro (1996, p. 24) concludes that

> The analysis has implications for the desirability of exporting democratic institutions from the advanced western economies to developing nations. The first lesson is that more democracy is not the key to economic growth. . . . The more general conclusion is that advanced western countries would contribute more to the welfare of poor nations by exporting their economic systems, notably property rights and free markets, rather than their political systems.

Yet there is widespread dissatisfaction with conventional empirical methods of growth analysis. Many critiques of growth econometrics have appeared in recent years. Typical examples include Pack (1994, pp. 68–69) who described a litany of problems with cross-country growth regressions:

> Once both random shocks and macroeconomic policy variables are recognized as important, it is no longer clear how to interpret many of the explanations of cross-country growth. . . . Many of the right hand side variables are endogenous. . . . The production function interpretation is further muddled by the assumption that all countries are on the same international production frontier . . . regression equations that attempt to sort out the sources of growth also generally ignore interaction effects. . . . The recent spate of cross-country growth regressions also obscures some of the lessons that have been learned from the analysis of policy in individual countries.

Another is Schultz (1999, p. 71): "Macroeconomic studies of growth often seek to explain differences in economic growth rates across countries in terms of levels and changes in education and health human capital, among other variables. However, these estimates are plagued by measurement error and specification problems." In fact, it seems no exaggeration to say that the growth literature in economics is notable for the large gaps that persist between theory and empirics. A recent (and critical) survey of the empirical literature, Durlauf and Quah (1999, p. 295), concludes that

> the new empirical growth literature remains in its infancy. While the literature has shown that the Solow model has substantial statistical power in explaining cross-country growth variation, sufficiently many problems exist with this work that the causal significance of the model is not clear. Fur-

ther, the new stylized facts of growth, as embodied in nonlinearities and distributional dynamics have yet to be integrated into full structural econometric analysis.

Our purposes in this article are threefold. First, we attempt to identify some general methodological problems that we believe explain the widespread mistrust of growth regressions. Although the factors we identify are not exhaustive, they do represent many of the most serious criticisms of conventional growth econometrics of which we are aware. These problems are important enough to at best seriously qualify and at worst invalidate many of the standard claims made in the new growth literature concerning the identification of economic structure. In particular, we argue that causal inferences as conventionally drawn in the empirical growth literature require certain statistical assumptions that may easily be argued to be implausible. This assertion holds from the perspective of both economic theory and the historical experiences of the countries under study. We further argue that a major source of skepticism about the empirical growth literature, and one that incorporates many of the usual criticisms, is the failure of certain statistical conditions representing forms of a property known as exchangeability to hold in conventional empirical growth exercises.

Second, we argue that the exchangeability failures underlying many criticisms of growth models may be constructively dealt with through explicit attention to model uncertainty in the formulation of growth regressions; see Temple (2000) for a complementary analysis. What we mean is the following: In estimating a particular regression, the inferences are made conditional both on the data and on the specification of the regression. The exchangeability objection to a regression amounts to questioning whether the specification of the regression is correct. The assumption that a particular specification is correct can be relaxed by treating model specification as an additional unknown feature of the data, that is, by explicitly incorporating model uncertainty in the statistical analysis. In taking this approach, we follow some important recent developments in the empirical growth literature—Fernandez, Ley, and Steel (1999) and Doppelhofer, Miller, and Sala-i-Martin (2000)—in endorsing the use of Bayesian methods to address explicitly the model uncertainty that we believe underlies the mistrust of conventional growth regressions. This analysis does not address all the criticisms we describe in the first part of the article. In particular, we argue that questions of causality versus correlation, which are of first-order importance in interpreting growth regressions, may only be addressed using substantive information that originates outside the models under analysis. Nevertheless, accounting for exchangeability can strengthen the confidence that may be attached to causal interpretations of regression exercises.

Third, we argue that the appropriate use of empirical growth analyses for policy analysis requires an explicit decision-theoretic formulation. Current empirical practice in growth is therefore not "policy-relevant" in the sense that the policy inferences of a given data analysis are decoupled from the analysis itself. For example, one often sees a statistically insignificant coefficient used as evidence that some policy is not important for growth or, conversely, the assertion that statistical significance establishes the importance of some policy. We argue that these types of claims are not appropriate. Ideally, empirical growth exercises should employ cross-country growth data to compute predictive distributions for the consequences of policy outcomes, distributions that can then be combined with a policymaker's welfare function to assess alternative policy scenarios. A decision-theoretic approach to evaluating growth regressions can provide a better measure of the level of the evidence inherent in the available data, especially for the construction of policy-relevant predictive structures through empirical growth analyses.

The title of this article intentionally echoes the classic Sims (1980) critique of macroeconometric models. The growth literature does not suffer from the exact type of "incredible" assumptions (Sims 1980) that were required to identify economic structure through 1960s-style simultaneous equation models and whose interpretation Sims was attacking. Yet this literature does rely on assumptions that may be argued to be equally dubious and whose implausibility renders the inferences typically claimed by empirical workers to be equally suspect.[1] As will be clear from our discussion, this article only begins to scratch the surface of a policy-relevant growth econometrics. Our hope is that the ideas herein will facilitate new directions in growth research.

At the same time, our purpose is not to argue that statistical analyses of cross-country growth data are incapable of providing insights. Regression and other forms of statistical analysis have several critical roles in the study of growth. One role is the identification of interesting data patterns, patterns that can both stimulate economic theory and suggest directions along which to engage in country-specific studies. Quah's work (1996a, 1996b, 1997) is exemplary in this regard. However, we focus explicitly on the role of empirical work in formulating policy recommendations. In particular, a second goal of this article is to explore how one can, by casting empirical analysis in an explicitly decision-theoretic framework, develop firmer insights into the growth process.

Throughout we will take an eclectic stance on how one should go about data analysis. Many of our ideas are derived from the Bayesian statistics literature. Yet the basic arguments we make are relevant to frequentist analyses. Our view of data analysis is essentially pragmatic. Data analyses of the sort that are conventional in economics should be thought of as evidence-gathering exercises aimed at facilitating the evaluation of hypotheses and the development of policy-relevant predictions for future trajectories of variables of interest. For example, one starts with a proposition such as "the level of democracy in a country causally influences the level of economic growth." Once this statement is mathemati-

1. A number of the issues we raise echo, at least in spirit, Freedman (1991, 1997), who has made serious criticisms of the use of regressions to uncover causal structure in the social sciences.

cally instantiated (which means that *ceteris paribus* conditions are formalized, a more or less convincing theoretical model or set of models of causal influence is formulated in a form suitable for econometric implementation, etc.), the purpose of an empirical exercise is to see whether the statement is more or less plausible once the analysis has been conducted. The success or failure of an empirical exercise rests on whether one's prior views of the proposition have been altered by the analysis and on whether the level of uncertainty around a conclusion is low enough for the conclusion to be of policy relevance. Our position is that one should evaluate statistical procedures on the basis of whether they successfully answer the questions for which they are employed; we are unconcerned, at least in this article, with abstract issues that distinguish frequentist and Bayesian approaches, for example.

Many of our criticisms of the empirical growth literature apply in principle to other empirical contexts. They take on particular force in the growth context because of the complexity of the objects under study, the poor data available for empirical growth work, and the qualitative nature of the theories that drive the new growth literature.

I. A Baseline Regression

The bulk of modern empirical work on growth has focused on cross-country growth regressions of the type pioneered by Barro (1991) and Mankiw, Romer, and Weil (1992). Although recent work has extended growth analysis to consider panels (Evans 1998; Islam 1995; Lee, Pesaran, and Smith 1997), the arguments we make relating to conventional empirical growth practice as well as our proposed alternative approach are generally relevant to that context too, so long as cross-section variation is needed for parameter identification. Hence we focus on cross-sections.

A generic form for various cross-country growth regressions is

$$g_i = X_i\gamma + Z_i\pi + \varepsilon_i \tag{1}$$

where g_i is real per capita growth in economy i over a given period (typically measured as the change in per capita income between the beginning and end of the sample divided by the number of years that have elapsed). We have divided the regressors into two types. X_i represents variables whose presence is suggested by the Solow growth model: a constant, initial income and a set of country-specific savings and population growth rate controls. The Solow model is often treated as a baseline from which to build up more elaborate growth models, hence these variables tend to be common across studies. Z_i, in contrast, consists of variables chosen to capture additional causal growth determinants that a researcher believes are important and so generally differs across analyses.[2]

2. See Galor (1996) for a discussion of the implications of different growth theories for convergence and Bernard and Durlauf (1996) for an analysis of the economic and statistical meanings of convergence.

Though this regression is typically applied to national aggregates, it can in principle be applied to regions or sectors once g_i is reinterpreted as a vector of growth rates within a country. This is particularly important for policy analysis when a given policy may affect different regions or population groups differently. Our conjecture is that such decompositions are important when evaluating growth policies with significant distributional consequences.

In our discussion we assume that the motivation for the estimation of a regression of the type given in equation 1 is policy driven. Specifically, we assume that a policymaker is interested in using this equation to advise some country i on whether it should change some policy instrument z. If the policymaker's objective function depends on the growth rate in country i, he will presumably need to understand the country's overall growth process and hence to make inferences about a number of aspects of equation 1 in addition to π_z, the coefficient on the policy instrument. We return to this issue in section VI.

II. Econometric Issues

In this section we discuss three problems with the use of the baseline equation 1 in policymaking or other exercises in which one wishes to give a structural interpretation to this regression. These problems all, at one level, occur because of violations of the assumptions necessary to estimate equation 1 using ordinary least squares (OLS) and interpret the estimated equation as the structural model of growth dynamics implied by the augmented Solow model. Each of these criticisms ultimately reduces to questioning whether growth regressions as conventionally analyzed can provide the causal inferences that motivate such analyses. As discussed in the introduction, growth regressions have been subjected to a wide range of criticisms from many authors. We do not claim that any of the criticisms are necessarily original to us; instead, we believe our contribution in this section lies in the way we organize and unify these criticisms.

Open-Endedness of Theories

A fundamental problem with growth regressions is determining what variables to include in the analysis. This problem occurs because growth theories are open-ended. By open-endedness, we refer to the idea that the validity of one causal theory of growth does not imply the falsity of another. So, for example, a causal relationship between inequality and growth has no implications for whether a causal relationship exists between trade policy and growth. As a result, well over 90 different variables have been proposed as potential growth determinants (Durlauf and Quah 1999), each of which has some ex ante plausibility. As there are at best about 120 countries available for analysis in cross-sections (the number may be far smaller as a result of missing observations on some covariates), it is far from obvious how to formulate firm inferences about any particular explanation of growth.

This issue of open-endedness has not been directly dealt with in the literature. Instead, a number of researchers have proposed ways to deal with the robustness of variables in growth regressions. The basic idea of this approach is to identify a set of potential control variables for inclusion in equation 1 as elements of Z_i. Inclusion of a variable in the final choice of Z_i requires that its associated coefficient prove to be robust with respect to the inclusion of other variables. Levine and Renelt (1992) introduced this idea to the growth literature, employing Edward Leamer's ideas on extreme bounds analysis (see Leamer 1983 and Leamer and Leonard 1983). In extreme bounds analysis, a coefficient is robust if the sign of its OLS estimate stays constant across a set of regressions representing different possible combinations of other variables. Sala-i-Martin (1997), arguing that extreme bounds analysis is likely to lead to the rejection of variables that do influence growth, proposes computing likelihood-weighted significance levels of coefficients across alternative regressions.

These proposals for dealing with the plethora of growth theories are useful, but neither is definitive as a way to evaluate model robustness.[3] The reason is simple. In these approaches a given coefficient will prove not to be robust if its associated variable is highly collinear with variables suggested by other candidate growth theories. This is obvious for the Sala-i-Martin approach, because collinearity affects significance levels. It is also true for extreme bounds analysis, in the sense that a given coefficient is likely to be highly unstable when alternative collinear regressors are included alongside its corresponding regressor. Hence these procedures will give sensible answers only if lack of collinearity is a "natural" property for a regressor that causally influences growth. Yet when one thinks about theories of how various causal determinants of growth are themselves determined, it is clear that collinearity is a property that one might expect to hold for important causal determinants of growth.[4] This is easiest to see by considering a recursive model for growth. Suppose that growth is causally determined by a single regressor, d_i, and that this regressor in turn depends causally on a third regressor, c_i, so that

$$
\begin{aligned}
g_i &= d_i \gamma_d + \varepsilon_i \\
d_i &= c_i \pi_c + \eta_i .
\end{aligned}
\tag{2}
$$

It is easy to construct cases (which will depend on the covariance structure of c_i, ε_i, and η_i) in which adding c_i to the growth equation will render d_i fragile.

Important recent papers by Doppelhofer, Miller, and Sala-i-Martin (2000) and Fernandez, Ley, and Steel (2001) have proposed ways to deal with regressor

3. Leamer's work on model uncertainty falls into two parts: a powerful demonstration of the importance of accounting for such uncertainty in making empirical claims, and a specific suggestion, extreme bounds analysis, for determining when regressors are fragile. The first constitutes a fundamental set of ideas. The second is a particular way of instantiating Leamer's deep ideas of accounting for model uncertainty and is more easily subjected to criticism. By analogy, Rawl's controversial use of minimax arguments to infer what rules are just in a society does not diminish the importance of his idea of the veil of ignorance. Economists have inappropriately used criticisms of extreme bounds analysis to ignore the conceptual issues raised by Leamer's work.

4. Leamer is quite clear on this point. See Leamer (1978, p. 172) for further discussion.

choice and hence at least indirectly with model open-endedness through the use of Bayesian model averaging techniques. We exploit the approach used in those papers and therefore defer discussion of them until section V.

Parameter Heterogeneity

A second problem with conventional growth analyses is the assumption of parameter homogeneity. The vast majority of empirical growth studies assume that the parameters that describe growth are identical across countries. This assumption is surely implausible. Does it really make sense to believe that a change in the level of a civil liberties index has the same effect on growth in the United States as in the Russian Federation? Although the use of panel data approaches to growth has addressed one aspect of this problem by allowing for fixed effects (Evans [1998] is particularly clear on this point), it has not addressed this more general question.

In some sense this criticism might seem unfair, as it presumably applies to any socioeconomic data set. After all, economic theory does not imply that individual units ought to be characterized by the same behavioral functions. That said, any empirical analysis necessarily will require a set of interpretable statistical properties that are common across observations; when homogeneity assumptions are or are not to be made is a matter of judgment. Our contention is that the assumption of parameter homogeneity is particularly inappropriate in studying complex heterogeneous objects, such as countries. See Draper (1997) for a general discussion of these issues.

Evidence of parameter heterogeneity has been developed in different contexts, such as in Canova (1999); Desdoigts (1999); Durlauf and Johnson (1995); Durlauf, Kourtellos, and Minkin (2000); Kourtellos (2000); and Pritchett (2000). These studies use very different statistical methods, but each suggests that the assumption of a single linear statistical growth model that applies to all countries is incorrect.[5] Put differently, the reporting of conditional predictive densities based on the assumption that all countries obey a common linear model may understate the uncertainty present when the data are generated by a family of models; Draper (1997) provides further analysis of this idea.

There has been substantial interest in the empirical growth literature in incorporating forms of parameter heterogeneity when panel data are available. Islam (1995) is an early analysis that allows constant terms to differ across country growth processes for a panel in which growth is measured in five-year intervals. In what appears to be the richest analysis of parameter heterogeneity to date, Lee, Pesaran, and Smith (1997) show how to allow for parameter heterogeneity for regressor slope parameters for a growth model employing annual data.

5. Conventional growth analyses give some attention to parameter heterogeneity between rich and poor countries: Barro (1996), for example, allows the effects of democracy on growth to differ between rich and poor countries.

The idea that panel data may be used to model rich forms of parameter heterogeneity is of course important; a comprehensive analysis is Pesaran and Smith (1995). However, this approach is of limited use in empirical growth contexts, because variation in the time dimension is typically small. This occurs for two reasons. First, many of the variables used as proxies for new growth theories do not vary over high frequencies. For some variables, such as political regime, this is true by their nature; for others, this is due to measurement. In any event, this means that cross-section variation must be used to uncover parameters. Second, there is a conceptual question of the appropriate time horizon over which to employ a growth model. High-frequency data will contain business cycle factors that are presumably irrelevant for long-run output movements. Hence it is difficult to see how annual or biannual data, for example, can be interpreted in terms of growth theories. In our view the use of long run averages has a powerful justification for identifying growth as opposed to cyclical factors.

Causality versus Correlation

A final source of skepticism about conventional growth empirics relates to a problem endemic to all structural inference in social science—the question of causality versus correlation. Many of the standard variables used to explain growth patterns—democracy, trade openness, rule of law, social capital, and the like—are as much outcomes of socioeconomic decisions and interrelationships as growth itself is. Hence there is an a priori case that the use of OLS estimates of the relationship between growth and such variables cannot be treated as structural any more than coefficients produced by OLS regressions of price on quantity can be. Yet the majority of empirical growth studies treat the various growth controls as exogenous variables and so rely on ordinary or heteroskedasticity-corrected least squares estimation. What is particularly ironic about the lack of attention to endogeneity is that it was precisely this lack of attention in early business cycle models that helped drive the development of rational expectations econometrics.

Recent econometric practice in growth has begun to employ instrumental variables to control for regressor endogeneity. This is particularly common for panel data sets where temporally lagged variables are treated as legitimate instruments. However, this trend toward using instrumental variables estimation has not satisfactorily addressed this problem. The reason is that the failure to account properly for the open-endedness of growth theories has important implications for the validity of instrumental variables methods.

What we mean by this is the following. For a regression of the form

$$y_i = R_i\gamma + \varepsilon_i \tag{3}$$

the use of some set of instrumental variables I_i as instruments for R_i requires, of course, that each element of I_i be uncorrelated with ε_i. In the growth literature this is not a condition typically employed to motivate the choice of instruments. Instead, instruments are typically chosen exclusively because they are in some sense exogenous, which operationally means that they are predetermined with respect to ε_i. Predetermined variables, however, are not necessarily valid instruments.

As discussed in Durlauf (2000), a good example of this pitfall can be found in Frankel and Romer (1996), which studies the relationship between trade and growth. Frankel and Romer argue that because trade openness is clearly endogenous, it is necessary to instrument the trade openness variable in a cross-country regression to consistently estimate the trade openness coefficient. To do this, they use a geographic variable, area, as an instrument and argue in favor of its validity that area is predetermined with respect to growth. Their argument that the instrument is predetermined is certainly persuasive. Nevertheless, it is hard to make an argument that it is a valid instrument. Is it plausible that country land size is uncorrelated with the omitted growth factors in their regression? The history and geography literatures are replete with theories of how geography affects political regime, development, and so on. For example, larger countries may be more likely to be ethnically heterogeneous, leading to attendant social problems. Alternatively, larger countries may have higher per capita military expenditures, which means relatively greater shares of unproductive government investment, higher distortionary taxes, or both. Our argument is not that any one of these links is necessarily empirically salient, but that the use of land area as an instrument presupposes the assumption that the correlations between land size and all omitted growth determinants are in total negligible. It is difficult to see how such an assumption can be defended when these omitted growth determinants are neither specified nor evaluated.

It is interesting to contrast the difficulties of identifying valid instruments in growth contexts with the relative ease with which this is done in rational expectations contexts. The reason for this difference is that rational expectations models are typically closed in the sense that a particular theory will imply that some combination of variables is a martingale difference with respect to some sequence of information sets. For the purposes of data analysis, rational expectations models therefore generate instrumental variables, that is, any variables observable at the time expectations are formed, whose orthogonality to expectation errors may be exploited to achieve parameter identification. Of course, rational expectation models can be faulted for imposing sufficiently wide-ranging restrictions on the economic behavior under study that some of the assumptions necessary for identification are not plausible; that is, for being insufficiently open-ended in the sense we have described. So the problems associated with theory open-endedness in growth are hardly nonexistent in other contexts.

III. Exchangeability

Inferences from any statistical model can only be made, of course, conditional on various prior assumptions that translate the data under study into a particular mathematical structure. One way to evaluate the plausibility of inferences drawn from empirical growth regressions is by assessing the plausibility of the assumptions made in making this translation. In the empirical growth literature it is easy to find examples where the assumptions employed to construct statistical models are clearly untenable. For example, researchers typically assume that the errors in a cross-section regression are jointly uncorrelated and orthogonal to the model's regressors.[6] Do they really wish to argue that no omitted factors exist that induce correlation across the innovations in the growth regressions associated with the model? More generally, it is easy to see that parameter heterogeneity and omitted variables, which, we argued in the previous section, are endemic to growth regressions, can each lead to a violation of the error uncorrelatedness assumption, the regressor orthogonality assumption, or both.

On the other hand, econometrics has a long tradition of identifying minimal sets of conditions under which coefficients and standard errors may be consistently estimated. Examples include the emphasis on orthogonality conditions between regressors and errors as the basis for OLS consistency (rather than the interpretation of the OLS estimators as the maximum likelihood estimates for a linear model with nonstochastic regressors and *i.i.d.* normal errors) or the use of mixing conditions to characterize when central limit theorems apply to dependent data (rather than the modeling of the series as a known autoregressive moving average process). Hence any critique of cross-country growth analyses that is based on the plausibility of particular statistical assumptions needs to argue that the violations of the assumptions invalidate the objectives of a given exercise.

In this section we argue that of the three econometric issues we have raised, the first two may be interpreted as examples of deviations of empirical growth models from a statistical "ideal" that allows for the sorts of inferences researchers wish to make in growth contexts. Our purpose is to establish a baseline for statistical growth models such that if a model does not meet this standard, a researcher needs to determine whether the reasons for this invalidate the goal of the empirical exercise. Hence the baseline does not describe a necessary requirement for empirical work, but instead helps define a strategy that we think empirical workers should follow in formulating growth models. When a model does not meet this standard, researchers should be prepared to argue that the violations of the standard do not invalidate the empirical claims they wish to make. This standard is based on a concept in probability known as *exchangeability*.[7]

6. In the subsequent discussion, we focus on OLS estimation of growth regressions. In the empirical growth literature examples can be found of heteroskedasticity corrections to relax assumptions of identical residual variances and instrumental variables to deal with violations of error/regressor orthogonality. Our discussion is qualitatively unaffected by either of these alternatives to OLS.

7. Bernardo and Smith (1994) provide a complete introduction to exchangeability. Draper and others (1993) develop a detailed argument on the importance of exchangeability to statistical inference. Our analysis is much indebted to their perspective.

Basic Ideas

A formal definition of exchangeability is as follows.

DEFINITION: EXCHANGEABILITY. A sequence of random variables η_i is exchangeable if, for every finite collection $\eta_1 \ldots \eta_K$ of elements of the sequence,

$$\mu(\eta_1 = a_1, \ldots, \eta_K = a_K) = \mu(\eta_{\rho(1)} = a_1, \ldots, \eta_{\rho(K)} = a_K)^8 \tag{4}$$

where $\rho(.)$ is any operator that permutes the K indices.

Exchangeability is typically treated as a property of the unconditional probabilities of random variables. In regression contexts, however, it is often more natural to think in terms of the properties of random variables conditional on some information set. For example, in a regression, one is interested in the properties of the errors conditional on the regressors. We therefore introduce a second concept, F-conditional exchangeability.[9]

DEFINITION: F-CONDITIONAL EXCHANGEABILITY. For a sequence of random variables η_i and a collection of associated random vectors F_i, η_i is F-conditionally exchangeable if, for every finite collection $\eta_1 \ldots \eta_K$ of elements of the sequence,

$$\mu(\eta_1 = a_1, \ldots, \eta_K = a_K | \underset{\sim}{F}) = \mu(\eta_{\rho(1)} = a_1, \ldots, \eta_{\rho(K)} = a_K | \underset{\sim}{F})$$

where $\rho(\cdot)$ is any operator that permutes the K indices and $\underset{\sim}{F} = \{F_1 \ldots F_K\}$.

If $F_i = \phi \; \forall \; i$, the empty set, F-conditional exchangeability reduces to exchangeability.

Associated with exchangeability and F-conditional exchangeability is the idea of partial exchangeability.

DEFINITION: PARTIAL EXCHANGEABILITY. A sequence of random variables η_i is partially exchangeable with respect to a sequence of random vectors Y_i if, for every finite collection $\eta_1 \ldots \eta_K$ of elements of the sequence,

$$\mu(\eta_1 = a_1, \ldots, \eta_K = a_K | Y_i = \bar{Y} \; \forall \; i \in \{1 \ldots K\}) = \mu(\eta_{\rho(1)} = a_1, \ldots, \eta_{\rho(K)} = a_K | Y_i = \bar{Y} \; \forall \; i \in \{1 \ldots K\}) \tag{6}$$

where $\rho(\cdot)$ is any operator that permutes the K indices.

The key difference between exchangeability and partial exchangeability is the

8. Throughout, $\mu(.)$ is used to denote probability measures.

9. F-conditional exchangeability was originally defined in Kallenberg (1982). Ivanoff and Weber (1996) provide additional discussion. The notion of F-conditional exchangeability is rarely employed in the statistics literature and is not mentioned in standard textbooks such as Bernardo and Smith (1994). We believe the reason for this is that exchangeability analyses in the statistics literature generally focus on whether the units under study are exchangeable, rather than whether they are conditional on certain characteristics, the more natural notion in economic contexts.

conditioning on common values of some random vectors Y_is associated with the η_is in the partial exchangeability case. If Y_i is a discrete variable, partial exchangeability implies that a sequence may be decomposed into a finite or countable number of exchangeable subsequences.

Even though F-conditional exchangeability of model errors constitutes a stronger assumption than is needed for many of the interpretations of OLS, this exchangeability condition is nevertheless useful as a benchmark in the construction and assessment of statistical models. We make this claim for two reasons. First, this exchangeability concept helps organize discussions of the plausibility of the invariance of conditional moments that lie at the heart of policy relevant predictive exercises. Draper (1987, p. 458) describes the critical role of exchangeability in any predictive exercise:

> Predictive modeling is the process of expressing one's beliefs about how the past and future are connected. These connections are established through *exchangeability judgments*: with what aspects of past experience will the future be more or less interchangeable, after conditioning on relevant factors? It is not possible to avoid making such judgments; the only issue is whether to make them explicitly or implicitly.

Put in the context of growth analysis, the use of cross-country data to predict the behavior of individual countries presupposes certain symmetry judgments about the countries, judgments that are made precise by forms of exchangeability.

Second, exchangeability is separately important because of its implications for the appropriate statistical theory to apply in growth contexts. The reason for this relates to a deep result in probability theory known as de Finetti's Representation Theorem.[10] This theorem, formally stated in the technical appendix, establishes that the sample path of a sequence of exchangeable random variables behaves *as if* the random variables were generated by a mixture of *i.i.d.* processes. For empirical practice, de Finetti's Representation Theorem is important because it creates a link between a researcher's prior beliefs about the nature of the data under analysis (specifically, the properties of regression errors) that permits the researcher to interpret OLS estimates and associated test statistics in the usual way.[11]

10. See Bernardo and Smith (1994, chs. 4 and 6) for an insightful discussion of the nature and implications of the theorem and Aldous (1983) for a comprehensive mathematical development of various forms of the theorem.

11. Caution is needed in using de Finetti's theorem to calculate the distributions of regression estimators. For linear regression models of the form of equation 1, with normally distributed errors and nonstochastic regressors, Arnold (1979, p. 194) shows that "many optimal procedures for the model with *i.i.d.* errors are also optimal procedures for the model with exchangeably distributed errors . . . in the univariate case the best linear unbiased estimator and the ordinary least squares estimator are equal . . . as long as the experimenter is only interested in hypotheses about (the slope coefficients of the regression) he may act as though the errors were *i.i.d.*" Further, if the errors are non-normal, de Finetti's theorem leads one to expect analogous asymptotic equivalences. Similarly, we believe that analogies to de Finetti's theorem can be developed for stochastic regressors and F-conditional exchangeability, although as far as we know no such results have been established.

Exchangeability and Growth

How does exchangeability relate to the assumptions underlying cross-country growth regressions? These models typically assume that once the included growth variables in the model are realized, no basis exists for distinguishing the probabilities of various permutations of residual components in country-level growth rates, that is, these residuals are F-conditionally exchangeable, where F is the modeler's information set. Notice that F may include variables beyond those included in a growth regression as well as knowledge about nonlinearities or parameter heterogeneity in the growth process.

Various forms of exchangeability appear, in our reading of the empirical growth literature, to implicitly underlie many of the regression specifications. An implicit (F-conditional) exchangeability assumption is made whenever the empirical implementation of the growth trajectory for a single country from a given theoretical model is turned into a cross-country regression (typically after linearizing) by allowing the trajectory's state variables to differ across countries and appending an error term. Such an assumption of exchangeability has substantive implications for how a researcher thinks about the relationship between a given observation and others in a data set. Suppose that a researcher is considering the effect of a change in trade openness on a country, for example, Tanzania, in Sub-Saharan Africa. How does the researcher employ estimates of the effects of trade on growth in other countries to make this assessment? The answer depends on the extent to which the causal relationship between trade and growth in Tanzania can be uncovered using data from other countries.

More generally, notice how a number of modeling assumptions that are standard in conventional growth exercises are conceptually related to the assumptions that the model errors ε_i are F-conditionally exchangeable and that the growth rates g_i are partially exchangeable with respect to available information. Specifically,

1. The assumption that a given regression embodies all of a researcher's knowledge of the growth process is related to the assumption that the errors in a growth regression are F-conditionally exchangeable.
2. The assumption that the parameters in a growth regression are constant is related to the assumption that country-level growth rates are partially exchangeable.
3. The justification for the use of ordinary (or heteroskedasticity-corrected) least squares, as is standard in the empirical growth literature, is related to the assumption that the errors in a growth regression are exchangeable (or are exchangeable after a heteroskedasticity correction).

Our general claim is that exchangeability, in particular, F-conditional exchangeability of model errors, is an "incredible" (Sims 1980) assumption in the context of the standard cross-country regressions of the growth literature. (By a standard regression, we refer to equation 1, in which a small number of regressors are assumed to explain cross-country growth patterns.[12]) For exchangeability to hold for a given regression and information set, the likelihood of a positive error for a given country—say, Japan—would need to be the same as that for any other country in the sample. In turn, for this to be true, no prior information could exist about the countries under study that would render the distribution of the associated growth residuals for these countries sensitive to permutations.

To repeat, exchangeability is not necessary to justify the estimation methods and structural interpretations conventionally given to cross-country growth regressions.[13] Hence our use of the term *related* in the three points above. What exchangeability does is provide a baseline, based on economic theory and a researcher's prior knowledge of the growth process, by which to assess cross-country regressions. Exchangeability is a valuable baseline for two reasons. First, the conditions under which various types of exchangeability do or do not hold for growth rates or model residuals can be linked to a researcher's substantive understanding of the growth process in ways that alternative sets of (purely statistical) assumptions on errors usually cannot be. In turn, once exchangeability is believed to be violated, a researcher can naturally link the reasons that exchangeability fails to hold to the question of whether the estimation methods used in the growth literature nevertheless can be expected to yield consistent parameter estimates and standard errors.

Second, exchangeability is important because it shifts the focus of specification analysis away from the question of theory inclusion (determining which variables need to be included in a growth regression to cover relevant structural growth determinants) to the identification of groups of countries that obey a common regression surface and hence can provide information on the growth process. This shift of emphasis is important for two reasons. First, for many growth determinants, the variables used to proxy for theories are very poor measures. For example, in the standard Gastil index of political rights, often used to measure levels of democracy, South Africa is ranked as high as or even higher than (depending on the period) the Republic of Korea for the period 1972–84. It is difficult to know what this means (political rights for whom?) and in what sense this rank ordering is relevant for the aspects of democracy conducive to growth. A more fruitful exercise is to identify groups of countries that obey a common, parsimonious growth model. Put differently, if, as seems plausible, many growth determinants such as political regime are common background variables for subsets of countries, a more productive empirical strategy may be to identify these subsets rather than to use crude empirical proxies for regime. Second, to the extent that nonquantifiable factors, such as "culture" (see Landes

12. To be fair, empirical growth papers often check the robustness of variables relative to a small number of alternative controls, but such robustness checks do not address exchangeability per se.

13. In section VI we return to the question of when the full force of exchangeability is useful for policy analysis.

2000), matter for growth, the identification of partially exchangeable subsets may be necessary for any sort of growth inferences.

To see how exchangeability plays a role in the leap from the identification of statistical patterns to structural inference, suppose that one runs the baseline Solow regression and observes that regression errors for the countries in Sub-Saharan Africa are predominately negative (as is the case). How does one interpret this finding? One can either attribute the finding to chance (the errors are, after all, zero mean with nonnegligible variance) or conclude that there was something about those countries that was not captured by the model. Easterly and Levine (1997), for example, develop a comprehensive argument on the role of ethnic divisions as a causal determinant of growth working from this initial fact. Or, put differently, Easterly and Levine (1997), from prior knowledge about the politics and cultures of these countries, developed their analysis on the basis that the Solow errors were not exchangeable, that is, that there was something about Sub-Saharan African countries that should have been incorporated into the Solow model.

Does the requirement of exchangeability imply the impossibility of structural inference whenever observational data are being studied? This would grossly exaggerate the import of our critique. Exchangeability of errors is conceivable for a wide range of models with observational data sets. For example, exchangeability seems to be a plausible assumption for statistical models based on the use of individual-level data sets, such as the Panel Study of Income Dynamics (PSID), once relevant information about the individuals under study is controlled for. One reason for this relates to the units of analysis. A basic difference between microeconomic data sets of this type and macroeconomic data sets of the type used in growth analysis is that macroeconomic observations pertain to large heterogeneous aggregates for which a great deal of information is known; information that can imply that exchangeability does not hold. In addition, the large size of individual-level data sets such as the PSID means that the range of possible control variables is much greater than that for growth. By this we mean something deeper than "the more data points, the more regressors may be included." Instead, we argue that large data sets of the type found in microeconomics will contain observations on groups of individuals who are sufficiently similar with respect to observables that they may be plausibly regarded as representing exchangeable observations.

That said, we fully accept that exchangeability for observations on objects as complicated as countries may well be problematic. Will our knowledge of the histories and cultures of the countries in cross-country regressions ever be embedded in the regressions to such an extent that the exchangeability requirement is met? This question is at the heart of many of the controversies about the empirical growth literature.

To summarize, conventional growth econometrics has failed to consider the ways in which appropriate exchangeability concepts may or may not hold for the specific models analyzed. This failure in turn renders these studies difficult if not impossible to interpret, because one must know whether any exchangeability violations that are present invalidate the statistical exercise being conducted. We therefore concur with Draper and others (1993, p. 1), who argue that

> statistical methods are concerned with combining information from different observational units and with making inferences from the resulting summaries to prospective measurements on the same or other units. These operations will be useful only when the units to be combined are judged to be *similar* (comparable or homogeneous) . . . judgments of similarity involve concepts more primitive than probability, and these judgments are central to preliminary activities that all statisticians must perform, even though probability specifications are absent or contrived at such a preliminary stage.

Exchangeability and Causality

Though exchangeability is a useful benchmark for understanding some of the major sources of skepticism about growth regressions, it does not bear in any obvious way on the third of our general criticisms, the lack of attention to causality versus correlation in growth analysis. For example, following a nice example due to Goldberger (1991), a regression of parental height on daughter height can have a perfectly well-defined set of exchangeable errors, so that parental heights are partially exchangeable, yet the interpretation of the associated regression coefficient is obviously noncausal. More generally, causality is a different sort of question than the other issues we have addressed, in that it cannot be reduced to a question of whether the data fulfill a generic statistical property. As Heckman (2000, p. 89) notes, "causality is a property of a model . . . many models may explain the same data and . . . assumptions must be made to identify causal or structural models." And (2000, p. 91):

> Some of the disagreement that arises in interpreting a given body of data is intrinsic to the field of economics because of the conditional nature of causal knowledge. The information in any body of data is usually too weak to eliminate competing causal explanations of the same phenomenon. There is no mechanical algorithm for producing a set of "assumption free" facts or causal estimates based on those facts.

In our subsequent discussion we do not address strategies for dealing with questions of causality. Instead, we focus on model uncertainty, which presupposes that causality uncertainty within a given model has been addressed by suitable assumptions by the analyst. In doing this, we are not diminishing the importance of thinking about causal inference; instead, we believe that causal arguments require judgments about economic theory and qualitative information about the problem at hand that represent issues separate from those we address.

IV. A Digression on Noneconometric Evidence

Regression analyses of the type conventionally done are useful mechanisms for summarizing data and uncovering patterns. These techniques are not, as currently employed, particularly credible ways to engage in causal inference. Before proceeding to econometric alternatives, we wish to point out the importance of integrating different sources of information in the assessment of growth theories. These sources are often the basis on which exchangeability can be questioned in a particular context.

The economic history literature is replete with studies that are of enormous importance in adjudicating different growth explanations, yet this literature usually receives only lip service in the growth literature.[14] An exemplar of historical studies that can speak to growth debates is Clark (1987), which explores the sources of productivity differences between cotton textile workers in New England and those workers in other countries in 1910. These differences were immense—a typical New England textile worker was about six times as productive as his counterpart in China or India and more than twice as productive as his counterpart in Germany. Clark painstakingly shows that these differences cannot be attributed to differences in technology, education, or management.[15] Instead, they seem to reflect cultural differences in work and effort norms. Such studies have important implications for understanding why technology may not diffuse internationally and how poverty traps may emerge, and should play a far greater role in the empirical growth literature.

Historical and qualitative studies also play a crucial role in the development of credible statistical analyses. One reason for this is that these sorts of studies provide information on the plausibility of identifying assumptions that are made to establish causality. Further, our discussion on exchangeability and growth analysis may be interpreted as arguing that a researcher needs to do one of two things to claim that a regression provides causal information. The researcher must make a plausible argument that, given the many plausible growth theories and plausible heterogeneity in the way different causal growth factors affect different countries, the errors in a particular growth regression are nevertheless exchangeable. Or, the researcher can make the argument that the violations of exchangeability in the regression occur in ways that do not affect the interpretation of the coefficients and standard errors from those that are employed. To some extent exchangeability judgments must be made prior to a statistical exercise, as Draper and others (1993) note above. Where does information of this type come from? Often from qualitative and historical work. Hence, the detailed study of individual countries that is a hallmark of work by the World Bank, for example, plays an invaluable role in allowing credible statistical analysis.

14. There are notable exceptions, such as Easterly and Levine (1997) and Prescott (1998).

15. See also Wolcott and Clark (1999), which provides detailed evidence that managerial differences cannot explain the low productivity in Indian textiles.

V. Modeling Model Uncertainty

The main themes of our criticisms of current econometric practice may be summarized as two claims:

1. The observations in cross-country growth regressions do not obey various exchangeability assumptions given the information available on the countries under study.

This implies that:

2. Model uncertainty is not appropriately incorporated into empirical growth analyses.

There are no panaceas for the interpretation problems we have described for growth regressions. Although our formulation of model uncertainty can reduce the dependence of empirical growth studies on untenable exchangeability or other assumptions, growth regressions will always rely on untestable and possibly controversial assumptions if causal or structural inferences are to be made. It may be impossible, for example, to place every possible growth theory in a common statistical analysis, so critiques based on theory open-endedness will apply, at some level, to our own suggestions. Further, we will not be able to model all aspects of uncertainty about partial exchangeability of growth rates. However, we do not regard this as a damning defect. Empirical work always relies on judgment as well as formal procedures, what Draper and others (1993, p. 16) refer to as "the role of leaps of faith" in constructing statistical models. What we wish to do is reduce the number and magnitude of such leaps.

General Framework

We assume that the structural growth process for country i obeys a linear structure that applies to all countries j that are members of class $J(i)$. Suppose that this model is described by a set of regressors S that we partition into a subset X and a scalar z. Our analysis focuses on how to employ data to uncover β_z, the coefficient that determines the effect of z_i on country i's growth. We work with models of the form:

$$g_j = S_j\zeta + \varepsilon_j = X_j\pi + z_j\,\beta_z + \varepsilon_j,\ j \in J(i). \tag{7}$$

When a given model represents the "true" or correct specification of the growth process for countries in $\boldsymbol{J}$, the sequence of residuals ε_j will be $\boldsymbol{F}$-exchangeable. The information set $\boldsymbol{F}$ comprises the total available information to a researcher about the countries. For our purposes, $\boldsymbol{F}$ will consist of a collection of regressors available to a researcher; S is a subset of these. The idea that a model consists of the specification of a set of growth determinants, (S_j), and the specification of a set of countries with common parameters, $J(i)$, that together render the associated model errors F-conditionally exchangeable, will, as we shall see, parallel our earlier discussion of the first two sources of criticisms of growth regressions.

It is skepticism about the claim that a particular model is correctly specified in the sense we have described that renders many of the empirical claims in the growth literature not credible.

The standard approach to statistical analysis in the growth literature can be thought of as using a single model M and given data set D to analyze model parameters. Suppose that the goal of the exercise is to uncover information about a particular parameter β_z. From a frequentist perspective, this involves calculating an estimate of the parameter β_z along with an associated standard error for the estimate. From a Bayesian perspective, this involves calculating the posterior density $\mu(\beta_z \mid D,M)$. We will employ the Bayesian framework in our subsequent discussion. That said, we will be interested in relating our analysis to frequentist analyses of growth. For this reason we shall often employ a "leading case" in the analysis. As described in the technical appendix, under some conditions the posterior mean of the set of regression coefficients in equation 7 equals the OLS estimates of the parameters and the posterior variance/covariance matrix equals the variance/covariance matrix of the OLS estimates. We will use this equivalence repeatedly in the next section.

Formulating Types of Model Uncertainty

Suppose that there exists a universe of models, $\mathbf{M}$ with typical element M_m, that are possible candidates for the "true" growth model that generated the data under study; the true model is assumed to lie in this set.[16] This universe is generated from two types of uncertainty. First, there is *theory uncertainty*. In particular, we assume that there is a set $\mathbf{X}$ of possible regressors to include in a growth regression whose elements correspond to alternative causal growth mechanisms. In our framework a theory is defined as a particular choice of regressors for a model of the form of equation 7. Second, there is *heterogeneity uncertainty*. By this we mean that there is uncertainty as to which countries make up $J(i)$, that is are partially exchangeable with country i.[17] In the presence of these types of uncertainty a researcher will be interested not in $\mu(\beta_z \mid D,M_m)$ for a particular M_m but in $\mu(\beta_z \mid D)$; the exception, of course, is when the correct growth theory and the set of countries that are partially exchangeable with country i are known with certainty to the modeler.

This dichotomy of model uncertainty can, at least in principle, incorporate other forms of uncertainty as well. Consider the question of nonlinearities in the growth process. One could attempt to deal with *functional form uncertainty* through the addition of regressors. Examples would include adding regressors that are nonlinear functions of the initial set of theory-based regressors (appealing to Taylor series-type or other approximations) or adding regressors whose

16. It is possible to consider contexts where no model in $\mathbf{M}$ is correct, as discussed in Bernardo and Smith (1994), but that is beyond the scope of this paper.

17. These two types of uncertainty are not independent; for example, theory uncertainty may induce heterogeneity uncertainty.

values are zero below some threshold and equal to a theory-based regressor above that threshold (as suggested by such models as Azariadis and Drazen [1990]. In this sense heterogeneity uncertainty is no different from theory uncertainty.

It is possible to integrate theory uncertainty and some forms of heterogeneity uncertainty into a common variable selection framework. Doing so has the important advantage that it allows us to draw on new developments in the statistics literature stemming from an important paper by Raftery, Madigan, and Hoeting (1997). By definition, theory uncertainty is a question of variable inclusion. To see how to interpret heterogeneity uncertainty in a similar way, we proceed as follows. For a given regressor set S, suppose that one believes that the countries under study may be divided into two subsets with associated subscripts A_1 and A_2 such that the countries within each subset are partially exchangeable, but that countries in one subset may not be partially exchangeable with countries in the other because of parameter heterogeneity. Each of these subsets is characterized by a linear equation so that

$$g_j = X_j\pi + z_j\beta_z + \varepsilon_j \text{ if } j \in A_1 \tag{8}$$

and

$$g_j = X_j\pi' + z_j\beta'_z + \varepsilon_j \text{ if } j \in A_2. \tag{9}$$

This last equation can be rewritten as

$$g_j = X_j\pi + z_j\beta_z + X_j(\pi' - \pi) + z_j(\beta'_z - \beta_z) + \varepsilon_j \text{ if } j \in A_2. \tag{10}$$

Therefore, the two equations can be combined into a single growth regression of the form:

$$g_j = X_j\pi + z_j\beta_z + X_j\,\delta_{j,A_2}(\pi' - \pi) + z_j\,\delta_{j,A_2}(\beta'_z - \beta_z) + \varepsilon_j, \text{ if } j \in A_1 \cup A_2 \tag{11}$$

where $\delta_{j,A_2} = 1$ if $j \in A_2$, 0 otherwise. The additional regressors $X_j\delta_{j,A_2}$ and $z_j\delta_{j,A_2}$ therefore produce a common regression for all observations.[18] Of course, this type of procedure is often done in empirical work; our purpose in this development here is to emphasize how heterogeneity uncertainty may be explicitly modeled in terms of variable inclusion. Notice that it is straightforward to generalize this procedure to multiple groups of partially exchangeable countries. This procedure is not completely general in that it restricts the sort of possible parameter heterogeneity allowed; for example, each country is not allowed a separate set of coefficients. To allow for this more general type of heterogeneity would require moving to an alternative structure, such as a hierarchical linear model (see Schervish 1995, ch. 8); we plan to pursue this in subsequent work.

18. When heterogeneity uncertainty is introduced, the variable z will be associated with different parameters for different countries. For ease of exposition we let β_z refer to the relevant parameter for the country i that is of interest.

Posterior Probabilities

Once a researcher has formulated a space of possible models, it is relatively straightforward to calculate posterior probabilities that do not rely on the assumption that one model is true. In the presence of model uncertainty the calculation of $\mu(\beta_z \mid D)$ requires integrating out the dependence of the probability measure $\mu(\beta_z \mid D,M_m)$ on the model M_m. By Bayes's rule, the posterior density of a given coefficient conditional only on the observed data is

$$\mu(\beta_z \mid D) = \sum_m \mu(\beta_z \mid D,M_m)\mu(M_m \mid D), \tag{12}$$

which can be rewritten as

$$\mu(\beta_z \mid D) \propto \sum_m \mu(\beta_z \mid D,M_m)\mu(D \mid M_m)\mu(M_m), \tag{13}$$

where $\propto$ means "is proportional to," $\mu(D \mid M_m)$ is the likelihood of the data given model M_m, and $\mu(M_m)$ is the prior probability of model M_m. This formulation gives a way of eliminating the conditioning of the posterior density of a given parameter on a particular model choice.

Calculations of this type originally appeared in Leamer (1978) and are reported in Draper (1995). Leamer (1978, p. 118) gives the following derivations of the conditional mean and variance of β_z given the data D:

$$E(\beta_z \mid D) = \sum_m \mu(M_m \mid D)E(\beta_z \mid D,M_m) \tag{14}$$

and

$$\begin{aligned} var(\beta_z \mid D) &= E(\beta_z^2 \mid D) - (E(\beta_z \mid D))^2 = \\ &\sum_m \mu(M_m \mid D)(var(\beta_z \mid M_m,D) + (E(\beta_z \mid D,M_m))^2) - (E(\beta_z \mid D))^2 = \\ &\sum_m \mu(M_m \mid D)var(\beta_z \mid M_m,D) + \sum_m \mu(M_m \mid D)(E(\beta_z \mid D,M_m) - (E(\beta_z \mid D))^2. \end{aligned} \tag{15}$$

As discussed in Leamer (1978) and Draper (1995), the overall variance of the parameter estimate β_z depends on the variance of the within-model estimates (the first term in equation 15) and the variance of the estimates across models (the second term in equation 15).

Equation 12 and the related expressions are all examples of Bayesian model averages. The methodology surrounding Bayesian model averaging is specifically developed for linear models with uncertainty about variable inclusion in Raftery, Madigan, and Hoeting (1997).[19] Doppelhofer, Miller, and Sala-i-Martin (2000), focusing on theory uncertainty only, compute a number of measures of variable robustness based on the application of this formula to growth regressions and conclude that initial income is the "most robust" regressor. Fernandez, Ley, and Steel (1999) also employ Bayesian model averaging for theory uncertainty, focusing on the explicit computation of posterior coefficient distributions. Our own development should be read as an endorsement and extension of the analyses in

19. The survey by Hoeting and others (1999) provides a nice introduction to model averaging techniques. See also Wasserman (2000).

these articles. Our formulation differs in two respects from previous work. First, we treat heterogeneity uncertainty as well as theory uncertainty as part of overall model uncertainty. Draper and others (1993) provide a general overview of the importance of accounting for heterogeneity uncertainty in constructing credible empirical exercises. As our discussion illustrates, heterogeneity uncertainty can be treated as a question of variable inclusion, so the ideas in Doppelhofer, Miller, and Sala-i-Martin (2000) and Fernandez, Ley, and Steel (1999) can be extended to this domain in a straightforward fashion. Second, we develop an explicit decision-theoretic approach to interpreting growth regressions. As far as we are aware, this analysis is new.

Outliers

One important concern in the empirical growth literature has revolved around the role of outliers in determining various empirical claims. A famous example is the role of the Botswana observation in determining the estimated magnitude of social returns to equipment investment (DeLong and Summers 1991, 1994; and Auerbach, Hassett, and Oliner 1994).[20]

Outliers can be dealt with in a straightforward fashion. There are three strategies one can pursue. First, one can always employ a within-model estimator that is designed to be robust to outliers. As Temple (2000) points out, one can employ a trimmed least square estimator (one that drops or downweights observations whose associated OLS residuals are large) in estimating each model's parameters and still employ whatever posterior analysis one wishes. Second, one can explicitly allow the density for model errors to accommodate outliers. For example, one can model errors as drawn from a mixture distribution. Third, and most promising in our view, one can employ a Bayesian bagging procedure due to Clyde and Lee (2000). Bayesian bagging ("bagging" is an abbreviation for bootstrap aggregating) was introduced by Breiman (1996) to improve the performance of what he called "unstable" prediction and modeling methods. A method is "unstable" when small changes in the data set lead to large changes in the method's output. Intuitively, the Clyde and Lee procedure constructs bootstrap data sets from the empirical distribution function of a data set, computes a model average for each sample, and then averages these results. (See their article for details.) Clyde and Lee provide reasons to think that this modification of model averaging will be robust to outliers.

That said, the ex post analysis of outliers, as was carried out in the Botswana case we described, is often problematic; as Leamer (1978, p. 265) remarks, the mechanical and typically ad hoc dropping of outliers both leads to invalid statistical conclusions and ignores valuable information.

20. The role of outliers in growth regressions has been somewhat overstated; for example, the DeLong and Summers results are far more robust to the inclusion or exclusion of Botswana than is often asserted, as a careful reading of DeLong and Summers (1991, 1994) and Auerbach, Hassett, and Offner (1994) clearly reveals. Temple (1998) is a more persuasive example of the importance of dealing with outliers.

Priors on the Space of Possible Models

An important issue in the implementation of the model averaging approach that we describe is the choice of the prior distribution on the space of models. For the problem of variable inclusion, this is typically handled by assuming that all 2^k possible models (where k is the number of regressors that may be placed in a given model) have equal probability; Fernandez, Ley, and Steel (1999) follow this procedure in their analysis. The procedure in essence assumes that the prior probability that a given regressor is in a model is 1/2. Doppelhofer, Miller, and Sala-i-Martin (2000) make the alternative assumption that for a regression whose expected number of included regressors is $\bar{k}$, the probability of inclusion of a given regressor is $\bar{k} / k$. They make this assumption to avoid "a very strong prior belief that the number of included variables should be large" (2000, pp. 15–16).

These alternative approaches to setting model priors are not very appealing from the perspective of economic theory. Clearly, the addition of a given regressor to the set of possible regressors should affect the probabilities with which other variables are included. It is unclear, for example, why the effect of ethnicity on growth should be independent of the effect of democracy, as it can easily be imagined that one will affect growth only if the other does as well. The conventional approaches to modeling the space of priors ignore this fact.

This problem is closely associated with a standard criticism of the "irrelevance of independent alternatives" assumption in choice theory, originally due to Debreu (1960) and later instantiated in the choice literature as the "red bus/blue bus problem" (see Ben-Akiva and Lerman 1985, sec. 3.7). In discrete choice theory irrelevance of independent alternatives means that the ratio of choice probabilities between any two alternatives should be unaffected by the presence of a third. As pointed out by Debreu, this assumption is untenable if the third choice is a close substitute for one of the other two. For the analysis of growth regressions, the priors we have discussed suffer from a similar problem, although the reasons are more complicated. As noted above, the likelihood that one growth theory matters may covary positively with whether another one matters. Further, because the variables employed to capture growth theories are often crude proxies for underlying theories, their inclusion probabilities could covary positively, as each helps measure some common growth determinants. For example, contra Doppelhofer, Miller, and Sala-i-Martin (2000, n. 15), the likelihood that political assassinations predict growth differences could be positively associated with the likelihood that revolutions predict growth differences, as each helps instrument the unobservable variable "political instability."

We have no advice to offer on how to deal with this problem, because its resolution will depend on one's priors on the space of underlying growth models, as determined by the interconnections between particular growth theories. In our view, it makes more sense at this stage of development to treat the prior distribution over models as a benchmark for reporting posterior statistics. (A number of Bayesians have developed a similar view of priors; see, for example, the discussion of "robust" priors in Berger 1987, p. 111.) Because the complexity of the growth process speaks to the strong likelihood that a large number of growth factors substantively matter, the uniform prior of Fernandez, Ley, and Steel (1999) makes the most sense at this stage in providing a benchmark.

That said, there is nothing theoretically compelling about the assumption that the inclusion probability of each regressor is 1/2. We therefore believe that it might make sense in future work to report values for some benchmark alternative probabilities, in order to help evaluate the robustness of results. By choosing inclusion probabilities lower than 1/2, it is possible to incorporate the spirit of the Doppelhofer, Miller, and Sala-i-Martin concerns without having to form prior beliefs on the expected number of regressors in a model, which seems extremely problematic.

VI. Toward a Policy-Relevant Growth Econometrics

The framework developed in the previous section provides a general way of describing model uncertainty in growth regressions. It does not, however, provide any guidance on how to determine what variables should be included in a regression, or on when to regard the sign or magnitude of a regression coefficient as robust. The reason is that the posterior densities embodied in equations 12 to 15 are nothing more than data summaries. As such, they can inform policy analysis only to the extent that they are integrated with a specific formulation of the decision problem of a policymaker. Hence it is necessary to develop an explicit decision-theoretic basis for assessing growth data. The decision-theoretic framework we describe explicitly incorporates the various forms of model uncertainty associated with possible violations of exchangeability, as discussed in the previous section. In this section we discuss the use of growth regressions to inform empirical analysis when one of the growth controls is under the control of a policymaker. Many of the purported policy variables included in growth regressions—for example, indices of political stability—are not necessarily tightly linked to the variables over which a policymaker has control. The framework we describe can be generalized to incorporate a more complicated relationship between growth determinants and policy than the one we analyze here.

The decision-theoretic perspective involves moving away from a specific concern with a particular hypothesis to an evaluation of the implications of a given set of data for a particular course of action. Kadane and Dickey (1980, p. 247) argue

> The important question in practice is not whether a true effect is zero, for it is already known not to do exactly zero, but rather, How large is the effect? But then this question is only relevant in terms of How large is important? This question in turn depends on the use to which the inference will be put, namely, the utility function of the concerned scientist. Approaches which attempt to explain model specification from the viewpoint of the inappro-

> priate question, Is it true that . . . ? have a common thread in that they all proceed without reference to the utility function of the scientist. And therefore, from the decision theory point of view, they all impose normative conditions on the utility function which are seldom explicit and often far from the case.

Substituting *policymaker* for *scientist* in this quotation makes it clear why policy-relevant growth econometrics needs to explicitly integrate policy objectives and empirical practice. Our approach is well summarized by Kass and Raftery (1994, p. 784): "The decision making problem is solved by maximizing the posterior expected utility of each course of action considered. The latter is equal to a weighted average of the posterior expected utilities conditional on each of the models, with the weights equal to the posterior model probabilities." In other words, we argue that policy-relevant econometrics needs explicitly to identify the objectives of the policymaker and then calculate the expected consequences of a policy change.

Policy Assessment: Basic Ideas

The basic posterior coefficient density described by equation 12 and the associated first and second moments described by equations 14 and 15 represent data summaries and as such have no implications for either inference or policy assessment. The goal of a policy analysis is not to construct such summaries but to assess the consequences of changes in a policy. Similarly, such data summaries do not imply the validity of particular rules for data evaluation or inference. For example, the assessment of whether regressors are robust, such as is in extreme bounds analysis or the comparison of models using Bayes factors,[21] may not be appropriate for certain policy exercises. Put differently, decisions on whether to treat regressors as robust and the like should, for the purposes of policy analysis, be derived from the policymaker's assessment of the expected payoffs associated with alternative policies.

In this section we explore policy assessment when model uncertainty has been explicitly accounted for. The purpose of this exercise is twofold. First, it captures what we believe is the appropriate way for policymakers to draw inferences from data. Second, it shows that various rules for the assessment of regressor fragility, such as extreme bounds analysis, will arise in such exercises. A critical feature of this approach to model assessment is it illustrates that the evaluation of regressor robustness can be derived from particular aspects of the policymaker's objective function.

For expositional purposes we initially suppose that the goal of an empirical exercise is to evaluate the effect of a change dz_i[22] in some scalar variable that is

21. For any two models M_m and $M_{m'}$, the Bayes factor $B_{m,m'}$ is defined as $\mu(D \mid M_m) / \mu(D \mid M_{m'})$. Kass and Raftery (1994) provide an extensive overview of the use of Bayes factors in model evaluation.

22. Without loss of generality, we generally assume that $dz_i > 0$.

under the control of a policymaker and believed to have some effect on growth. Therefore, the decisionmaker's set of actions A is $\{0, dz_i\}$. This decision rule is based on a vector observable data D. This means that a decisionmaker chooses a rule $\phi(\cdot)$ that maps D to A so that

$$\begin{aligned} \phi(D) &= dz_i \text{ if } D \in D_1 \\ \phi(D) &= 0 \text{ otherwise.} \end{aligned} \tag{16}$$

D_1 is therefore the acceptance region for the policy change. We assume that the "true" linear growth model is a causal relationship that will allow evaluation of the effect of this change.

Because we restrict ourselves to linear models, the analysis of the policy decision is particularly straightforward, as $\mu(\beta_z \mid D)$ will describe the posterior distribution of the effect of a marginal change in z on growth in a given country. A marginal policy intervention in country i can be evaluated as follows. Let z_i denote the level of a policy instrument in country i. This instrument appears as one of the regressors in the linear model that describes cross-country growth. Suppose one has the option of either keeping the policy instrument at its current value or changing it by a fixed amount dz_i. Let g_i denote the growth rate in the country in the absence of the policy change, and $g_i + \beta_z dz_i$ the growth rate with the change. Finally, let $V(g_i, O_i)$ denote the utility value of the growth rate to the policymaker. O_i is a placeholder vector that contains any factors relating to country i that affect the policymaker's utility.

An expected utility assessment of the policy change can be based on the comparison

$$E(V(g_i + \beta_z dz_i, O_i) \mid D) - E(V(g_i, O_i) \mid D). \tag{17}$$

Calculations of the expected utility differential in equation 17 implicitly contain all information relevant to a policy assessment. From the perspective of policy evaluation, the various rules that have been proposed for the assessment of regressor robustness should be an implication of this calculation. Notice that this calculation requires explicitly accounting for model uncertainty, because the conditioning is always done solely with respect to the data.

Policy Assessment under Alternative Utility Functions

In this section we consider the implications of some alternative utility functions for the analysis of growth regressions. Our goal is to show how particular utility functions will lead a policymaker to decide whether or not to implement a policy on the basis of aspects of the posterior distribution of β_z. We do not claim that the utility functions we examine are particularly compelling. We have chosen them to illustrate what sort of utility functions can justify some of the standard ways of interpreting growth regressions.

RISK NEUTRALITY. Suppose that V is linear and increasing in the level of growth, that is,

$$V(g_i, O_i) = \alpha_0 + \alpha_1 g_i, \alpha_1 > 0. \tag{18}$$

For this policymaker the relevant statistic is the posterior mean of the regressor coefficient. In this case it is straightforward to see that the policy change is justified if the expected value of the change in the growth rate is positive, that is,

$$\sum_m \mu(M_m \mid D) E(\beta_z \mid D, M_m) > 0. \tag{19}$$

When the prior model probabilities are equal, this is equivalent to the condition

$$\sum_m \mu(D \mid M_m) E(\beta_z \mid D, M_m) > 0 \tag{20}$$

so the likelihoods $\mu(D \mid M_m)$ determine the relative model weights.

MEAN/VARIANCE UTILITY OVER CHANGES IN THE GROWTH RATE. Suppose that a policymaker has preferences that relate solely to changes in the growth rate, as opposed to its level. The idea here is that a policymaker assesses a policy relative to the baseline g_i. Operationally, we assume that one chooses the elements of O_i and the functional form of $V(\cdot\,,\cdot)$ so that

$$\begin{gathered} E(V(g_i + \beta_z dz_i, O_i) \mid D) - E(V(g_i, O_i) \mid D) = \\ \alpha_0 E(\beta_z dz_i \mid D) + \alpha_1 var(\beta_z dz_i \mid D)^{1/2}, \alpha_0 > 0, \alpha_1 < 0. \end{gathered} \tag{21}$$

When $|\alpha_0/\alpha_1| = 1/2$, this utility specification implies that the policymaker will act only if the t-statistic (the posterior expected value of β_z divided by its posterior standard deviation) is greater than 2. Hence this specification, at least qualitatively, corresponds to the standard econometric practice of ignoring regressors whose associated t-statistics are less than 2.

From a decision-theoretic perspective, the conventional practice of ignoring "statistically insignificant" coefficients (by which we mean coefficients whose posterior standard errors are more than twice their posterior expected values) can be justified only in very special cases. First, it is necessary to assume that the form of risk aversion of the policymaker applies to the standard deviation rather than to the variance of the change in growth. Otherwise, the desirability of the policy will depend on the magnitude of dz_i. For example, if the utility function is

$$\begin{gathered} E(V(g_i + \beta_z dz_i, O_i) \mid D) - E(V(g_i, O_i) \mid D) = \\ \alpha_0 E(\beta_z dz_i \mid D) + \alpha_1 var(\beta_z dz_i \mid D), \alpha_0 > 0, \alpha_1 < 0 \end{gathered} \tag{22}$$

with $|\alpha_0/\alpha_1| = 1/2$, there will be a threshold level T such that for all $0 < dz_i \leq T$ a policy change increases the policymaker's utility.[23] Therefore, the rule of ignoring regressors with t-statistics less than 2 presupposes a very specific assumption about how risk affects the policymaker's utility. Second, if equation 22 is the correct utility function, the policymaker may still choose to act with the fixed dz_i level we started with under (conventionally defined) statistical insignificance

23. This is an example of the famous result of Pratt that one will always accept a small amount of a fair bet. We plan to address the question of the optimal choice of dz_i in future work.

or, alternatively, may decline to act when the coefficient is statistically significant. These possibilities can be generated through appropriate choices of $|\alpha_0/\alpha_1|$.

KNIGHTIAN UNCERTAINTY AND MAXIMIN PREFERENCES. In the examples we have studied thus far we have allowed all uncertainty about the correct model M_m to be reflected in the posterior model probabilities $\mu(M_m \mid D)$. An alternative approach to model uncertainty, one in the tradition of Knightian uncertainty, assumes that an additional layer of uncertainty exists in the environment under study that may be interpreted as a distinct type of risk, sometimes called ambiguity aversion, as will be seen below.

As before, let $\boldsymbol{M}$ denote the universe of possible growth models. A risk sensitive utility function for the policymaker can be defined as

$$(23) \qquad (1-e)E(V(g_i,O_i) \mid D) + e(inf_{M_m \in \boldsymbol{M}}E(V(g_i,O_i) \mid D,M_m)).$$

In this equation e denotes the degree of ambiguity.

This equation is motivated by recent efforts to reconceptualize utility theory in light of results such as the classic Ellsberg paradoxes. For example, if experimental subjects are given a choice between (1) receiving \$1 if they draw a red ball at random from an urn that they know contains 50 red balls and 50 black balls and (2) receiving \$1 if they draw a red ball when the only information available is that the urn contains 100 red and black balls, the subjects typically choose the first, "unambiguous" urn (Camerer 1995, p. 646). Clearly, if subjects were Bayesians who placed a flat prior on the distribution of the balls in case 2, they would be indifferent between the two options.[24]

Experimental evidence of ambiguity aversion has led researchers—including Anderson, Hansen, and Sargent (1999); Epstein and Wang (1994); and Gilboa and Schmeidler (1989)—to consider formal representations of preferences that exhibit ambiguity aversion. One popular representation, studied in Epstein and Wang (1994), replaces expected utility calculations of the form $\int u(\omega)dP(\omega)$ with $inf_{p \in \boldsymbol{P}} \int u(\omega)dP(\omega)$, where $\boldsymbol{P}$ is a space of possible probability measures. When this space contains a single element, this second expression reduces to the first, which is the standard expected utility formulation. A variant of this formulation is to assume that $\boldsymbol{P}$ consists of a set of mixture distributions $(1-e)P_0 + eP_1$, where P_0 is a baseline probability measure that a policymaker believes to be true, P_1 is the least favorable of all possible probability measures for the policymaker, and e represents the strength of the possibility that this measure applies. When the universe of alternative processes for growth is the space of linear models that we have described, one can replace P with M_m and $\boldsymbol{P}$ with $\boldsymbol{M}$ and obtain the second part of equation 23. In using this specification, we do not claim that it is the only sensible way to model ambiguity aversion by policymakers. We introduce

24. See Camerer (1995) for additional examples of ambiguity aversion as well as a survey of the implications of different results in the experimental economics literature for utility theory.

it to illustrate how recent developments in decision theory may be linked to econometric practice.

We can explore the effects of this additional uncertainty on our analysis by considering the two specifications of V studied above. First, assume that V is linear and increasing, while equation 23 characterizes the ambiguity aversion we have described. In this case the policy change dz is justified if

$$(24) \qquad (1-e)E(\beta_z \mid D) + e(inf_{M_m \in M} E(\beta_z \mid D, M_m)) > 0.$$

When $e = 1$, the policy action will be taken only when $E(\beta_z \mid D, M_m) > 0$ for all $M_m \in \mathbf{M}$. This has an interesting link to OLS coefficients for different models in M. In the leading case described in the technical appendix, the posterior expectation $E(\beta_z \mid D, M_m)$ equals $\hat{\beta}_{z,m}$, the OLS coefficient associated with the regressor z for model M_m. If $e = 1$, this utility function would then mean that a policymaker will choose to implement dz_i if the OLS coefficient estimate of β_z is positive for every model in $\mathbf{M}$.

Alternatively, assume that the policymaker is risk-averse in the sense that equation 21 describes his utility function. In this case the policy change should be implemented if

$$(25) \qquad (1-e)(\alpha_0 E(\beta_z \mid D) + \alpha_1 var(\beta_z \mid D)^{1/2}) + e(inf_{M_m \in M} (\alpha_0 E(\beta_z \mid D, M_m) + \alpha_1 var(\beta_z \mid D, M_m)^{1/2})) > 0.$$

Again, this rule has an interesting link to OLS parameter estimates. If $e = 1$ and $\mid \alpha_0/\alpha_1 \mid = 1/2$, then for the leading case in the technical appendix, the policymaker will not act unless the OLS regression coefficient $\hat{\beta}_{z,m}$ is positive and statistically significant (in the sense that the t-statistic is at least 2) for each model in $\mathbf{M}$. (Here, we rely on the additional fact that for the leading case, $var(\beta_z \mid D, M_m)$ equals the OLS variance of $\hat{\beta}_{z,m}$.)

The policy rules that hold for $e = 1$ are closely related to the recommendations made by Leamer for assessing coefficient fragility through extreme bounds analysis (see Leamer 1983 and Leamer and Leonard 1983). In extreme bounds analysis, recall that when a regressor "flips signs" across specifications, this is argued to imply that the regressor is fragile. From the perspective of policy recommendations, we interpret this notion of fragility to mean that no policy change should be made when there is a model of the world under which the policy change can be expected to make things worse off. This suggests that extreme bounds analysis is based on a maximin assumption of some type. Our derivations show that this intuition can be formalized.

This derivation of extreme value analysis appears to complement a number of the objections raised against it by Granger and Uhlig (1990) and McAleer, Pagan, and Volker (1985). Both these articles argue that extreme bounds analysis can lead to spurious rejections of regressors as a result of changes in sign induced by regressions that are, by standard tests, misspecified. In our view these criticisms need to be developed from the perspective of the objectives of the empirical exercise. Put differently, the salience of these critiques of extreme bounds analysis requires that one reject the utility functions we have described as supporting extreme bounds analysis.

Further, we believe that our derivations provide an appropriate way of modifying extreme bounds analysis—through the use of utility functions, such as equation 23 for $0 < e < 1$. For such cases the relative goodness of fit of different models will be relevant to the empirical exercise. As is well known (see Wasserman 2000, p. 94 for a nice exposition), when Bayesian model selection between two models is based on posterior odds ratios and the prior odds on the models are equal, the posterior odds ratios will equal the ratio of their likelihoods, that is, the posterior odds will reflect the relative likelihoods of the data under the alternative models. Further, as the amount of data becomes large enough, for this special case of equal prior odds, the model with the minimum Kullback-Leibler Information Criterion (KLIC) distance to the "truth" will be revealed. If the set of models under scrutiny includes the true model, the true model will be revealed in large samples.[25] Thus in our context, under our assumption of equal prior odds across models, we may expect the data in large samples to ultimately place greater weight on models whose KLIC distance is closer to the true model.[26]

By choosing $0 < e < 1$ for policymaker utility functions such as equations 24 and 25, one can retain the ambiguity aversion that justifies (in a limiting case) extreme bounds analysis. In particular, one can reflect a policymaker's desire to avoid harm when he faces scientific ambiguity, but at the same time prevent him from being so ambiguity-averse that he fails to take welfare-enhancing actions that are supported by relatively good posterior odds under available scientific evidence (especially when samples are large enough to contain *some* policy-relevant predictive information). Notice that our treatment avoids the criticism of Bayes factors in Kadane and Dickey (1980) that the weights do not account for the purpose of the empirical exercise.

ALTERNATIVE UTILITY FUNCTIONS. In the previous section we assumed that the policymaker cares only about the level of or change in growth induced by a policy change. It is of course possible to imagine other plausible utility functions for a policymaker. One possibility is to assume that a policymaker evaluates a utility on the basis of changes in the expected value of growth within a regime; formally, one assumes that there exists a function ψ such that

$$\psi(E(g_i + \beta_z dz_i \mid D, M_m) - E(g_i \mid D, M_m)) \tag{26}$$

25. See White (1994) for a discussion of measures of closeness based on KLIC and how various estimators achieve minimum KLIC distance to the true mode in large samples.

26. There are some subtleties involved in making the argument we have sketched precise. For example, regularity conditions need to be assumed to justify assertions about the relationship between KLIC distance minimization and quasi-maximum likelihood estimation. Furthermore, as Fernandez, Ley, and Steel (2001) point out, the form of priors for parameters *within* models raises thorny issues. Nevertheless, we believe that this heuristic argument is useful.

measures the utility for a policy change conditional on a particular growth model. Again assuming the leading case where $E(\beta_z \mid D,M_m)$ equals the OLS coefficient $\hat{\beta}_{z,m}$, linearity of the expected growth process implies that the expected utility from dz_i, once one accounts for model uncertainty, is

$$(27) \qquad E(\psi(E(\beta_z dz_i \mid D,M_m)) \mid D) = \sum_m \mu(M_m \mid D)\psi(E(\beta_z dz_i \mid D,M_m)) = \sum_m \mu(M_m \mid D)\psi(\hat{\beta}_{z,m} dz_i)$$

When $\psi(\cdot)$ is linear, this reduces to the risk-neutral case discussed earlier. However, alternative utility functions can produce very different decision rules. For example, suppose that either $\psi(c) = -\infty$ if $c < 0$, $\psi(\cdot)$ bounded otherwise, or $\psi(c) = -\infty$ if $c > 0$, $\psi(\cdot)$ bounded otherwise. One will then have the implied decision rule that a single sign change in the OLS coefficient estimate $\hat{\beta}_{z,m}$ as one moves across models is sufficient to imply that the policymaker should not act to either increase or decrease z_i by dz. This type of utility function induces behavior mimicking that found under Knightian uncertainty.

At first glance this might appear to be an unreasonable utility function for a policymaker. This conclusion is at least partially incorrect. Suppose that each state of the world is indexed by the growth process that is "true" under it. The utility of the policymaker will then depend on both the growth rate that is expected to prevail and the state of the world under which it transpires. For example, suppose that there is a model of the world in which the expected effect of democracy on growth is negative. Such a model could be one whose features imply that a policymaker is particularly wary of reducing growth by changing a given policy instrument. For example, if there is a (positive probability) model of the world in which democracy is especially fragile and may not survive a growth reduction, a policymaker might be especially wary of the policy change for fear this would prove to be the correct model of the world.

This type of argument can be formalized by considering model-dependent utility specifications. Suppose that conditional on model M_m, the utility from a policy change is equal to

$$(28) \qquad U(E(\beta_z \mid D,M_m)dz_i,M_m) = U(\hat{\beta}_{z,m}dz_i,M_m)$$

so that the posterior expected utility of the policy change is

$$(29) \qquad E(U(\hat{\beta}_{z,m}dz_i,M_m) \mid D) = \sum_m \mu(M_m \mid D)U(\hat{\beta}_{z,m}dz_i,M_m).$$

Manipulating $U(\cdot,\cdot)$, one can produce (under the leading case) a result that is consistent with refusing to act whenever the posterior mean $\hat{\beta}_{z,m}$ is negative for at least one M_m, thereby producing extreme bounds–like behavior in the sense that one would not choose $dz_i > 0$, even though for all other models $\hat{\beta}_{z,m}$ is positive.

Policy Analysis and Exchangeability

A decision-theoretic approach of the type we have advocated makes clear the importance of a growth model being rich enough for a researcher to plausibly regard the observations as F-conditionally exchangeable. Suppose that a researcher is using data from I countries to provide a recommendation for the optimal choice of z_i subject to some constraint set $\mathbf{Z}_i$ for country i. In other words, a researcher is attempting to solve the problem

$$max_{z_i \in Z_i} E(V(g_i, z_i)) \tag{30}$$

where information in computing this expression is taken from the regression described by equation 1.

What information in equation 1 is relevant to this calculation? The answer depends on the shape of V. Suppose that V is linear in growth rates, as in equation 18 above. The only information needed about the growth process as described by equation 1 is the posterior expected value of β_z. In our leading case, the OLS coefficient in a growth regression will be sufficient for policy analysis as long as all countries are described by a common linear model. Growth rates need not be partially exchangeable, because partial exchangeability requires symmetry with respect to all moments of the growth process. Similarly, suppose that V is quadratic. In this case one will need only the second moments from the posterior densities generated by equation 1 to apply to country i; partial exchangeability is still not necessary.

However, if V is arbitrary, one will need to employ equation 1 to obtain information on the full conditional distribution $F(\varepsilon_i \mid X_i, Z_i)$. To reveal this type of statistic from cross-country data, one will require full F-conditional exchangeability of the type we have discussed.

VII. An Empirical Example

In this section we reconsider an important growth study, Easterly and Levine (1997), which examines the role of ethnic conflict in growth.[27] We chose to reexamine this study for three reasons. The study is widely regarded as quite important in the growth literature. It has important implications for policy and the sorts of advice and advocacy an international organization would engage in. And the authors of the study have done an admirable job of making their data and programs publicly available.

Easterly and Levine's analysis is designed to explain why in standard cross-country growth regressions the performance of Sub-Saharan Africa[28] is so much worse than that of the rest of the world. Rather than remain content with modeling this phenomenon as a fixed effect (a dummy variable) for these countries, Easterly and Levine argue that a major cause of the poor growth performance is the presence of ethnic conflict in these countries. They construct a measure of ethnic diversity to proxy for this conflict. This variable is substantially larger for

27. We thank Duncan Thomas for suggesting to us that the findings in Easterly and Levine (1997) warrant reexamination.

28. See the data appendix for the list of the countries in Sub-Saharan Africa.

Sub-Saharan Africa than for the rest of the world. Inclusion of the variable in a cross-country growth regression reduces the size of the African fixed effect and is itself statistically significant. Easterly and Levine (1997, p. 1241) conclude that "the results lend support to the theories that interest group polarization leads to rent-seeking behavior and reduces the consensus for public goods, creating long-run growth tragedies."

Our reexamination of this study has an explicitly narrow focus. In our view it is important to see whether and how the influence of ethnolinguistic heterogeneity on growth depends on what other variables are included in the regression. Further, a natural alternative to the claim that the African growth experience is different because of an omitted variable, ethnolinguistic heterogeneity, is that other growth determinants influence Africa differently than they do the rest of the world. Put differently, parameter heterogeneity is a natural alternative explanation.

We therefore conduct the following analysis to account for the effect of model uncertainty on Easterly and Levine's results. A data appendix describes the variables we employ; these are identical to those used in Easterly and Levine (1997). The data are based on decade-long average observations for the 1960s, 1970s, and 1980s, except where indicated in the appendix. We focus on a reexamination of Easterly and Levine's equation 3, table IV, which by conventional measures (such as the statistical significance of all included variables) is arguably their strongest regression in support of the role of ethnic diversity in growth. Our results using this regression are reported in column 1 of our table 1. The key variable of interest is *ELF60*, a measure of ethnic diversity in each country in 1960.

We explore the role of model uncertainty in two ways. We first consider the impact of theory uncertainty on inferences about the determinants of growth. We do this by constructing a universe of models that consists of all possible combinations of the variables in Easterly and Levine's baseline regression. This exercise should be interpreted as a robustness check for Easterly and Levine's results. To perform this exercise, we employ an approximation algorithm whereby posterior model probabilities are replaced with their maximum likelihood estimates. We perform the subsequent calculation of the posterior mean and standard deviation of each regression coefficient using formulas 14 and 15.[29]

Our results incorporating theory uncertainty are reported in column 2 of table 1. Interestingly, we find that the evidence of a role for ethnic diversity in the growth process is slightly strengthened through the model averaging technique. Specifically, the posterior mean of *ELF60* is –0.02 under model averaging, compared with the –0.017 estimate reported by Easterly and Levine. Our primary conclusion from this exercise is that Easterly and Levine's main result is robust to theory uncertainty as we have characterized it.

29. See the computational appendix for details on the calculation of these quantities. Ethnolinguistic heterogeneity is not, of course, directly subject to a policymaker's control, so we do not explore the issues raised in section VI. The policy importance of the variable stems from the implications of its importance to questions of institutional design.

TABLE 1. *OLS* and Bayesian Model Averaging Coefficient Estimates and Standard Errors Using Data from Easterly and Levine (1997)

	[1]	[2]	[3]	[4]	[5]	[6]
Intercept term	—	—	—	—	0.4013	0.1382
	—	—	—	—	(0.3985)	(0.0336)
Dummy for Sub-	−0.0113	−0.0031	0.9558	0.0761	—	—
Saharan Africa	(0.0048)	(0.0053)	(0.3704)	(0.0302)	—	—
Dummy for Latin	−0.0191	−0.0197	−0.0197	−0.0184	—	—
America and the Caribbean	(0.0036)	(0.0042)	(0.0035)	(0.0037)	—	—
Dummy for 1960s	−0.2657	−0.2200	−0.3643	−0.0028	—	—
	(0.0998)	(0.1765)	(0.1328)	(0.0326)	—	—
Dummy for 1970s	−0.2609	−0.2154	−0.3520	0.0009	0.0080	0.0050
	(0.0997)	(0.1745)	(0.1332)	(0.0325)	(0.0134)	(0.0079)
Dummy for 1980s	−0.2761	−0.2298	−0.3650	−0.0143	−0.0038	−0.0024
	(0.0996)	(0.1751)	(0.1336)	(0.0325)	(0.0132)	(0.0058)
Log of initial	0.0870	0.0756	−0.1090	0.0218	−0.0696	−0.0004
income	(0.0254)	(0.0444)	(0.0986)	(0.0088)	(0.1171)	(0.0027)
Log of initial	−0.0063	−0.0056	0.0070	−0.0022	0.0044	−0.0000
income squared	(0.0016)	(0.0029)	(0.0067)	(0.0006)	(0.0088)	(0.0002)
Log of schooling	0.0117	0.0130	−0.0220	0.0130	−0.0131	−0.0017
	(0.0042)	(0.0056)	(0.0216)	(0.0045)	(0.0194)	(0.0077)
Assassinations	−12.8169	−3.3629	−377.3810	−30.6120	−306.4870	−343.4434
	(9.2709)	(7.8137)	(165.5661)	(86.9027)	(158.4484)	(181.6948)
Financial depth	0.0162	0.0111	0.1010	0.0129	0.0774	0.0104
	(0.0058)	(0.0083)	(0.0497)	(0.0075)	(0.0483)	(0.0278)
Black market	−0.0188	−0.0219	−0.0130	−0.0207	−0.0171	−0.0039
premium	(0.0045)	(0.0053)	(0.0098)	(0.0043)	(0.0107)	(0.0081)
Fiscal surplus/GDP	0.1210	0.1717	0.1200	0.1382	0.1654	0.0948
	(0.0314)	(0.0411)	(0.0874)	(0.0357)	(0.0986)	(0.1071)
Ethnic diversity	−0.0169	−0.0222	−0.2020	−0.1437	−0.1516	−0.1595
(*ELF60*)	(0.0060)	(0.0066)	(0.0376)	(0.0279)	(0.0353)	(0.0327)

[1] Ordinary least squares estimates for model "*ALL*".
[2] Bayesian model averaging estimates for model "*ALL*".
[3] Ordinary least squares estimates for model "*ALL* + *ALL**I(*AFRICA*)"; composite coefficient estimates and standard errors reported. *AFRICA*, *LATINCA*, and *DUM60* dropped from *AFRICA*-specific set of regressors.
[4] Bayesian model averaging estimates for model "*ALL* + *ALL***I*(*AFRICA*)"; composite coefficient estimates and standard errors reported. *AFRICA*, *LATINCA*, and *DUM60* dropped from *AFRICA*-specific set of regressors.
[5] Ordinary least squares on *AFRICA* subsample.
[6] Bayesian model averaging on *AFRICA* subsample.
Note: Standard errors are in parentheses.

As we have emphasized, theory uncertainty is not the only form of model uncertainty that needs to be accounted for in cross-country analysis. We therefore next incorporate heterogeneity uncertainty. Following equation 11, we do this by constructing for each regressor x_i in the baseline regressors a corresponding variable $x_i\delta_{j,A_A}$, where $\delta_{j,A_A} = 1$ if country j is in Sub-Saharan Africa and 0 otherwise. This allows for the possibility that the Sub-Saharan African countries have different growth parameters than the rest of the world. Column 3 in table 1 reports the OLS values and standard errors of the regressor coefficients for the African countries; column 4 reports the same statistics when model averaging is done over the augmented variable set. Column 5 reports OLS estimates of the growth regression coefficients and standard errors when the African subsample is analyzed in isolation; column 6 reports the corresponding model average results.

Our explorations of the role of heterogeneity uncertainty provide a rather different picture of the role of ethnicity in African growth than of its role in the rest of the world. The coefficient estimates for Africa are about 7–10 times greater than the corresponding estimates for the world.[30] This result is extremely striking and makes clear that the operation of ethnic heterogeneity on growth is different in Africa, not just the levels of ethnic heterogeneity. Further, a comparison of the other regressor coefficients for Africa with those of the rest of the world makes clear that the growth observations for African countries should not be treated as partially exchangeable with the growth rates of the rest of the world.

These results in no way diminish the importance of Easterly and Levine's findings. In fact, our exercises show that their basic claims are robust to a limited variable uncertainty exercise. Our finding of parameter heterogeneity with respect to ethnolinguistic heterogeneity suggests a direction along which to extend their research. Our results illustrate how additional insights can be obtained by explicitly controlling for model uncertainty.

Finally, we again note that this reexamination is quite narrow. A full-scale study should at a minimum include explicit calculations and presentation of the predictive distribution of the effects of the policy change on growth. Fernandez, Ley, and Steel (1999) provide a good illustration of how to present results of this type. More generally, the reporting of results should always include the information necessary to calculate the posterior expected utility changes of the policymaker. Our own reporting is useful for mean/variance utility functions, but not for the others we have discussed. In addition, we have not allowed for parameter heterogeneity for countries outside Sub-Saharan Africa; doing so is a natural extension of this exercise. The results we report should be treated as suggestive, in this sense, of more elaborate examinations of the role of ethnic heterogeneity in the growth process.

VIII. Conclusions

This paper has had two basic aims. First, we attempted to delineate the major criticisms of cross-country growth regressions and to show how to interpret two

30. Similar results are obtained when one compares Sub-Saharan Africa with the rest of the world. When the Sub-Saharan African countries are dropped from the data set, the OLS estimate for the *ELF60* regressor is –0.0115 with an associated standard error of 0.006. The associated values when model averaging is done across different regressor combinations (to check for robustness to theory uncertainty) are –0.013 and 0.009. By conventional levels, one would conclude that ethnicity is marginally statistically significant outside Sub-Saharan Africa.

of these criticisms, theory uncertainty and parameter uncertainty, as violations of a particular assumption—F-conditional exchangeability—in the residual components of growth models. Second, we outlined a framework for conducting and interpreting growth regressions. For conducting regressions, we advocated an explicit modeling of theory and heterogeneity uncertainty and the use of model averaging to condition out strong assumptions. For interpreting regressions, we argued that the policy objectives associated with a given exercise must be made explicit in the analysis. We outlined a decision-theoretic approach to growth regressions and explored its relationship to conventional approaches to assessing model robustness. Finally, in an empirical application we showed how attention to model uncertainty can provide new insights into the relationship between ethnicity and growth.

To amplify some earlier remarks, we do not believe that there is a single privileged way to conduct statistical or, for that matter, empirical analysis in the social sciences. Persuasive empirical work always requires judgments and assumptions that cannot be falsified or confirmed within the statistical procedure being employed.[31] Indeed, this is the reason that we have not included a treatment of how to provide more robust arguments in favor of causality in this article. What we hope is that this article has provided some initial steps toward the development of a language in which policy-relevant empirical growth research may be better expressed.

Computational Appendix

All model averaging calculations were done using the program bicreg, which was written in SPLUS by Adrian Raftery and is available at *www.research.att.com/~volinsky/bma.html*. Given the large number of possible models, this program, as is standard in the model averaging literature, uses a search algorithm that explores only a subset of the model space; the key feature of the design of the algorithm is that it ensures that the search proceeds along directions such that it is likely to cover models that are relatively strongly supported by the data. We follow the procedure suggested by Madigan and Raftery (1994); see Raftery, Madigan, and Hoeting (1997); and Hoeting and others (1999) for additional discussion. Though the reader should see those papers for a full description of the search algorithm, Hoeting and others (1999, p. 385) provide a nice intuitive description:

> First, when the algorithm compares two nested models and decisively rejects the simpler model, then all submodels of the simpler model are rejected. The second idea, "Occam's window," concerns the interpretation of the ratio of posterior model probabilities $\mathrm{pr}(M_0/D)/\mathrm{pr}(M_1/D)$. Here M_0 is "smaller" than M_1. . . . If there is evidence for M_0 then M_1 is rejected, but rejecting M_0 requires strong evidence *for* the larger model M_1.

31. See Draper and others (1993) and Mallows (1998) for valuable discussions of such issues.

In implementing the model averaging procedure, the algorithm we employ uses an approximation, due to Raftery (1995), based on the idea that, because for a large enough number of observations, the posterior coefficient distribution will be close to the maximum likelihood estimator, and so one can use the maximum likelihood estimates to avoid the need to specify a particular prior. We refer the reader to Raftery (1994) as well as to Tierney and Kadane (1986) for technical details. While some evidence exists that this approximation works well in practice, more research is needed on the specification of priors for model averaging; an important recent contribution is Fernandez, Ley, and Steel (2001).

Data Appendix: Variable Definitions

All data are the same as those used in Easterly and Levine (1997).

- *AFRICA:* Dummy variable for Sub-Saharan African countries, as defined by the World Bank. These countries are Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, the Central African Republic, Chad, Comoros, Democratic Republic of Congo, Republic of Congo, Côte d'Ivoire, Djibouti, Equatorial Guinea, Ethiopia, Gabon, The Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mozambique, Namibia, Niger, Nigeria, Rwanda, São Tomé and Principe, Senegal, Seychelles, Sierre Leone, Somalia, South Africa, Sudan, Swaziland, Tanzania, Togo, Uganda, Zambia, and Zimbabwe.
- *ASSASS:* Number of assassinations per 1,000 population.
- *BLCK:* Black market premium, defined as log of 1 + decade average of black market premium.
- *DUM60:* Dummy variable for 1960s.
- *DUM70:* Dummy variable for 1970s.
- *DUM80:* Dummy variable for 1980s.
- *ELF60:* A measure of ethnic diversity, equalling an index of ethnolinguistic fractionalization in 1960. This variable measures the probability that two randomly selected individuals from a given country will not belong to the same ethnolinguistic group.
- *GYP:* Growth rate of real per capita GDP.
- *LATINCA:* Dummy variable for countries in Latin America and the Caribbean.
- *LLY:* Financial depth, measured as the ratio of liquid liabilities of the financial system to GDP, decade average. Liquid liabilities consist of currency held outside the banking system plus demand and interest-bearing liabilities of banks and nonbank financial intermediaries.
- *LRGDP:* Log of real per capita GDP measured at the start of each decade.
- *LRGDPSQ:* Square of LRGDP.

- *LSCHOOL:* Log of 1 + average years of school attainment, quinquennial values (1960–65, 1970–75, 1980–85).
- *SURP:* Fiscal surplus/GDP: Decade average of ratio of central government surplus to GDP, both in local currency, local prices.

Technical Appendix

1. *De Finetti's Representation Theorem*

De Finetti's theorem establishes that the symmetry inherent in the concept of the exchangeability of errors leads to a representation of the joint distribution of the errors in terms of an integral of the joint product of identical marginal distributions against some conditional distribution function. The theorem is as follows.

If η_i is an infinite exchangeable sequence with associated probability measure P, there exists a probability measure Q over F, the space of all distribution functions on R, such that the joint distribution function $F(\eta_{i-j} \ldots \eta_i \ldots \eta_{i+k})$ for any finite collection $\eta_{i-j} \ldots \eta_i \ldots \eta_{i+k}$ may be written as

$$F(\eta_{i-j} \ldots \eta_i \ldots \eta_{i+k}) = \int \prod_{r=-j}^{k} F(\eta_{i+r}) dQ(F). \tag{A-1}$$

See Bernardo and Smith (1994, p. 177), for this formulation of de Finetti's theorem as well as a proof.

2. *Some Relations between OLS Estimates and Bayesian Posteriors*

For the linear model

$$g_i = S_i\zeta + \varepsilon_i \; i = 1 \ldots I \tag{A-2}$$

suppose that (1) conditional on $S_1 \ldots S_I$, the ε_is are independent and identically distributed and jointly normal; the marginal distribution of the typical element is $N(0,\sigma_\varepsilon^2)$, (2) σ_ε^2 is known, and (3) prior information on ζ is characterized by the noninformative (improper) prior

$$\mu(\zeta) \propto c \tag{A-3}$$

where c is a constant. Denote the OLS estimate (as well as the classical maximum likelihood estimate) of ζ as $\hat{\zeta}$, and denote the data matrix of regressors in equation A-2 as S.

As shown for example in Box and Tiao (1973, p. 115), the posterior density of the parameter vector ζ given the available data D, $\mu(\zeta \mid D,M)$, is, under our assumptions, multivariate normal. Specifically,

$$\mu(\zeta \mid D) \sim N(\hat{\zeta}, (S'S)^{-1}\sigma_\varepsilon^2) \tag{A-4}$$

The posterior density of any particular coefficient can of course be calculated from this vector density. Under the assumptions justifying A-4, the posterior mean and variance of ζ therefore correspond to the standard OLS estimates of the parameter vector and its associated covariance matrix.

When σ_ε^2 is unknown, the posterior density of ζ can also be characterized and related to OLS estimates. Formally, if σ_ε^2 is unknown and has a noninformative prior

$$\mu(\sigma_\varepsilon^2) \propto \sigma^{-2}, \tag{A-5}$$

then it can be shown (Box and Tiao 1973, p. 117) that

$$\mu(\zeta \mid D,\sigma_\varepsilon^2) \sim N(\hat{\zeta},(S'S)^{-1}\sigma_\varepsilon^2). \tag{A-6}$$

For reasonably large samples, σ_ε^2 can be replaced with the OLS estimate $\hat{\sigma}_\varepsilon^2$ so that, approximately,

$$\mu(\zeta \mid D) \sim N(\hat{\zeta},(S'S)^{-1}\hat{\sigma}_\varepsilon^2) \tag{A-7}$$

and again the posterior mean and variance of ζ may be equated with the corresponding OLS estimates. We refer to this as the "leading case" in the text.

In our evaluation of growth models, we have emphasized the role of ***F***-exchangeable, as opposed to independent and identically distributed errors. De Finetti's theorem provides a link between exchangeability and independence and so motivates our use of this leading case.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Aldous, D. 1983. "Exchangeability and Related Topics." In *École d'Été de Probabilités de Saint Flour XIII*. Lecture Notes in Mathematics Series, no. 1117. New York: Springer-Verlag.

Anderson, E., L. Hansen, and T. Sargent. 1999. "Risk and Robustness in Equilibrium." Department of Economics, Stanford University. Processed.

Arnold, S. 1979. "Linear Models with Exchangeably Distributed Errors." *Journal of the American Statistical Association* 74:194–99.

Auerbach, A., K. Hassett, and S. Oliner. 1994. "Reassessing the Social Returns to Equipment Investment." *Quarterly Journal of Economics* 109:789–802.

Azariadis, C., and A. Drazen. 1990. "Threshold Externalities in Economic Development." *Quarterly Journal of Economics* 105:501–26.

Barro, R. 1991. "Economic Growth in a Cross-Section of Countries." *Quarterly Journal of Economics* 106:407–43.

———. 1996. "Democracy and Growth." *Journal of Economic Growth* 1:1–27.

Ben-Akiva, M., and S. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, Mass.: MIT Press.

Berger, J. 1987. *Statistical Decision Theory and Bayesian Analysis*. New York: Springer-Verlag.

Bernard, A., and S. Durlauf. 1996. "Interpreting Tests of the Convergence Hypothesis." *Journal of Econometrics* 71:161–72.

Bernardo, J., and A. Smith. 1994. *Bayesian Theory*. New York: John Wiley and Sons.

Box, G., and G. Tiao. 1973. *Bayesian Inference in Statistical Analysis*. New York: John Wiley and Sons. Reprinted 2000.

Breiman, L. 1996. "Bagging Predictors." *Machine Learning* 26:123–40.

Camerer, C. 1995. "Individual Decision Making." In J. Kagel and A. Roth, eds., *Handbook of Experimental Economics*. Princeton, N.J.: Princeton University Press.

Canova, F. 1999. "Testing for Convergence Clubs in Income Per-Capita: A Predictive Density Approach." Department of Economics, University of Pompeu Fabra, Spain. Processed.

Clark, G. 1987. "Why Isn't the Whole World Developed? Lessons from the Cotton Mills." *Journal of Economic History* 47:141–73.

Clyde, M., and H. Lee. 2000. "Bagging and the Bayesian Bootstrap." Duke University, Department of Statistics, Durham, N.C. Processed.

Debreu, G. 1960. "Review of R. D. Luce, *Individual Choice Behavior: A Theoretical Analysis.*" *American Economic Review* 50:186–88.

DeLong, J. B., and L. Summers. 1991. "Equipment Investment and Economic Growth." *Quarterly Journal of Economics* 106:445–502.

———. 1994. "Equipment Investment and Economic Growth: Reply." *Quarterly Journal of Economics* 109:803–7.

Desdoigts, A. 1999. "Patterns of Economic Development and the Formation of Clubs." *Journal of Economic Growth* 4:305–30.

Doppelhofer, G., R. Miller, and X. Sala-i-Martin. 2000. "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach." Working Paper no. 7750, National Bureau of Economic Research, Cambridge, Mass.

Draper, D. 1987. "Comment: On the Exchangeability Judgments in Predictive Modeling and the Role of Data in Statistical Research." *Statistical Science* 2:454–61.

———. 1995. "Assessment and Propagation of Model Uncertainty." *Journal of the Royal Statistical Society*, Series B 57:45–70.

———. 1997. "On the Relationship between Model Uncertainty and Inferential/Predictive Uncertainty." School of Mathematical Sciences, University of Bath. Processed.

Draper, D., J. Hodges, C. Mallows, and D. Pregibon. 1993. "Exchangeability and Data Analysis." *Journal of the Royal Statistical Society*, Series A 156:9–28.

Durlauf, S. 2000. "Econometric Analysis and the Study of Economic Growth: A Skeptical Perspective." In R. Backhouse and A. Salanti, eds., *Macroeconomics and the Real World*. Oxford: Oxford University Press.

Durlauf, S., and P. Johnson. 1995. "Multiple Regimes and Cross-Country Growth Behavior." *Journal of Applied Econometrics* 10:365–84.

Durlauf, S., and D. Quah. 1999. "The New Empirics of Economic Growth." In J. Taylor and M. Woodford, eds., *Handbook of Macroeconomics*. Amsterdam: North Holland.

Durlauf, S., A. Kourtellos, and A. Minkin. 2000. "The Local Solow Growth Model." Forthcoming in the *European Economic Review, Papers and Proceedings*.

Easterly, W., and R. Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *Quarterly Journal of Economics* 112:1203–50.

Epstein, L., and T. Wang. 1994. "Intertemporal Asset Pricing Behavior under Knightian Uncertainty." *Econometrica* 62:283–322.

Evans, P. 1998. "Using Panel Data to Evaluate Growth Theories." *International Economic Review* 39:295–306.

Fernandez, C., E. Ley, and M. Steel. 1999. "Model Uncertainty in Cross-Country Growth Regressions." Department of Economics, University of Edinburgh (also forthcoming, *Journal of Applied Econometrics*).

———. 2001. "Benchmark Priors for Bayesian Model Averaging." *Journal of Econometrics* 100:381–427.

Frankel, J., and D. Romer. 1996. "Trade and Growth: An Empirical Investigation." Working Paper 5476, National Bureau of Economic Research. Cambridge, Mass.

Freedman, D. 1991. "Statistical Models and Shoe Leather." In P. Marsden, ed., *Sociological Methodology 1991*. Cambridge: Basil Blackwell.

———. 1997. "From Association to Causation via Regression." In V. McKim and S. Turner, eds., *Causality in Crisis*. South Bend, Ind.: University of Notre Dame Press.

Galor, O. 1996. "Convergence? Inferences from Theoretical Models." *Economic Journal* 106:1056–69.

Gilboa, I., and David Schmeidler. 1989. "Maximin Expected Utility with Nonunique Prior." *Journal of Mathematical Economics* 18:141–53.

Goldberger, A. 1991. *A Course in Econometrics*. Cambridge, Mass.: Harvard University Press.

Granger, C., and H. Uhlig. 1990. "Reasonable Extreme-Bounds Analysis." *Journal of Econometrics* 44:159–70.

Heckman, J. 2000. "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective." *Quarterly Journal of Economics* 115:45–97.

Hoeting, J., D. Madigan, A. Raftery, and C. Volinsky. 1999. "Bayesian Model Averaging: A Tutorial." *Statistical Science* 14:382–401.

Islam, N. 1995. "Growth Empirics: A Panel Data Approach." *Quarterly Journal of Economics* 110:1127–70.

Ivanoff, B., and N. Weber. 1996. "Some Characterizations of Partial Exchangeability." *Journal of the Australian Mathematical Society*, Series A 61:345–59.

Kadane, J., and J. Dickey. 1980. "Bayesian Decision Theory and the Simplification of Models." In J. Kmenta and J. Ramsey, eds., *Evaluation of Econometric Models*. New York: Academic Press.

Kallenberg, O. 1982. "Characterizations and Embedding Properties in Exchangeability." *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 60:249–81.

Kass, R., and A. Raftery. 1994. "Bayes Factors." *Journal of the American Statistical Association* 90:773–95.

Kourtellos, A. 2000. "A Projection Pursuit Approach to Cross-Country Growth Data." Department of Economics, University of Wisconsin. Processed.

Landes, D. 2000. "Culture Makes Almost All the Difference." In L. Harrison and S. Huntington, eds., *Culture Matters*. New York: Basic Books.

Leamer, E. 1978. *Specification Searches*. New York: John Wiley.

———. 1983. "Let's Take the Con Out of Econometrics." *American Economic Review* 73:31–43.

Leamer, E., and H. Leonard. 1983. "Reporting the Fragility of Regression Estimates." *Review of Economics and Statistics* 65:306–17.

Lee, K., M. H. Pesaran, and R. Smith. 1997. "Growth and Convergence in a Multi-Country Stochastic Solow Model." *Journal of Applied Econometrics* 12:357–92.

Levine, R., and D. Renelt. 1992. "A Sensitivity Analysis of Cross-Country Growth Regressions." *American Economic Review* 82:942–63.

Lucas, R. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22:3–42.

Madigan, D., and A. Raftery. 1994. "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window." *Journal of the American Statistical Association* 89:1535–46.

Mallows, C. 1998. "The Zeroth Problem." *American Statistician* 52:1–9.

Mankiw, N. G., D. Romer, and D. Weil. 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107:407–37.

McAleer, M., A. Pagan, and P. Volker. 1985. "What Will Take the Con Out of Econometrics?" *American Economic Review* 75:293–307.

Pack, H. 1994. "Endogenous Growth Theory: Intellectual Appeal and Empirical Shortcomings." *Journal of Economic Perspectives* 8:55–72.

Pesaran, M. H., and R. Smith. 1995. "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels." *Journal of Econometrics* 68:79–113.

Prescott, E. 1998. "Needed: A Theory of Total Factor Productivity." *International Economic Review* 39:525–52.

Pritchett, L. 2000. "Patterns of Economic Growth: Hills, Plateaus, Mountains, and Plains." *World Bank Economic Review* 14:221–50.

Quah, D. 1996a. "Convergence Empirics across Economies with Some Capital Mobility." *Journal of Economic Growth* 1:95–124.

———. 1996b. "Empirics for Growth and Economic Convergence." *European Economic Review* 40:1353–75.

———. 1997. "Empirics for Growth and Distribution: Polarization, Stratification, and Convergence Clubs." *Journal of Economic Growth* 2:27–59.

Raftery, A. 1995. "Bayesian Model Selection in Social Research." In P. Marsden, ed., *Sociological Methodology 1995*. Cambridge: Blackwell.

Raftery, A., D. Madigan, and J. Hoeting. 1997. "Bayesian Model Averaging for Linear Regression Models." *Journal of the American Statistical Association* 92:179–91.

Raiffa, H., and R. Schlaifer. 1961. *Applied Statistical Decision Theory*. New York: John Wiley.

Romer, P. 1986. "Increasing Returns and Long Run Growth." *Journal of Political Economy* 94:1002–37.

Romer, P. 1990. "Endogenous Technical Change." *Journal of Political Economy* 98:S71–S102.

Sala-i-Martin, X. 1997. "I Just Ran Two Million Regressions." *American Economic Review, Papers and Proceedings* 87:178–83.

Schervish, M. 1995. *Theory of Statistics*. New York: Springer-Verlag.

Schultz, T. P. 1998. "Inequality in the Distribution of Personal Income in the World: How It Is Changing and Why." *Journal of Population Economics* 11:307–44.

———. 1999. "Health and Schooling Investments in Africa." *Journal of Economic Growth* 13:67–88.

Sims, C. 1980. "Macroeconomics and Reality." *Econometrica* 48:1–48.

Temple, J. 1998. "Robustness Tests of the Augmented Solow Growth Model." *Journal of Applied Econometrics* 13:361–75.

———. 2000. "Growth Regressions and What the Textbooks Don't Tell You." *Bulletin of Economic Research* 52:181–205.

Tierney, L., and J. Kadane. 1986. "Accurate Approximations for Posterior Moments and Marginal Densities." *Journal of the American Statistical Association* 81:82–6.

Wasserman, L. 2000. "Bayesian Model Selection and Model Averaging." *Journal of Mathematical Psychology* 44:97–102.

White, H. 1994. *Estimation, Inference and Specification Analysis*. Cambridge: Cambridge University Press.

Wolcott, S., and G. Clark. 1999. "Why Nations Fail: Managerial Decisions and Performance in Indian Cotton Textiles, 1890–1938." *Journal of Economic History* 59:397–423.

THE WORLD BANK ECONOMIC REVIEW, VOL. 15, NO. 2 273–275

What have we learned from a decade of empirical research on growth?

Comment on "Growth Empirics and Reality," by William A. Brock and Steven N. Durlauf

Lant Pritchett

World Bank economists are mostly practical people—people who try to answer the question, "What exactly should this particular country do right now?" But if they had hoped that the growth regression lessons summarized in William Brock and Steven Durlauf's article would enhance their practical advice giving, they might feel some dissatisfaction. How would they change their advice to, say, Brazil? But that is why this article is important conceptually. It goes to the heart of the matter by proposing a change in the empirical growth literature's fundamental methodology—from model testing to decision theoretic.

The article's valiant but flawed attempt reveals the difficulties in making this shift, however. I'd like to make three points: There is a tension between the interests of academics and practitioners in growth regressions. Output response heterogeneity is a huge practical problem. And policy decisions can be guided only in broad outlines by growth regressions.

The Tensions between Academic and Practical Interests

What Paul Romer said about the intellectual history is on point: First there was this received model, the Solow model, and then along came others—but nearly all were essentially models of Organisation of Economic Co-operation and Development (OECD) countries. The interest in empirically testing these competing models of the evolution of GDP in technologically advanced countries led in two directions, reflecting two sources of pressure.

The first source of pressure is that even though a model may be about Germany or the United States (technological leaders), data from Guyana and Papua New Guinea and Senegal get recruited into increasing the degrees of freedom for the statistical tests among models. But that should not lead anyone to believe that the model is about Guyana.

The second is that the model testing perspective focuses attention on hypotheses that present a clean separation between alternative models—but not necessarily on what is empirically or practically most important. This may account for the seemingly casual approach to the specification of "policy." Robert Hall,

Lant Pritchett is economist on leave from the World Bank at the Kennedy School of Government.

in a nice paper in honor of Professor Solow's birthday, (Hall 1990) basically points out that if we take Solow's "toy" model literally, everything should be orthogonal to total factor productivity (TFP). The reason is that given the model's assumptions, there cannot be incentives to augment TFP because factor payments to labor and capital exhaust product. From the narrow perspective of model testing, proving that *any* policy affected steady-state or TFP growth by altering incentives provides a clean separation of the models. Empirical work that couldn't possibly contribute to the practical advice but provided a clean separation between competing models became academically important.

The Problem of Output Response Heterogeneity

Assume that we are considering a permanent shift in the value of policy P from level P to P' at time t. A key question is, "What is the difference between properly measured GDP with policy P' and that with policy P at time t, $t + 1$, $t + 2$, out to t + N?" A huge problem with empirical estimation of an output response function, however, is that there is no reason to believe that it looks the same—across countries or over time.

Brock and Durlauf's article does a good job of emphasizing this problem by focusing on parameter heterogeneity. But even characterizing the problem as parameter heterogeneity—and, especially, limiting heterogeneity to a small number of linear interactions—artificially narrows the problem, as output response function could vary in many ways for many reasons. It could depend on structural differences across economies, on economic structure, on institutional differences that would mediate the policy change, or on complementary policies that could affect the output response of a given policy.

This proliferation of parameters for characterizing output response makes the empirical situation seem bad—but it is much worse than that. Even for countries observationally equivalent in terms of structure, output response dynamics could depend on timing, expectations, and history. So a policy that may be perfect for bringing a country out of a recession might be neutral, or even counterproductive, at the peak of the cycle. Expectations play a huge role in output response—observationally equivalent policy changes can potentially have enormously different impacts depending on whether all people believe that the policy change will persist or all believe that it will not persist. Finally, there may be output response function hysteresis, in which the output response function depends not just on conditions today but on an economy's entire history of policy changes and their impacts.

The Limits of Growth Regressions as a Source of Policy Advice

The policy variables that go into growth regressions are at a level of abstraction far greater than that at which policy recommendations and decisions are made and implemented in the real world. So although it is possible to come up with a growth regression that says that, on average, countries that are more open tend to grow faster, that leaves a million questions about trade policy reform unanswered: Should the tariff be lowered on this good or that set of goods? Should tariff reductions be *concertina* or from the top down? Should changes be carried out in one stroke or phased in?

Suppose we take the decision-theoretic approach to empirics seriously. The inevitable problem is that the level of specificity at which most growth economists need to work is far greater than can ever be adequately informed by growth regressions. Some people act disappointed that we haven't learned more from growth regressions—but we have to live with the fact that growth regressions are not going to tell us what the tariff on capital goods in Brazil should be in 2001.

Growth regressions are incredibly useful in providing a general empirical background of stylized facts about the world. The partial associations of policy variables with growth provide a grounding in reality from which policy discussions can build. But none of us is in any danger in the near term of being replaced by an automaton based on growth regressions that takes in country conditions, searches the data, and then spits out policy solutions. Policy decisions draw on a variety of information and remain the domain of that most elusive of qualities: good human judgment.

Reference

Hall, Robert. 1990. "Invariance Properties of Solow's Productivity Residual." In Peter Diamond, ed., *Growth/Productivity/Unemployment: Essays to Celebrate Bob Solow's Birthday*. Cambridge, Mass: MIT Press, pp. 71–112.

THE WORLD BANK ECONOMIC REVIEW, VOL. 15, NO. 2 277–282

What have we learned from a decade of empirical research on growth?

Comment on "Growth Empirics and Reality," by William A. Brock and Steven N. Durlauf

Xavier Sala-i-Martin

William Brock and Steven Durlauf's article nicely summarizes some of the recent research on Bayesian model averaging. They make a number of important points. One is that the empirics of growth face three key problems: model uncertainty, parameter uncertainty, and endogeneity. They argue that theory uncertainty can be dealt with using Bayesian model averaging methods. Their key equations are 16, 17, and 18, for which the interpretation is as follows. Suppose you are interested in the distribution of the partial derivative of the growth rate with respect to variable z, b_z. Let each set of every possible combination of explanatory variables be called a "model." Conditional on each model there is a distribution of b_z for a given data set. Equation 17 says that the posterior distribution of b_z is a weighted average of all these individual distributions, where the weights are proportional to the likelihoods of the models. Equation 18 says that the mean of this distribution is the weighted average of the ordinary least squares (OLS) estimates of all these models, where the weights are proportional to the likelihoods. Equation 19 makes a similar claim about the variance.

The assumption that weights are proportional to the likelihoods is an important one. In fact, it may drive the authors' first key empirical result—that Easterly and Levine's (1997) regression of growth on ethnolinguistic fractionalization (ELF) is "robust" to Bayesian model averaging analysis. It is important to remember that models with more explanatory variables have larger likelihoods. It is also important to remember that Brock and Durlauf perform Bayesian model averaging analysis by combining the explanatory variables of the Easterly and Levine paper in all possible ways: sets of one right-hand-side variable, sets of two, sets of three, and, eventually, one set with all the right-hand-side variables. This last model is the one run by Easterly and Levine and the largest model run by Brock and Durlauf (and therefore the one that is likely to have the largest likelihood and that gets the largest weight).

Hence, it is not surprising that the weighted average of all the models is similar to that for Easterly and Levine's model, because most of the weight of the average goes to Easterly and Levine's specification, by construction. In other words, the finding that Easterly and Levine's regression results (column 1 in table

Xavier Sala-i-Martin is at Columbia University and UPF.

1) are "robust" to the Bayesian model averaging analysis because the weighted average of models (column 2) is virtually identical is likely to be an artifact of the weights used.

I should confess that these are also the weights I used in a 1997 paper (which has equations 16 and 17 in exactly the same form). However, in that work I averaged only regressions with a fixed set of explanatory variables, so I did not have the problem that I am pointing out here. Doppelhofer, Miller, and Sala-i-Martin (2000) derive an alternative weighting scheme. The posterior density of model M_m is proportional to the likelihood (sum of squares of residuals, or $SSE_m^{-T/2}$), multiplied by $T^{-km/2}$, where T is the number of observations and k_m is the number of explanatory variables in model m:

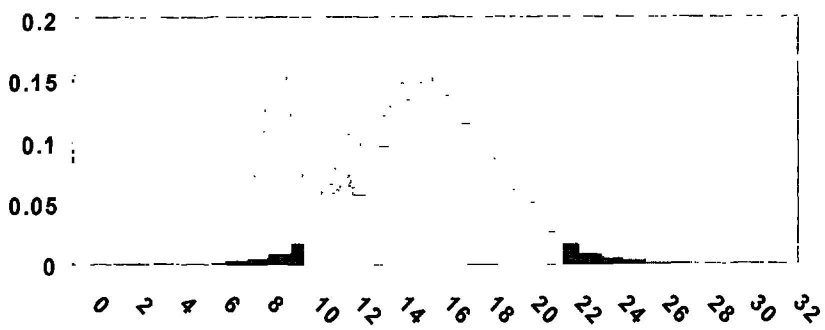
$$\mu\langle M_m | D\rangle = \frac{\mu(M_n)T^{-k_m/2}.SSE_m^{-T/2}}{\sum_{i=1}^{2^K}\mu(M_i)T^{k_m/2}.SSE_i^{-T/2}}$$

Note that this weighting scheme penalizes larger models. It would be interesting to see whether column 1 still looks very much like column 2 when these alternative weights are used.

A second important assumption is the prior that allows Brock and Durlauf to eliminate the $\mu(M_m)$ from equation 15 to derive equation 16. They use the prior that "all models are equally likely." Imagine that we had 32 possible right-hand-side variables. If we believe that all models are equally likely, the prior distribution of model sizes is as shown in figure 1. The average model size is 16. If instead we had 10 explanatory variables, the implicit assumption would be that the average model size of the prior distribution of cross-country regressions is 5.

The problem is that Brock and Durlauf propose that when analyzing (or discussing) a paper like Easterly and Levine (1997), we take the key regression in that paper and perform Bayesian model averaging analysis with it. If we take

FIGURE 1. Prior Probabilities by Model Size: Equal Model Probabilities



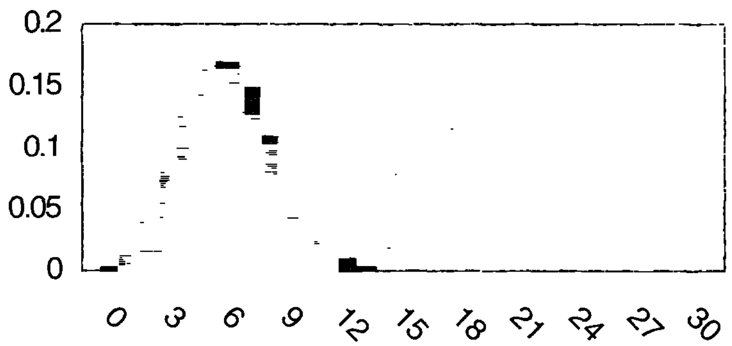
this proposal literally, we would implicitly assume that the average model size of "the growth regression" is 5 when the original paper had 10 variables, and 16 when the original paper had 32 variables. Besides being arbitrary, this assumption does not make sense: The prior model size should be invariant to the paper being discussed.

One solution to this problem, following Doppelhofer, Miller, and Sala-i-Martin (2000), would be to specify the model prior probabilities by choosing a prior mean model size, k, with each variable having a prior probability κ/K of being included, independent of the inclusion of any other variables, where K is the total number of potential regressors (figure 2). Equal probability for each possible model is the special case in which $\kappa = K/2$. The prior distribution of model sizes would be invariant to the paper analyzed. Moreover, the robustness of this prior could be checked by redoing the Bayesian model averaging exercise (or better yet, Bayesian averaging of classical estimates) for different values of κ.

My third comment relates to the treatment of parameter uncertainty. I agree with the authors that this problem is analogous to that of theory uncertainty. But if so, why do they propose a different solution? If we think that Africa needs a different slope for variable z, all we need to do is to construct a new variable (z times one for countries in Africa and z times zero otherwise) and put this new variable in the pool of potential variables to be included in the Bayesian model averaging analysis. Rather than columns 3–6, table 2 should include a row presenting the distribution of the $\beta_{(j<2)}$ for this new variable, as a regular additional variable subject to theory uncertainty.

When we think of parameter uncertainty as another form of theory uncertainty, an additional problem comes to mind. Why do we think that Africa needs its own slope? Why don't we have a special slope for Christian countries? Or hot countries? Or small countries? Of course, we do not know whether or not special slopes are needed (we do not have a theory, or we can have many open-ended theories that would call for a special slope for each of these country groups). However, in the spirit of Durlauf and Johnson (1995), shouldn't we then per-

FIGURE 2. Prior Probabilities by Model Size ($\kappa = 7$)



form Bayesian model averaging or Bayesian averaging of classical estimates for each group of countries? How would we go about that?

A perhaps related question is that of nonlinearities, which Brock and Durlauf do not allow for in their article. It is clear that African countries have both lower average growth and greater ethnolinguistic fractionalization. The conditional data might therefore look like figure 3. If we think about the implications of figure 3, we arrive at the conclusion that if we could somehow reduce ELF for African countries, Africa will conditionally grow faster than the rest of the world forever (that is, we would move the African data points to the left along the steeper regression line). Because we do not have a theory of ELF, we do not know whether this is sensible or not.

Alternatively, we could think that the partial relationship between growth and ELF looks like figure 4. In fact, the data points in figures 3 and 4 are exactly the same. The only thing that differs is the functional form of the regression curve. Under this interpretation, if Africa manages to get the same ELF as the rest of the world, its growth rate will also be similar. Hence the economic implications of a separate slope for Africa are very different from those of a nonlinear relationship. It would have been interesting to incorporate nonlinearities in the analysis.

Finally, the claim that growth economists have not dealt with parameter uncertainty is not quite true. In fact, parameter uncertainty is a particular form of what economists usually label *interaction terms*. For example, suppose a claim is made that the partial derivative of growth with respect to z depends on variable y:

$$\frac{\partial g_i}{\partial z_{ji}} = \beta_z + \beta_{z,yi} y_i$$

FIGURE 3.

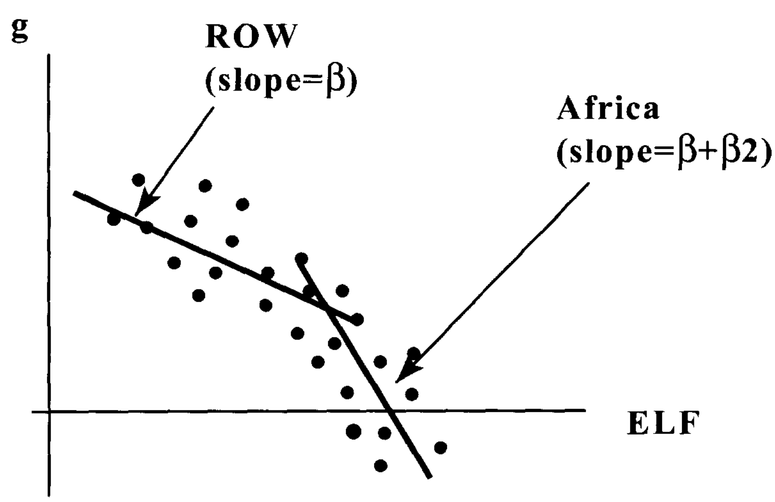
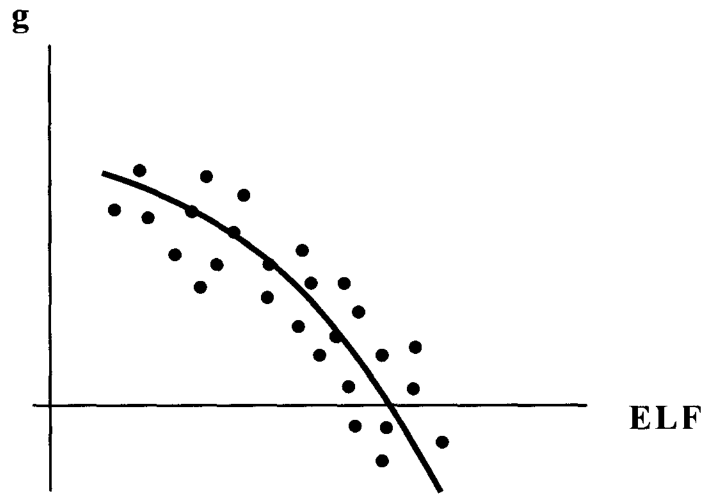


FIGURE 4.



The way to test this claim would be to run a regression of growth with z as an explanatory variable and with an additional variable that is a country-by-country product of z times y. That is, we should introduce interaction terms. It should be clear that parameter uncertainty is nothing but an interaction term when variable y is simply a dummy variable for a region (in this case, Sub-Saharan Africa). To the extent that growth economists have introduced interaction terms, therefore, they have allowed for parameter heterogeneity.

I conclude with two sources of disappointment about this otherwise excellent article. First, the article is not really about the empirics of economic growth. All empirical analyses are subject to the problems it discusses, especially those forced to use small data sets. In this sense the title, though cute, is highly misleading and, to the extent that it leads future researchers away from economic growth analysis, potentially damaging. A more appropriate title would be "Small-Sample Econometrics," because the problems discussed are common to all empirical analyses with small samples (which include all cross-country analyses in any field). After all, if we had a huge data set with zillions of observations, we could simply throw in all potential variables, with particular slopes for each potential set of countries, with all potential nonlinearities, and so on—and the data would tell us which coefficients are zero and which are not. The fact that we have more potential variables than we have countries prevents us from following this strategy, and this is where the problem starts. But this is a problem of small samples, not growth econometrics.

Second, although the authors introduce endogeneity as an important problem early in their article, I was disappointed to find that they went no further. Given the authors' reputation, I was excited when I started reading the article about the prospect of a potential solution, perhaps along the lines of Bayesian model averaging. But no solution was offered.

References

Doppelhofer, G., R. Miller, and X. Sala-i-Martin. 2000. "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach." NBER Working Paper no. 7750, National Bureau of Economic Research, Washington, D.C.

Durlauf, S., and P. Johnson. 1995. "Multiple Regimes and Cross-Country Growth Behavior." *Journal of Applied Econometrics* 10:365–84.

Easterly, W., and R. Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *Quarterly Journal of Economics* 112(4):1203–50.

Sala-i-Martin, X. 1997. "I Just Ran Two Million Regresssions." *American Economic Review, Papers and Proceedings* 87:178–83.

THE WORLD BANK ECONOMIC REVIEW, VOL. 15, NO. 2 283–288

What have we learned from a decade of empirical research on growth?

Applying Growth Theory across Countries

Robert M. Solow

I am broadly in sympathy with the spirit of the article by William Brock and Steven Durlauf and that by William Easterly and Ross Levine. They are trying to move the literature in the right direction. I say this even though I have been skeptical from the beginning about the interpretation of cross-country growth regressions.

The potential problem of reverse causality has been obvious to everyone. It has usually been met with the standard econometric dodge: using lagged values of slow-moving variables as instruments. But this cannot be a serious solution to the problem. The causality issue points to a deeper question: Do cross-country regressions define a meaningful surface along which countries can move back *and* forth at will? If this is the idea, what mechanism could underlie such a surface? Brock and Durlauf call such a regression a "model." I suppose in a statistical sense it is. But an economic model should have some internal structure; its causal arrows should rest on some sort of behavioral mechanism, and that seems to be missing in this literature.

I think I had this prejudice even before cross-country regressions became fashionable. I thought of growth theory as the search for a dynamic model that could explain the evolution of one economy over time. There were no explicit cross-sectional implications. Were there implicit ones? Certainly, and my comments bear on the question of what they might be.

A Justification for Multivariable Cross-Country Growth Regressions?

In my view growth theory was conceived as a model of the growth of an industrial economy. Its parameters certainly could not be regarded as fixed forever, but maybe they would need to be reconsidered only over intervals of 30–50 years, long enough so that the differences between endpoints could not be dominated by demand-driven business cycles. So far as I can remember, I have never applied such a model to a developing economy, because I thought the underlying machinery would apply mainly to a planned economy or a well-developed market economy. This is not a matter of principle, just wariness.

Robert M. Solow is with the Department of Economics at the Massachusetts Institute of Technology. An earlier version of this article was presented at the World Bank conference "What Have We Learned from a Decade of Empirical Research on Growth?" held on 26 February 2001.

Now suppose that you want to compare several such economies and make inferences from the comparison. I use these general terms to emphasize that the cross-country regression is not the only way to make comparisons. You could intend just to interpret parallel time series for a small group of countries, with the goal of understanding the source of differences among them. What background assumptions do you have to make if such a comparison is to be sensible?

The economies you are comparing must have something in common, some part of the driving mechanism. If Robert Summers and Alan Heston were suddenly to discover national income and product accounts from an economy that existed on Mars a million years ago, you would not expect that economy to fit neatly into a Barro regression. So the economies must have some things in common, but not everything. One possible specification that early writers tended to make in this spirit, perhaps automatically, and that some perhaps still make, is that the economies in question share common technological knowledge. The normal justification for this assumption was that technological handbooks were easily available everywhere, even before the Internet. So the basic commonality could be the knowledge of a production function $F(K,L,H;A)$, not necessarily Cobb-Douglas in form. Of course, countries would have different values of K, L, and H, but a strict interpretation would give them the same value of $A(t)$.

Within the model, which leaves only the saving-investment pattern, the growth rate of employment, and the rate of depreciation—as well as initial conditions, of course—to differ from country to country. Those are the implications to be explored. We are all familiar with exercises like this. Right or wrong, they are coherent. But then what is the role of all the other right-hand-side variables—openness to trade, size of government, black market premium, and degree of inequality, to name just a few that appear frequently? Maybe we should regard them as purely descriptive, a search for empirical correlations with no analytical implications. That would be a respectable occupation. But to stay within the model, we have to think about the role of total factor productivity (TFP) or $A(t)$.

When I used to teach growth theory, I would always begin by saying, "Let's imagine a toy economy that produces only one homogeneous good, using as inputs just (the services of) a stock of the good itself and a flow of labor." In that context it is natural to think of $A(t)$ in purely technological terms. That may be how the habit was established of supposing that the shape of the production function *and* the path of $A(t)$ were common across countries. But soon questions would arise; if they didn't, I would raise them myself.

For example, someone would be sure to wonder if there really is significant substitutability between labor and capital, at least enough to make the model interesting. The routine reply is, "That is an empirical question, so we have to go outside the narrow model." Empirically speaking, every industrial economy produces not one but thousands of different goods and services. Even if each of them operates with a Leontief technology, they exhibit a wide variety of capital-labor ratios, from the very labor-intensive, such as personal services, to the very capital-intensive, like electricity generation. The economy as a whole—what the toy economy symbolizes—can substitute between capital and labor by changing the composition of output in the obvious way. Moreover—and this is very important—there is an elementary market mechanism to make this happen: If capital becomes scarce, its rental price will rise relative to the wage; the price of capital-intensive goods will rise relative to the price of labor-intensive goods; demand will shift to labor-intensive goods; and the aggregate capital-labor ratio will fall. (And that is not the end of the process.)

Once you go down that road, you have to rethink the production function and especially the role of $A(t)$. It is certainly unwise to assume that all economies are equally efficient at reallocating inputs across industries. This difference in efficiency would be reflected in $A(t)$, and maybe not only there. As soon as that thought enters your mind, it immediately occurs to you that there are many other nontechnological factors that could influence the level and growth of TFP. They would include the intensity of competition—domestic or foreign—because that would influence the amount of waste and slack in various industries, the alacrity with which the national economy adopts new technology, and thus the level and growth of TFP. You can just as easily imagine that the amount and nature of regulation in a country can affect the efficiency of resource allocation, and thus the "effective" level of TFP and quite possibly its rate of growth. Even among Organization for Economic Co-operation and Development (OECD) countries analysts have found that substantial international differences in productivity can persist even within a narrowly defined industry.

It is easiest to think of such institutional differences as if they could be summed up as international differences in TFP, but the situation could be more complicated. We usually model TFP as if it were a Hicks-neutral multiplicative factor. That might be harmless if the production function is Cobb-Douglas. But that assumption is rarely tested. (Michael Boskin and Lawrence Lau [2000] have tested it in a cross-country panel context involving a dozen or so countries and a translog technology, and they reject it strongly.) One obvious generalization is to allow arbitrary factor augmenting technological change, so that the production function would be written as $F(A(t)K, B(t)L, C(t)H)$.

Does this line of thought provide a justification for multivariable cross-country regressions? Probably. But it also suggests focusing more directly on TFP or factor augmentation functions as the proper left-hand-side variables in empirical work and thinking more seriously about right-hand-side variables that might legitimately account for differences in TFP or in $A(t)$, $B(t)$, and $C(t)$. Current practice seems to be much too haphazard.

Growth Theory as the Theory of the Evolution of Potential Output

One might protest that only pure labor augmentation allows the existence of a steady-state growth path. Maybe that should not be a concern. Easterly and Levine point out that observed growth paths do not look like steady states, except perhaps in the United States. This could be related to the nature of factor augmentation. In heavily agricultural economies, for instance, weather and disease could play the role of disturbances to steady-state growth, and their effects would not be represented as Hicks-neutral. But an old Keynesian like me suspects that deviations from steady-state growth are often related to deviations of output from potential output—often demand failures with sticky prices, possibly export failures with sticky prices.

A serious attempt to study the nontechnological components of TFP should also get away from the indefensible presumption that actual output is always close to potential output. In poor countries, especially monocultural primary producers, the choice of off-potential end points, even 20–30 years apart, can materially distort the measured growth of output and maybe even more so the measured growth of TFP. Industrial countries have been less vulnerable to this problem in the postwar period. But Japan has produced nowhere near its potential output for a decade, and I have my doubts about contemporary Europe.

This point may be worth developing. I think of growth theory as precisely the theory of the evolution of potential output. So it is mostly concerned with the supply side of the macroeconomy. Deviations are demand driven. In advanced industrial countries we have ways of estimating the growth of potential output. Some depend on a version of Okun's Law; others work directly with production functions. But all are explicitly trying to measure potential output. That is why I think of real-business-cycle models and the indiscriminate use of the Hodrick-Prescott filter as intellectually backward steps.

For other contexts we need other methods. To take the simplest example, agricultural economies, especially single-crop ones, are subject to large fluctuations in output and output growth stemming from such things as droughts and pests. Easterly and Levine provide examples. How should we deal with such observations? We can define potential output as the output that could be produced with normal weather. An estimate of that quantity (or the TFP derived from it) should be on the left-hand side of a growth regression. Or else we need a weather or disease prevalence variable on the right-hand side. In the first case weather-induced fluctuations are just removed; in the second case they are included in TFP but segregated. Without such makeshifts, we are asking for trouble in cross-country studies that include many countries at different stages of development. If those studies are growth-oriented, they should be aimed at explaining potential output or TFP.

In effect I am agreeing strongly with the advice of Easterly and Levine: Comparative growth studies should focus on understanding and analyzing the various sources of differences in TFP and the policies that might affect them. This goes for both technological and nontechnological factors in TFP. My residual doubt is whether using cross-country regressions involving large numbers of countries with different institutional histories is the best way to go about this task. I would prefer to start with qualitative studies of basically similar countries, extended over time and space.

This could be taken as a step toward endogenous growth theory, and so it is. The good thing about the fox that Paul Romer started chasing more than 15 years ago is that it leads us to focus on the analysis of the economic incentives to create new technology. I have two suggestions for those engaged in this hunt. First, close attention to what goes on in research and development enterprises might pay off more at this stage than simple mechanical modeling that may be off the mark. I once had the opportunity to observe the General Motors research laboratories at close hand; the incentives and responses were anything but simple.

Second, the nontechnological sources of differences in TFP may be more important than the technological ones. Indeed, they may control the technological ones, especially in developing countries. Obvious examples include things like the security of contracts, the intensity of competition, and respect for instrumental rationality as a mode of behavior. There may be less obvious ones.

Conclusion

I conclude by returning to an old hobbyhorse of mine. Our modeling exercises are usually carefully tailored to lead, all too transparently, to a steady-state growth rate. This habit induces analysts to make gratuitous linearity assumptions and to impose other more or less arbitrary restrictions on models. But maybe, as Easterly and Levine argue, steady-state behavior is a rarity outside a few successful advanced industrial countries. Nicholas Kaldor's stylized facts may still have relevance, but not everywhere.

One of the advantages of vast computer power is that theory does not need easy special cases, such as purely labor augmenting technological change or the analogous assumption that allows a steady state to be an attractor even without constant returns to scale. Numerical integration or iteration can answer the kinds of questions we want to ask of a model, even if the model is not tractable with pencil and paper. Correspondingly, it is probably not a good idea to set cross-country regressions the task of explaining the rate of growth or some other stationary characteristic. It is time paths that need to be modeled and studied.

Circumstances have given these remarks a discursive character. So I would like to end by distilling my three main points:

- If they mean anything at all, those many right-hand-side variables in growth regressions are determinants of TFP. But then they should be selected with that function in mind, and TFP (or its growth rate) should be the left-hand-side variable. Moreover, some of those factors could affect other characteristics of the aggregate production function. Allowing for separate factor augmentation functions would be a first step, but there are other possibilities. It may be necessary to think about genuine estimation of the underlying production function.
- The proper measure of output underlying the left-hand-side variable is potential output. In industrial countries there are standard methods of

approximation (not all of which are satisfactory). The deviation of actual from potential output is mainly demand driven. Some of the sharpest deviations of actual from potential output occur in primary producing countries. These deviations may be related to weather fluctuations or disease, but export failures may be a demand-side source. There are various thinkable ways to deal with this issue; they need to be systematized.

- The exponential steady state is a theoretical convenience. But many countries, much of the time, are nowhere near steady-state growth. This suggests that comparative studies should focus less on the growth rate and more on comparing and understanding whole time paths.

Reference

Boskin, Michael, and Lawrence J. Lau. 2000. "Generalized Solow-Neutral Technical Progress and Postwar Economic Growth." Discussion Paper 00-12. Stanford Institute for Economic Policy Research, Stanford University, and Working Paper 8023, National Bureau of Economic Research, Cambridge, Mass.

THE WORLD BANK ECONOMIC REVIEW, VOL. 15, NO. 2 289–314

Crisis Transmission: Evidence from the Debt, Tequila, and Asian Flu Crises

José De Gregorio and Rodrigo O. Valdés

This article analyzes how external crises spread across countries. The authors analyze the behavior of four alternative crisis indicators in a sample of 20 countries during three well-known crises: the 1982 debt crisis, the 1994 Mexican crisis, and the 1997 Asian crisis. The objective is twofold: to revisit the transmission channels of crises, and to analyze whether capital controls, exchange rate flexibility, and debt maturity structure affect the extent of contagion. The results indicate that there is a strong neighborhood effect. Trade links and similarity in precrisis growth also explain (to a lesser extent) which countries suffer more contagion. Both debt composition and exchange rate flexibility to some extent limit contagion, whereas capital controls do not appear to curb it.

The increasing globalization of the economy has put the issue of transmission of crises across countries in the front line. Although the word contagion is a rather new concept in international finance, it has been the focus of a large number of policy-oriented seminars and debates. Both regional and time clustering of currency crises are at the heart of the discussion. There are several important questions that need to be answered. In this article, we focus on two of them. First, what are the propagation channels of international crises across countries (other than common shocks)? Second, are there useful policy instruments for shielding countries from contagion? In particular, do capital controls, exchange rate flexibility and the external debt maturity structure affect contagion? We seek to answer these questions using evidence from three key events: the 1982 debt crisis, the 1994 Mexican devaluation, and the 1997 Asian crisis.

There is an ongoing discussion about the proper definition of contagion (see, for example, Kaminsky and Reinhart 1998; Forbes and Rigobón 1999). Here we simply refer to it as the co-movement suffered by countries during crisis periods and that is unexplained by initial conditions or common shocks. It is a characteristic of crises because it is precisely during these periods in which the

José De Gregorio is with the Ministry of Economics, Mining and Energy in Chile, and the Department of Industrial Engineering at the Universidad de Chile (*jdegregorio@minecon.cl*). Rodrigo O. Valdés is with the Ministry of Finance in Chile (*rvaldes@minhda.cl*). This article is part of the WIDER/World Bank research project "Contagion: How It Spreads and How It Can Be Stopped." Rodrigo O. Valdés acknowledges partial financial support from FONDECYT grant 1990338. The authors thank Ilan Goldfajn, Leonardo Hernández, Guillermo Perry, Carmen Reinhart, Roberto Rigobón, and two anonymous referees for helpful suggestions, and Pamela Mellado for excellent research assistance.

issue is important from a policy perspective. Nevertheless, as Rigobón (1999) emphasizes, contagion could be confused with the presence of a large common shock. In our empirical investigation, we attempt to separate the effects of contagion from other large common shocks. However, because we select crisis periods, we cannot strictly compare whether they are essentially of a different nature than tranquil times. This issue has led many to question the view that contagion is a particular phenomenon during crisis and is different from simple interdependence. We do not solve this problem, although we compare different transmission mechanisms through which interdependence across countries occurs.[1]

This article is closely related to other studies of contagion, particularly those that analyze the existence of contagion and the likelihood of alternative propagation channels by examining a number of currency crises. According to Eichengreen, Rose, and Wyplosz (1997) and Glick and Rose (1998), trade links are the key transmission channel of crises across countries. While the first study focuses on Organisation for Economic Co-operation and Development (OECD) countries, the second studies five international crises using a narrower form of contagion than the one we use, namely, contagion originating from "ground zero." Kaminsky and Reinhart (1998) claim that financial links are potentially an important transmission mechanism. However, they argue that because of the high correlation between trade and financial links, it is difficult to distinguish between both channels. We revisit the existence of contagion as well as the most likely transmission channels.

Instead of focusing on transmission from ground-zero countries to the rest of the world, we look at the impact of crises elsewhere on the likelihood that a country will suffer a crisis. This allows us to study the fact that many times contagion happens from country A to country B, but what may cause problems in country C is not a crisis in A, but the problems in B. A typical case we have in mind is that a crisis in Mexico may affect Chile more through its impact on Argentina and Brazil than through the crisis in Mexico itself. For this reason, focusing on ground-zero countries could give an incomplete picture of the evidence.

Section I discusses our basic empirical approach. Section II provides evidence of the existence of contagion and investigates the transmission channels behind this phenomenon. Section III investigates the extent to which capital controls, exchange rate flexibility, and debt structure shield countries against contagion effects. Section IV presents concluding remarks.

I. Empirical Approach

This section describes our empirical methodology. To measure contagion or transmission of crises across countries, we follow an approach that combines previous work by Sachs, Tornell, and Velasco (1996); Eichengreen, Rose, and Wyplosz

1. We use indistinctly the expressions contagion, interdependence, and co-movements.

(1997); and Glick and Rose (1998). In particular, we try to explain the cross-sectional variation in alternative crisis indicators during particular events using (i) a set of initial macroeconomic conditions, and (ii) a weighted average of the evolution of the crisis indicator in other countries. With (i), we seek to control for country-specific characteristics that may directly explain the extent of crises as well as common factors that affect countries differently depending on macroeconomic characteristics (for example, an international interest rate shock). With (ii), we seek to measure and characterize contagion. Because alternative weighting schemes can be associated a priori with different transmission channels, we are able to study what may drive contagion.

We focus the analysis on three important events of the past 25 years from the perspective of developing countries: crisis 1, the 1982 debt crisis; crisis 2, the 1994 Mexican devaluation; and crisis 3, the 1997 Asian crisis. In the spirit of Glick and Rose (1998), we identify a ground-zero country for each crisis and date the episode accordingly. This is used just to date the beginning of the crisis, not to define how it spreads to other countries. We assume that when the crisis begins, all countries are subject to contagion. We use a dummy to control only for the ground-zero country, which captures the fact that this country by definition cannot suffer from contagion.

In the case of the debt crisis, we use Mexico as the ground-zero country and date the initial period of the crisis in August 1982, when Mexico announced a moratorium on its external debt. In the case of the tequila crisis, the ground-zero country is naturally Mexico and the initial date is December 1994. Finally, we consider that the Asian crisis started in Thailand in July 1997.

We analyze the performance of four alternative crisis indicators in 20 countries, 8 from Latin America, 6 from Asia, and 6 controls (small, open OECD countries). Appendix table A-1 lists the countries as well as their neighborhood codes.

Measuring Contagion

To measure contagion, we explain the performance of crisis indicators in the countries, using particular averages of what happens in other countries. More formally, indexing countries by i ($i = 1, 2, \ldots, 20$) and crises by j ($j = 1, 2, 3$), we estimate cross-section models of the following form:

$$(1) \qquad \Delta CI_{i,t,j} = \beta_0 + \beta_1 X_{i,j} + \beta_2 \sum_{k \neq i} M_{i,k,j} \Delta CI_{k,t,j} + \beta_3 \sum_{k \neq i} M_{1,1} \Delta CI_{k,t,j} + \in_{i,t,j},$$

where $\Delta CI_{i,t,j}$ denotes the change in crisis indicator CI in country i, during crisis j, between one month before that crisis and month t; $X_{i,j}$ is a vector of initial macroeconomic conditions in country i prior to crisis j; $M_{i,k,j}$ is a fixed number that weights ex ante the importance of country k in explaining the performance of country i; $M_{1,1}$ is a fixed number that weights equally all countries different from i; and $\varepsilon_{i,t,j}$ is a random shock.

We construct a series of matrixes with weights $M_{i,k,j}$ to calculate particular linear combinations of other countries' returns. Each linear combination represents a particular theory of contagion.

The $M_{1,1}$ allows us to control for the effect of the size of each crisis. In other words, it controls for the effect of the common shock that occurs elsewhere. After normalizing the weights, this is equivalent to adding for each country the average crisis in all other countries. If we had a very large sample, this could be approximated by the average across countries, and solved by including a dummy variable for each crisis. However, in our sample, this could lead to biases as long as countries subject to large shocks—that is, large changes in the crisis indicator—also have a large weight in the average change in the crisis indicator. There would be an obvious and strong upward bias because the country with a large weight would be included in both the left- and right-side variables. For this reason, we exclude the country when computing the average external shock for each observation.

When the true β_2 is positive (that is, there is contagion) and the $M_{i,k,j}$ weights are nonnegative, the ordinary least squares (OLS) estimation of equation 1 has a positive bias.[2] A shock in $\in_{i,t,j}$ that triggers a crisis in a country will affect, through contagion, the performance of other countries; the other countries, in turn, will affect country i's performance, introducing a positive correlation between the error term ($\in_{i,t,j}$) and one of the regressors ($\Sigma_{k\neq i} M_{i,k,j}\ \Delta CI_{k,t,j}$). However, because this bias is monotonic in β_2 and hence there is no bias when β_2 is zero (and there is negative bias when $\beta_2 < 0$), the issue is not a serious problem for our particular purposes. As long as we focus on comparing alternative models, it is valid to compare different OLS estimates of β_2. The same is true when we compare alternative measures for curbing contagion. In a very large sample, this effect would not exist because the feedback from a single country to others would be small. Here we presume this is also small; as long as there are about 20 countries per episode, the effect of a particular $\varepsilon_{i,t,j}$ should be small.

We consider the following four crisis indicators:

- A foreign exchange market pressure index at a three-month horizon after the crisis, denoted by *PI-3*.
- A foreign exchange market pressure index at a 12-month horizon after the crisis, denoted by *PI-12*.
- The level of the real exchange rate 12 months after the crisis, denoted by *RER*.
- A credit rating indicator, denoted by CR.

When using indicators with the same time horizon in different crises, we are implicitly assuming that the three crises have similar contagion patterns in the time dimension. This does not need to be the case. The credit rating measure partially takes into account this issue.

In constructing *PI-3* and *PI-12*, we follow the standard procedure of calculating a weighted average of changes in the real exchange rate and the stock of international reserves in each country/observation. In the case of crises 2 and 3, we also include (minus) the change in the real interest rate with respect to the 12-

2. We consider only nonnegative $M_{i,k,j}$ weights.

month average level observed prior to the crisis. As in Kaminsky and Reinhart (1999), we weight each component of the index such that each one has equal (crisis-specific) volatility. A negative change in *PI* shows an increase in market pressure that may arise from any of the three components.[3] We use data from International Monetary Fund (IMF; various years) for international reserves, interest rates (short-run deposits), and inflation. We use the JP Morgan database for real exchange rates, in which a downward movement in *RER* means depreciation.[4]

For credit rating, we use the credit risk indicator compiled by *Institutional Investor*. Because it is published only in March and September of each year, we are not able to have a perfect dating for each crisis. However, this allows us to select the horizon we consider more appropriate in each crisis. For crisis 1, we use the 1-year change in the index published in March 1983; for crisis 2, we use the 6-month change published in September 1995 (which seems to better capture the Mexican downgrade); and for crisis 3, we use the 12-month change published in March 1998.

The 60 × 60 matrix with weights $M_{i,t,j}$ can take several forms. However, because cross-crisis contagion makes little economic sense, we restrict it to a block diagonal with three 20 × 20 submatrixes. Moreover, because we are not interested in explaining contagion suffered by ground-zero countries, the matrixes have zeros in the respective row. Furthermore, to avoid running regressions in which an independent variable is a function of that same dependent variable, we restrict the main diagonal to be zero. We follow the same procedure when constructing the $M_{1,1}$ matrix of equal weights. In any case, the concept of own contagion does not make sense.

Depending on the exact definition of contagion, there are two alternative classes of weighting matrixes. If contagion is defined as occurring exclusively from the ground-zero country to other countries, then the matrix has to have nonzero elements only in the columns corresponding to the ground-zero country. This is the approach taken by Glick and Rose (1998). Alternatively, if contagion is defined more broadly as transmission of crises from a particular set of countries to others, then the nonzero elements could appear anywhere in the 20 × 20 matrixes, except in the row of the ground-zero country. This is the approach followed by Eichengreen, Rose, and Wyplosz (1997) in trying to explain the probability of crisis (a binary variable) in a group of OECD countries. They consider that there is contagion as long as a weighted "crises elsewhere" variable affects the probability of crisis in an individual country.[5] We focus our analysis on the second type of contagion, although we also analyze the first type.

3. None of the results change in any important way if we exclude from *PI* interest rates for crises 2 and 3.

4. Because of dramatic jumps unrelated to the crises, we excluded international reserves from the indicators for South Africa in crises 2 and 3 and the real interest rate for Brazil in crisis 2.

5. The approach taken by Kaminsky and Reinhart (1998) is conceptually similar although formally different. They estimate the incidence of crises as a function of fundamentals and the number of crises in alternative clusters of countries. This is equivalent to having matrixes with ones in particular entries.

To test for the presence of contagion, we check whether β_2 in equation 1 is significantly different from zero. To compare the strength of contagion across different weighting matrixes (of the second type), we rescale them such that each row adds up to one. Thus, β_2 shows the impact of a particular weighted average of crisis indicators elsewhere in the crisis indicators of the average (not ground-zero) country. Then different weighting matrixes allow us to identify the most important transmission channels.

Macroeconomic Fundamentals

The vector $X_{t,j}$ of initial macroeconomic conditions includes country-specific characteristics that may explain the extent of the crises in each country. Specifically, we consider a set of variables that are typically related to currency attacks and balance of payments crises according to standard models (first, second, and later generations) and the existing empirical evidence.[6] The list of variables is the following:

1. *Credit boom 1.* Total credit to the private sector (as a percentage of gross domestic product, GDP) in excess of the long-run trend of the ratio credit/ GDP calculated using a Hodrik-Prescott filter (see Gourinchas, Landerretche, and Valdés 2001). We consider 1981, 1994, and 1996 as the initial conditions for crises 1, 2, and 3, respectively.
2. *Credit boom 2.* Total credit (as a percentage of GDP) in excess of the long-run trend of the ratio credit/ GDP, for the same years as for credit boom 1.
3. RER *overvaluation.* Twelve-month average of *RER* misalignment prior to each crisis calculated using as equilibrium *RER* an HP filter with information up to the month before each crisis (therefore the filter is one-sided).
4. *Fiscal balance/GDP.* Fiscal balance as a percentage of GDP, for the same years as for credit booms.
5. *Current account/GDP.* Current account balance as a percentage of GDP, for the same years as for credit booms.
6. *GDP growth.* GDP annual growth rate, for the same years as for credit booms.
7. *Debt/GDP.* Debt to GDP ratio. For OECD countries, we estimate the stock of debt by adding up current account deficits since 1950. This is for the same years as for credit booms.
8. *Inflation.* Consumer price index 12-month inflation measured in the month before each crisis (measured as p/(1 + p), where p is the rate of inflation).

Before analyzing the presence of contagion, it is interesting to evaluate whether these macroeconomic fundamentals matter in explaining which countries suffer stronger crises (or a crisis at all) during an international crisis. Sachs, Tornell, and Velasco (1996) address this issue, although they focus only on the Tequila crisis. Their main result is that excess credit creation and *RER* misalignment are

6. See Eichengreen, Rose, and Wyplosz (1997); Kaminsky, Lizondo, and Reinhart (1998); and the comprehensive study by Berg and Pattillo (1998) for details.

the most important variables in explaining the extent of crises across countries. They do not find any relevant role for the current account deficit. Berg and Pattillo (1998) find similar results using several alternative methodologies. They find that the most important indicators of vulnerabilities are the rate of growth of domestic credit, a measure of real exchange rate overvaluation, and the ratio of reserves to the M2 money supply. They find that the current account deficit, the budget deficit, and the composition of external liabilities are good predictors of external fragilities only in some cases (estimations).

Table 1 presents the results of estimating equation 1 without contagion effects. In the equation, we include ground-zero countries, so we estimate a standard crisis-prediction equation. In our estimations, the current account balance appears as a highly significant explanatory variable in *PI-3*, *PI-12*, and *RER* (the "objective" indicators). Credit boom (private credit), *RER* overvaluation, fiscal balance, and GDP growth are significant in some of the crisis indicators. In the case of the *RER* depreciation indicator, it is interesting to note that the signs of the current account balance and the fiscal balance are opposite. This indicates that an increase in the current account increases the real depreciation 12 months later, but the converse occurs with the fiscal balance. The interpretation is not straightforward. By accounting, we can decompose the current account deficit into private and public components, the latter being the budget balance. An in-

TABLE 1. Crisis Indicators and Initial Conditions

	Crisis indicator			
Variable	Change in *PI-3*[a]	Change in *PI-12*[b]	Change in credit rating	Change in *RER*[c]
Constant	−0.08	−4.62	−1.92	−1.25
	(−0.04)	(−2.94)	(−2.41)	(−0.68)
Credit boom	−30.82	—	−15.92	−44.35
	(−1.64)		(−1.75)	(−2.27)
RER overvaluation	−0.24	−0.45	—	—
	(−1.43)	(−2.63)		
Fiscal budget/GDP	—	—	—	−0.77
				(−2.14)
Current account/GDP	0.44	0.67	—	1.04
	(1.70)	(2.54)		(3.49)
GDP growth	—	1.50	—	—
		(3.22)		
R^2	0.17	0.31	0.05	0.29
F-statistic p-value	0.02	0.00	0.09	0.00
Observations	60	60	60	60

Note: Data are for 20 countries for three crisis periods. See table A-1 for countries and text for crisis periods. Values are from OLS regressions with constants (not reported). White's robust t-tests are in parentheses. We report variables with at least 80 percent significance.

[a]Foreign exchange market pressure index three months after the crisis.

[b]Foreign exchange market pressure index 12 months after the crisis.

[c]Level of the real exchange rate 12 months after the crisis.

Source: Authors' calculations.

crease in the budget deficit would raise the current account deficit, deteriorating the RER indicator, but there is a direct effect partially offsetting the current account effect.

An interesting result is that, other than credit boom, macro-variables do not explain changes in credit rating. Credit rating is a "subjective" crisis indicator because it is based on the assessment of vulnerabilities assigned by the market.

Neither the debt/GDP ratio nor inflation has significant effects in explaining any of the crisis indicators. As shown by the R^2 statistics, the macroeconomic fundamentals we consider have a limited capability for explaining the cross-country experience during crisis periods, a result consistent with the already large literature on crisis forecasting.

II. Contagion and Transmission Channels

This section investigates the presence of contagion in the three crises we study and analyzes the likelihood of alternative transmission channels. It discusses the construction of alternative weighting matrixes and presents some empirical results.

Weighting Matrixes

There are several potential channels for the propagation of contagion. The most important are direct trade links, trade competition in third markets, macroeconomic similarities, and financial links. Eichengreen, Rose, and Wyplosz (1997) and Glick and Rose (1998) find evidence that trade links are the most important channel of propagation. Kaminsky and Reinhart (1998) also find strong evidence of regional contagion. They conclude that this pattern could be associated with trade links as well as with financial links. A key problem is that the two are correlated. An additional problem is that measures to control for financial links are limited.

Controlling for the average shock elsewhere is a form of controlling for the international environment. In addition, we may capture the channels through which interdependence or contagion occurs by weighting the shocks elsewhere by some characteristics of the relationship among countries. Thus, different weighting matrixes $M_{i,k,j}$ allow us to investigate the importance of alternative transmission channels of contagion (from country i to country k). We consider the following matrixes:

1. Equal weights for all countries k, allowing us to control for differences across crises.
2. Direct trade links measured by the ratio of bilateral trade between countries i and k to total trade of country i. This set of weights is motivated by trade-based contagion theories, such as competitive devaluation.
3. Trade competition in third markets measured through a similarity index of the trade pattern based on the relative importance in total exports of

six sectors (agriculture, food, fuel, ores, high-tech manufacturing, and low-tech manufacturing). This matrix has the same motivation as in point 2.

4. Neighborhood (regional) dummies for Latin American, Asian, and industrial countries (see appendix table A-1 for details). This matrix is motivated by the presumption that contagion is regional (explained primarily by financial links after controlling for trade links).
5. An overall macroeconomic similarity index that combines RER misalignment, current account balance, credit boom, fiscal balance, and GDP growth. Macroeconomic similarities may explain contagion if, for instance, investors learn and update their priors during a crisis (that is, there is a "wake-up call" during crisis).
6. Specific macroeconomic similarity indexes, including external similarity (encompassing RER and current account), credit boom, and GDP growth.
7. All of the above measures, but with respect to only neighboring countries. This allows us to evaluate the alternative contagion channels at the regional level.

Both trade-pattern similarity, because of data availability, and neighbor dummy matrixes, by definition, are constant across crises. The rest of the matrixes are crisis-specific. All the matrixes are symmetric, except the one with direct trade links. The reason for the lack of symmetry of the trade-link matrix is that trade is measured with respect to total trade of the country; thus, bilateral trade is symmetric, not its importance with respect to each country.

To construct a similarity index between countries i and k when considering a single variable (for example, GDP growth or credit boom), we calculate:[7]

$$\theta_{i,k,j} = \exp(-|x_{i,j} - x_{k,j}|), \tag{2}$$

where x_i is the standardized variable under analysis in country i. The standardization is based on cross-country, crisis-specific observations.[8]

When constructing similarity indexes that combine multiple variables (for example, trade pattern, external conditions, and overall macroeconomic similarity), we calculate:

$$\theta_{i,k,j} = \exp(-\sum_s |x_{s,i,j} - x_{s,k,j}|), \tag{3}$$

where s indexes the different variables entering the index and $x_{s,i,j}$ is the standardized variable s in country i and crisis j.

To facilitate comparability across different matrixes, we rescale the $\theta_{i,k,j}$'s so that maximum similarity takes the value 1 and minimum similarity takes the value 0. Thus, we calculate the weight $M_{i,k,j}$ as follows:

7. The procedure for constructing similarity indexes is somewhat ad hoc because it introduces some nonlinear transformations in the data; however, it allows us to reduce the effect of outliers.

8. By standardized variable, we refer to a variable in a given crisis minus its mean divided by its standard deviation.

$$M_{i,k,j} = \frac{\theta_{i,k,j} - \min(\theta_{i',k',j})}{\max(\theta_{i',k',j}) - \min(\theta_{i',k',j})} \tag{4}$$

where i', k', and j represent all possible country combinations in crisis j. Furthermore, for a straightforward interpretation of the results, we rescale $M_{i,k,j}$ again so that $\Sigma_i M_{i,k,j} = 1$. Thus, β_2 reflects the impact of a weighted average of what is happening elsewhere on the average country.

Empirical Results

Tables 2 to 5 present the estimation of equation 1 using *PI-3*, *PI-12*, *RER*, and *CR*, respectively, and with alternative weighting matrixes for each crisis indicator. The variable "contagion index" corresponds to β_2, while "equal weight" corresponds to β_3. All regressions include a constant and dummies for the ground-zero countries (not reported).

The results for the *PI-3* indicator show that contagion is strongly and almost exclusively driven by neighborhood and direct trade effects. None of the "wider" matrixes (those considering not only neighbors) yields a significant coefficient that could indicate the presence contagion. Indeed, when constraining weighting matrixes to neighboring countries, most of the results are significant. The point estimate of direct trade links is smaller than that of the neighbor dummies, and, because we are constraining weights to be one, we can conclude that the neighbor effect is quantitatively stronger than that of direct trade. This probably reflects the close trade links that exist between neighbors rather than a proper propagation channel. In fact, when we consider direct trade with neighboring countries only, the estimate is highly significant, but the point estimate is still smaller than what the neighbor dummy matrix yields. Interestingly, neither macroeconomic similarities nor the common shock proxy plays any role in explaining the cross-country propagation of contagion at this three-month horizon.

None of the parameters corresponding to the variables measuring macroeconomic initial conditions, except for credit boom, changes in any important way when we incorporate the contagion index. In fact, credit boom ceases to be significant in all specifications. Consequently, once the effects of interdependence across crises are included, the R^2 increases from 0.17 in table 1 to values around 0.5. This reveals the importance that contagion and transmission of crisis across countries have on the vulnerability to external crisis.[9]

The results for *PI-12* show a different picture (table 3). For this indicator, we observe that a real exchange rate overvaluation, a current account deficit, and low growth increase the (absolute) value of the crisis indicator, that is, increase the incidence of crisis. After controlling for the equal-weight matrix, the R^2s increase with respect to the value reported in table 1, but the marginal explanatory power of this variable is not as large as that of the three-month exchange

9. It is also worth mentioning that, aside from the *PI-12* indicator, results do not change if we exclude the $M_{1,1}\Delta CI_{k,t,j}$ term in the regressions.

TABLE 2. Three-Month Change in Foreign Exchange Market Pressure Index and Total Contagion

	Weighting matrix								
Variable	Direct trade	Trade pattern	Neighbor dummy	Macro similarity	External similarity	Credit similarity	Growth similarity	Trade with neighbors	Trade pattern with neighbors
Credit boom	–5.32	–13.22	0.68	–13.07	–15.45	–21.92	–12.55	–2.09	–3.34
	(–0.33)	(–0.78)	(0.05)	(–0.68)	(–0.91)	(–0.96)	(–0.74)	(–0.14)	(–0.21)
RER overvaluation[a]	–0.24	–0.26	–0.21	–0.25	–0.25	–0.26	–0.25	–0.21	–0.21
	(–1.73)	(–1.78)	(–1.75)	(–1.71)	(–1.74)	(–1.79)	(–1.72)	(–1.63)	(–1.61)
Current account/GDP	0.45	0.24	0.41	0.28	0.27	0.26	0.27	0.49	0.40
	(1.98)	(1.00)	(2.05)	(1.18)	(1.18)	(1.15)	(1.20)	(2.30)	(1.87)
Contagion index	0.63	–0.54	0.71	–0.18	–3.40	–1.24	–0.73	0.61	0.47
	(2.37)	(–0.52)	(4.29)	(–0.12)	(–1.14)	(–0.65)	(–0.38)	(3.50)	(2.91)
Equal weight	–0.06	1.05	–0.08	0.71	3.80	1.91	1.21	–0.02	0.10
	(–0.14)	(1.01)	(–0.26)	(0.45)	(1.31)	(0.89)	(0.67)	(–0.07)	(0.31)
R^2	0.51	0.46	0.60	0.46	0.47	0.46	0.46	0.56	0.53
F-statistic p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Observations	60	60	60	60	60	60	60	60	60

Note: Data are for 20 countries for three crisis periods. See table A-1 for countries and text for crisis periods. Values are from OLS regressions with constants and dummy variables in the three ground-zero countries (not reported). White's robust t-tests are in parentheses. External similarity combines current account and RER overvaluation similarity.

[a]See text for definition.

Source: Authors' calculations.

TABLE 3. Twelve-Month Change in Foreign Exchange Market Pressure Index and Total Contagion

	Weighting matrix								
Variable	Direct trade	Trade pattern	Neighbor dummy	Macro similarity	External similarity	Credit similarity	Growth similarity	External similarity of neighbors	Growth of neighbors
RER overvaluation[a]	−0.46	−0.42	−0.45	−0.51	−0.44	−0.46	−0.46	−0.52	−0.43
	(−2.66)	(−2.60)	(−2.98)	(−2.96)	(−2.32)	(−2.68)	(−2.64)	(−3.17)	(−2.79)
Current account/GDP	0.45	0.48	0.39	0.53	0.47	0.48	0.47	0.39	0.41
	(1.56)	(1.87)	(1.59)	(1.94)	(1.73)	(1.75)	(1.68)	(1.50)	(1.63)
GDP growth	1.31	1.18	1.53	1.49	1.26	1.26	1.12	1.36	1.84
	(2.62)	(2.59)	(3.51)	(2.97)	(2.63)	(2.64)	(1.46)	(2.98)	(3.94)
Contagion index	−0.13	−2.80	−1.72	−3.03	0.15	−0.49	0.42	−1.08	−1.41
	(−0.34)	(−2.46)	(−3.50)	(−1.33)	(0.06)	(−0.33)	(0.22)	(−2.42)	(−3.37)
Equal weight	0.72	3.31	2.28	3.51	0.45	1.10	0.18	1.60	2.00
	(1.63)	(2.93)	(4.26)	(1.59)	(0.18)	(0.72)	(0.10)	(3.31)	(4.18)
R^2	0.40	0.46	0.52	0.42	0.40	0.40	0.40	0.46	0.50
F-statistic p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Observations	60	60	60	60	60	60	60	60	60

Note: Data are for 20 countries for three crisis periods. See table A-1 for countries and text for crisis periods. Values are from OLS regressions with constants and dummy variables in the three ground-zero countries (not reported). White's robust t-tests are in parentheses. External similarity combines current account and RER overvaluation similarity.

[a]See text for definition.

Source: Authors' calculations.

market pressures indicator. We find that for this indicator, co-movement is almost exclusively driven by the common shock (proxied by the equal-weight matrix, that is, crisis elsewhere). Transmission through trade, neighbor effects, and similarities do not appear to play an important additional role. In fact, none of the weighting matrixes yields significantly positive parameters. If we do not control for the equal-weight matrix, the results change dramatically, with several weighting matrixes having significantly positive results. However, this follows from the fact that the equal weight and other matrixes are collinear across crises. In what follows, we no longer consider *PI-12* in the analysis and conclude that there is no particular form of contagion in this indicator beyond the existence of common shocks (although there is a high degree of co-movement across countries).[10]

In the case of the indicator based on 12-month RER depreciation (table 4), we find that contagion indexes are significantly positive when we consider direct trade links, neighbors, and growth similarity. The strong negative sign for trade pattern similarity indicates that there is evidence against third-market competition being an important transmission mechanism of crises.

Conventional wisdom indicates that when a country has a currency crisis, a real depreciation will hurt competitors in those markets, leading to competitive devaluations. However, because a crisis in a country is usually coupled with an output collapse, it may create opportunities for the country's main competitors. This may be what is happening with the reverse sign we find, at least at the one-year horizon. It might also be that trade pattern similarity is not appropriately measuring third-market competition, and perhaps third-market competition could be better proxied by some regional effect. We still find that initial conditions measured by the current account deficit and budget deficit help to explain 12-month RER depreciation. Credit boom is the only initial macroeconomic variable that looses significance in the RER equation when we include contagion.

Finally, in the case of change in credit rating (table 5), we find that the direct trade links, neighbors, overall macro similarity, and growth similarity matrixes yield significant contagion coefficients. When considering only similarities with neighboring countries, we find that both trade and external macroeconomic similarity appear to be very important channels of contagion. As in the previous case, initial conditions measured by credit boom looses significance when we include contagion. With the CR index, we find no initial condition to be significant when we include contagion.

The evidence presented so far is not able to discriminate completely among (statistically significant) competing weighting matrixes. Following Eichengreen, Rose, and Wyplosz (1997), table 6 presents the results of estimating equation 1

10. We look again at *PI-12* only when examining contagion from ground-zero countries because the specification and the implication of the results are different. In addition, in the remaining results, we exclude the equal-weight matrix from the analysis because it is not significant for indicators other than *PI-12*.

TABLE 4. Real Exchange Rate Depreciation and Total Contagion

	Weighting matrix								
Variable	Direct trade	Trade pattern	Neighbor dummy	Macro similarity	External similarity	Credit similarity	Growth similarity	External similarity of neighbors	Growth of neighbors
Credit boom	−14.77	−23.81	−9.55	−10.78	−20.24	−13.95	−11.33	−7.25	−9.34
	(−0.77)	(−1.23)	(−0.49)	(−0.51)	(−1.02)	(−0.60)	(−0.61)	(−0.37)	(−0.48)
Fiscal budget/GDP	−0.65	−0.96	−0.68	−0.77	−0.90	−0.86	−0.52	−0.63	−0.68
	(−1.95)	(−2.90)	(−2.06)	(−2.26)	(−2.60)	(−2.54)	(−1.57)	(−1.92)	(−2.05)
Current account/GDP	1.08	0.94	1.09	1.07	1.03	1.02	1.14	1.08	1.09
	(4.02)	(3.47)	(4.12)	(3.84)	(3.72)	(3.69)	(4.34)	(4.11)	(4.07)
Contagion index	0.65	−2.39	0.60	1.52	−1.57	0.67	4.14	0.64	0.56
	(2.17)	(−1.99)	(2.36)	(1.14)	(−0.68)	(0.48)	(2.81)	(2.53)	(2.13)
Equal weight	0.10	3.09	0.26	−0.76	2.37	0.06	−3.08	0.21	0.33
	(0.26)	(2.58)	(0.72)	(−0.55)	(1.01)	(0.04)	(−2.21)	(0.60)	(0.95)
R^2	0.49	0.48	0.49	0.45	0.44	0.44	0.51	0.50	0.48
F-statistic p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Observations	60	60	60	60	60	60	60	60	60

Note: Data are for 20 countries for three crisis periods. See table A-1 for countries and text for crisis periods. Values are from OLS regressions with constants and dummy variables in the three ground-zero countries (not reported). White's robust t-tests are in parentheses. External similarity combines current account and RER overvaluation similarity.

Source: Authors' calculations.

TABLE 5. Change in Credit Rating and Total Contagion

Variable	Weighting matrix								
	Direct trade	Trade pattern	Neighbor dummy	Macro similarity	External similarity	Credit similarity	Growth similarity	Trade with neighbors	External similarity of neighbors
Credit boom	–0.95	–8.21	1.28	0.41	–7.10	–24.71	–3.80	1.58	3.63
	(–0.12)	(–1.06)	(0.19)	(0.05)	(–0.86)	(–2.41)	(–0.49)	(0.24)	(0.58)
Contagion index	0.75	–2.09	0.74	2.15	–0.38	–6.63	2.33	0.82	0.83
	(2.70)	(–2.01)	(5.34)	(1.82)	(–0.20)	(–2.59)	(2.16)	(5.11)	(6.10)
Equal weight	0.10	2.59	0.04	–1.23	0.99	7.46	–1.51	0.09	0.02
	(0.33)	(2.59)	(0.16)	(–1.16)	(0.58)	(2.82)	(–1.47)	(0.39)	(0.10)
R^2	0.48	0.45	0.62	0.45	0.41	0.48	0.46	0.61	0.65
F-statistic p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Observations	60	60	60	60	60	60	60	60	60

Note: Data are for 20 countries for three crisis periods. See table A-1 for countries and text for crisis periods. Values are from OLS regressions with constants and dummy variables in the three ground-zero countries (not reported). White's robust t-tests are in parentheses. External similarity combines current account and RER overvaluation similarity.

Source: Authors' calculations.

TABLE 6. Contagion and Competing Weighting Matrixes

	Pressure indicator								
	Change in PI-3[a]		Change in RER[b]			Change in credit rating			
Variable	(1)	(2)	(1)	(2)	(3)	(1)	(2)	(3)	(4)
Credit boom	0.00	0.01	−0.13	−0.10	−0.08	0.01	0.02	0.07	0.04
	(0.02)	(0.05)	(−0.69)	(−0.54)	(−0.44)	(0.12)	(0.26)	(1.2)	(0.59)
RER overvaluation[b]	−0.21	−0.22	—	—	—	—	—	—	—
	(−1.74)	(−1.78)							
Fiscal budget/GDP	—	—	−0.68	−0.62	−0.68	—	—	—	—
			(−2.07)	(−1.92)	(−2.10)				
Current account/GDP	0.37	0.38	1.08	1.10	1.10	—	—	—	—
	(1.95)	(1.92)	(4.06)	(4.19)	(4.17)				
Direct trade matrix	−0.22	—	0.50	0.37	—	−0.12	—	—	—
	(−0.74)		(1.47)	(1.12)		(−0.42)			
Neighbor dummy matrix	0.82	0.76	—	0.41	0.50	0.81	0.71	−2.00	—
	(3.48)	(2.22)		(1.26)	(1.91)	(4.36)	(5.09)	(−2.62)	
Macro similarity matrix	—	—	—	—	—	—	0.13	—	—
							(0.48)		
Growth similarity matrix	—	—	0.40	—	0.47	—	—	—	—
			(0.89)		(1.22)				
Trade with neighbors	—	−0.08	—	—	—	—	—	—	1.04
		(−0.24)							(2.83)
External similarity with neighbors matrix	—	—	—	—	—	—	—	2.90	−0.25
								(3.64)	(−0.61)
R^2	0.60	0.60	0.49	0.50	0.50	0.58	0.62	0.69	0.66
F-statistic p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Observations	60	60	60	60	60	60	60	60	60

Note: Data are for 20 countries for three crisis periods. See table A-1 for countries and text for crisis periods. Values are from OLS regressions with constants and dummy variables in the three ground-zero countries (not reported). White's robust t-tests are in parentheses. External similarity combines current account and RER overvaluation similarity.

[a]Foreign exchange market pressure index three months after the crisis.

[b]Level of the real exchange rate 12 months after the crisis.

Source: Authors' calculations.

simultaneously including competing relevant contagion indexes. We consider some of the matrixes that appeared as more relevant in tables 2–5 in pairs, using the same initial macroeconomic conditions as before.

The results show that in the cases of indicators based on PI-3 and country CR, the identification is straightforward. In both cases, the neighborhood effect appears as the most relevant propagation mechanism for contagion. In the second case, we also observe that external similarities with respect to neighbors appears to be a strong mechanism (which is a particular form of a neighborhood effect). Trade links no longer appear important in these two cases when we control for the effect of neighbors. Although trade links and neighbor effects are highly correlated, our results suggest that the prime candidate for contagion is not trade, as documented in other papers, but geographical proximity.[11] The results are less clear-cut in the case of the indicators based on RER. Because of strong collinearity, some times we observe that a pair of matrixes is highly significant when considered individually, but is no longer significant (individually) when considered together. Despite this issue, it is possible to exclude some explanations and rank others informally according to point estimates. Direct trade links and neighbors appear as the two most relevant matrixes.[12]

Contagion from Ground-Zero Countries

An alternative way of defining contagion is to limit it to propagating from ground-zero countries only. In this case, we try to explain the cross-country variation of our crisis indicators using different weights of ground zero for each country. This definition of contagion is obviously more restrictive than the previous approach. Moreover, it is potentially misleading if the ground-zero country is not correctly identified. However, this exercise is useful for testing the robustness of our results.

Because the temporal evolution of the ground-zero country can be very different from what actually happened in other countries, we modify our strategy slightly. In particular, we analyze whether a weighted change in *PI-3* at ground zero is able to explain changes in *PI-12*, RER, and CR. The weighting matrixes are similar to those we used in the previous subsection, although we no longer have the straightforward intuition for the estimated parameter we had before (a weighted average of what is happening elsewhere). Therefore, we use standardized parameters.

Table 7 presents the results for the cases in which we find statistically significant contagion. It shows that with the *PI-12* indicator, contagion marginally arises only when we consider the equal-weight matrix. This result is proof of comovement, perhaps caused by a large shock, which is different across crises, but it is not necessarily evidence of contagion. With the indicator based on the RER, direct trade ties between countries and the ground-zero country appear to generate contagion. Finally, changes in credit rating can be explained for countries that are neighbors of the ground-zero country (especially if they have similar initial external macroeconomic conditions) or have direct trade links with it.

11. We cannot avoid making references to the case of Chile, which suffered contagion from Asia due to high trade links, but is also dependent on movements in Latin America, a region with weak trade links. Chile's trade with Argentina and Brazil, its main trade partners in the region, is well below 10 percent.

12. One can further analyze this issue of collinearity by estimating a model of the following form:

$$\Delta CI_{i,t,j} = \beta_0 + \beta_1 X_{i,j} + \beta_2 \times (\gamma \sum_{k \neq i} M_{i,k,j} \Delta CI_{k,t,j} + (1-\gamma) \sum_{k \neq i} M'_{i,k,j} \Delta CI_{k,t,j}) + \epsilon_{i,t,j}$$

where γ measures the relative importance of $M_{i,k,j}$ vis-à-vis $M'_{i,k,j}$. The results for RER (not reported) show a significant β_2 but very imprecise estimates of γ, showing that any combination of the two matrixes would be valid.

TABLE 7. Contagion from a Ground-Zero Country

Variable	Pressure indicator and weighting matrix					
	Change in PI-12/equal weights[a]	Change in RER/direct trade[b]	Change in credit rating/ direct trade	Change in credit rating/ neighbor dummy	Change in credit rating/ external similarity of neighbors	Change in credit rating/ trade with neighbors
Credit boom	—	−0.23	−0.06	−0.04	−0.02	−0.05
		(−1.12)	(−0.80)	(−0.62)	(−0.46)	(−0.69)
RER overvaluation[b]	−0.48	—	—	—	—	—
	(−2.79)					
Fiscal budget/GDP	—	−0.57	—	—	—	—
		(−1.58)				
Current account/GDP	0.50	1.04	—	—	—	—
	(1.77)	(3.67)				
GDP growth	1.44	—	—	—	—	—
	(3.03)					
Contagion index	0.24	1.78	0.27	0.43	0.45	0.29
	(1.69)	(2.04)	(2.53)	(4.38)	(4.72)	(2.76)
R^2	0.37	0.41	0.41	0.52	0.54	0.43
F-statistic p-value	0.00	0.00	0.00	0.00	0.00	0.00
Observations	60	60	60	60	60	60

Note: Data are for 20 countries for three crisis periods. See table A-1 for countries and text for crisis periods. Values are OLS regressions with constants and dummy variables in the three ground-zero countries (not reported). White's robust t-tests are in parentheses. Contagion index corresponds to the standardized parameter of a weighted average of change in PI-3 according to a particular matrix M. External similarity combines current account and RER overvaluation similarity.

[a]Foreign exchange market pressure index three months after the crisis.

[b]Level of the real exchange rate 12 months after the crisis.

Source: Authors' calculations.

III. Policies to Curb Contagion

One key policy question is how countries can curb (or even stop) contagion. A leading prescription is to limit financial integration. Other policy prescriptions to limit the extent of contagion are exchange rate flexibility and avoiding short-term debt. The issue of contagion and alternative policies is an empirical one. This section evaluates the usefulness of these three policy measures in curbing contagion.

Capital Controls and Contagion

Capital controls could curb contagion if financial links are an important propagation channel. However, the usefulness of limiting financial integration is less clear if contagion arises due to trade links, or if initial similarity in macroeconomic conditions and crises are the consequence of real shocks. Nevertheless, it could be argued that capital controls might help an orderly adjustment, avoiding typical problems that an unregulated financial sector often produces, such as overshooting the exchange rate. Of course, capital controls have costs in tranquil times because the country does not take full advantage of capital movements. However, defenders of capital controls point to contagion as one of the reasons for having capital controls as a preventive measure.

Edwards (1999) evaluates whether capital controls in Chile were a useful device for avoiding contagion. He measures contagion as the correlation between domestic and Asian interest rates (specifically, interest rates in Hong Kong), controlling for domestic devaluation and exchange rates in the United States. He concludes that controls on capital inflows may have been able to protect Chile from relatively small shocks, but were not able to prevent contagion stemming from large external shocks.

It should be mentioned that the objective of capital control measures goes beyond avoiding contagion. Among other objectives, capital controls have been used to avoid excess real exchange rate appreciation, to curb capital inflows, and to modify the foreign debt term structure.[13]

To evaluate whether financial integration facilitates contagion, we use a standard capital control index and analyze whether contagion is weaker in countries with a higher index. In particular, we estimate models of the following form:

$$\Delta CI_{i,t,j} = \beta_0 + \beta_1 X_{i,j} + [\beta_2 + \beta_3 CC_{i,j}] \sum_{k \neq i} M_{i,k,j} \Delta CI_{k,t,j} + \epsilon_{i,t,j}, \tag{5}$$

where $CC_{i,j}$ is a capital control index of country i during crisis j. If capital controls were effective in curbing contagion, the estimation should yield a negative and significant β_3.

To construct the capital control index, we use the standard dummy variables that appear in IMF (various years). For restrictions on payments on capital transactions and the surrender requirement of export proceeds, we assign values of

13. See De Gregorio, Edwards, and Valdés (2000) for an evaluation of the Chilean experience.

0, 1, or 2, depending on whether neither, one, or both of the restrictions apply. We consider the status as of December in 1981, 1994, and 1996 for the corresponding crises.

Table 8 presents the results of the estimation of equation 5 for our three crisis indicators that show contagion and for the same weighting matrixes used in last section. The results show that capital controls do not have any relevant effect in limiting contagion. Indeed, the associated parameter is generally not significantly different from zero. It has to be noted, however, that we use a broad definition of capital controls, and the most commonly used and specific forms of controls or regulations cannot be captured with these 0, 1, 2 indicators. However, the results indicate that countries that had more pervasive forms of control did not avoid contagion more than countries with looser controls.

Exchange Rate Flexibility and Contagion

Exchange rate flexibility is expected to reduce contagion by avoiding some of the overvaluation episodes to begin with and limiting the scope of speculation. To evaluate the effect of exchange rate flexibility on contagion, we use the same approach as with capital controls. In particular, we estimate an equation similar to equation 5, but with an indicator of exchange rate flexibility for country i in crisis j instead of $cc_{i,j}$. We use a 0, 1, 2 indicator (2 is maximum flexibility) based on data gathered by Goldfajn and Valdés (1999). The data were constructed using IMF (various years). That report groups exchange rate regimes into three categories: fixed (including narrow bands), flexible, and floating.

Table 9 presents the results. They show that flexibility has a significant effect in limiting contagion only when we measure contagion using changes in credit ratings. Point estimates show a large effect: Moving from a fixed exchange rate regime to a floating one reduces contagion by two-thirds. This result is robust to alternative weighting matrixes. It is interesting because it indicates that the market evaluates better and is less vulnerable to economies with flexible exchange rate regimes.

When measuring contagion with real depreciation, we find that flexibility *increases* contagion, although this result is marginally significant under only two weighting matrixes. This latter result is not surprising because the exchange rate is the variable that adjusts when external shocks hit the economy. Moreover, part of the adjustment may be an overshooting of the real exchange rate. We do not find significant effects of flexibility in the case of *PI-3*.

Overall, we can conclude only for the *CR* indicator that having a flexible exchange rate may reduce contagion.

Debt Maturity Structure and Contagion

Having debt maturity tilted toward the long run would limit the scope of financial runs against a particular country. To evaluate whether the debt maturity structure has any impact on the extent of contagion, we run an equation similar to equation 5, but with the ratio of short-term debt to total debt for country i in

TABLE 8. Capital Controls and Contagion

Variable	Pressure indicator and weighting matrix								
	Change in PI-3/direct trade[a]	Change in PI-3/ neighbors[a]	Change in PI-3/trade with neighbors[a]	Change in RER/direct trade[b]	Change in RER/ neighbors[b]	Change in RER/ growth similarity[b]	Change in credit rating/ direct trade	Change in credit rating/ trade with neighbors	Change in credit rating/ external similarity of neighbors
Credit boom	−0.04	0.01	−0.02	−0.16	−0.13	−0.17	−0.01	0.02	0.04
	(−0.23)	(0.07)	(−0.12)	(−0.86)	(−0.70)	(−0.84)	(−0.11)	(0.26)	(0.58)
RER overvaluation[b]	−0.26	−0.23	−0.22	—	—	—	—	—	—
	(−1.86)	(−1.76)	(−1.65)						
Fiscal budget/GDP	—	—	—	−0.57	−0.56	−0.76	—	—	—
				(−1.69)	(−1.75)	(−2.26)			
Current account/GDP	0.38	0.37	0.46	1.09	1.18	1.05	—	—	—
	(1.67)	(1.79)	(2.15)	(4.10)	(4.42)	(3.87)			
Contagion index	0.48	0.64	0.56	0.82	0.94	1.00	0.84	1.12	0.92
	(1.65)	(2.77)	(2.36)	(3.19)	(3.71)	(2.71)	(2.64)	(4.95)	(4.69)
Contagion × capital controls	0.14	0.04	0.04	−0.15	−0.35	−0.06	−0.03	−0.23	−0.08
	(0.64)	(0.23)	(0.23)	(−0.69)	(−1.45)	(−0.20)	(−0.15)	(−1.53)	(−0.56)
R^2	0.51	0.60	0.56	0.49	0.51	0.47	0.48	0.62	0.66
F-statistic p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Observations	60	60	60	60	60	60	60	60	60

Note: Data are for 20 countries for three crisis periods. See table A-1 for countries and text for crisis periods. Values are from OLS regressions with constants and dummy variables in the three ground-zero countries (not reported). White's robust t-tests are in parentheses. External similarity combines current account and RER overvaluation similarity.

[a]Foreign exchange market pressure index three months after the crisis.

[b]Level of the real exchange rate 12 months after the crisis.

Source: Authors' calculations.

TABLE 9. Exchange Rate Flexibility and Contagion

	Pressure indicator and weighting matrix								
Variable	Change in PI-3/direct trade[a]	Change in PI-3/ neighbors[a]	Change in PI-3/trade with neighbors[a]	Change in RER/direct trade[b]	Change in RER/ neighbors[b]	Change in RER/ growth similarity[b]	Change in credit rating/ direct trade	Change in credit rating/ trade with neighbors	Change in credit rating/ external similarity of neighbors
Credit boom	−0.05	0.02	−0.01	−0.18	−0.11	−0.23	−0.21	−0.05	−0.02
	(−0.30)	(0.14)	(−0.07)	(−0.92)	(−0.60)	(−1.19)	(−0.28)	(−0.67)	(−0.26)
RER overvaluation[b]	−0.22	−0.22	−0.23	—	—	—	—	—	—
	(−1.58)	(−1.79)	(−1.75)						
Fiscal budget/GDP	—	—	—	−0.60	−0.62	−0.72	—	—	—
				(−1.85)	(−1.96)	(−2.27)			
Current account/GDP	0.43	0.40	0.50	1.04	1.10	1.00	—	—	—
	(2.06)	(2.09)	(2.50)	(3.90)	(4.23)	(3.78)			
Contagion index	0.74	0.62	0.44	0.40	0.18	0.35	1.54	1.36	1.23
	(1.97)	(2.42)	(1.62)	(1.03)	(0.50)	(0.79)	(4.50)	(4.92)	(5.87)
Contagion × exchange rate flexibility	−0.12	0.07	0.17	0.23	0.46	0.52	−0.68	−0.54	−0.43
	(−0.42)	(0.31)	(0.71)	(0.94)	(1.74)	(1.81)	(−2.67)	(−2.12)	(−2.27)
R^2	0.51	0.60	0.57	0.49	0.52	0.50	0.54	0.64	0.69
F-statistic p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Observations	60	60	60	60	60	60	60	60	60

Note: Data are for 20 countries for three crisis periods. See table A-1 for countries and text for crisis periods. Values are from OLS regressions with constants and dummy variables in the three ground-zero countries (not reported). White's robust t-tests are in parentheses. External similarity combines current account and RER overvaluation similarity.

[a]Foreign exchange market pressure index three months after the crisis.

[b]Level of the real exchange rate 12 months after the crisis.

Source: Authors' calculations.

crisis j instead of $CC_{t,j}$. We use data from the Bank of International Settlements (BIS) (various years) and consider the short term to be less than a year. Two of the countries in our sample (Sweden and Finland) have positive net external assets and report to the BIS from "within," and one country (Singapore) is considered a banking center and thus is highly leveraged. For these countries, we consider a zero in the ratio short debt/total debt and include a special dummy variable in the equation multiplying the contagion index.

Table 10 shows that a tilt toward short-term financing increases contagion when we measure it using changes in credit rating. The effects are economically relevant, highly significant, and robust to alternative weighting matrixes. With 12-month real depreciation and direct trade, there is a marginally significant positive effect.

IV. Concluding Remarks

This article has examined the channels through which crises spread across countries. For this purpose, we examined the behavior of crisis indicators as a function of initial conditions and the average of crisis indicators elsewhere. The latter variable attempts to capture interdependence or co-movements. This relationship could be simply the result of common shocks hitting a number of countries. To understand how these external common shocks and shocks originating in other countries spread to other places, we constructed a weighted average of crisis indicators elsewhere. The weighting schemes attempt to capture different transmission mechanisms. We used the importance of bilateral (also called direct) trade, competition in third markets, regional relationship, and indexes of similarities.

We found that the channel of propagation of crises depends on both indicators and horizons. Three months after a crisis, there are strong neighborhood effects. Rather than trade links and/or macroeconomic similarities, what seems to better explain cross-country correlation is the proximity of countries or regional effects. The same happens when we analyze changes in country credit ratings at longer horizons (6 to 12 months).

Thus the regional weighting scheme is the strongest quantitatively and is statistically the most robust. This implies that crisis spread mainly, but not uniquely, as the Russian crisis in 1998 witnessed, through regions. No wonder the debt crisis was centered in Latin America and the 1997 crisis in Asia. Part of this could be explained by direct trade links, because regions tend to have important trade relationships. But the effect of trade links, although important, cannot account for the whole regional effect. Another candidate for explaining this regional effect is financial links, through cross-border ownership of assets, stock market links, and others. At this stage, we do not have good indicators for constructing weighting matrixes to control for financial links. This is clearly an area that deserves further research.

A question that arises in most of the literature on currency crisis and contagion is whether crises are triggered by bad sentiments or by self-fulfilling prophecies. In

TABLE 10. Composition of Capital Inflows and Contagion

Variable	Pressure indicator and weighting matrix								
	Change in PI-3/direct trade[a]	Change in PI-3/ neighbors[a]	Change in PI-3/trade with neighbors[a]	Change in RER/direct trade[b]	Change in RER/ neighbors[b]	Change in RER/ growth similarity[b]	Change in credit rating/ direct trade	Change in credit rating/ trade with neighbors	Change in credit rating/ external similarity of neighbors
Credit boom	−0.05	−0.01	−0.03	−0.09	−0.07	−0.17	0.04	0.01	0.03
	(−0.28)	(−0.04)	(−0.16)	(−0.45)	(−0.34)	(−0.78)	(0.60)	(0.21)	(0.51)
RER overvaluation	−0.23	−0.22	−0.21	—	—	—	—	—	—
	(−1.72)	(−1.76)	(−1.65)						
Fiscal budget/GDP	—	—	—	−0.60	−0.55	−0.74	—	—	—
				(−1.83)	(−1.69)	(−2.22)			
Current account/GDP	0.47	0.34	0.45	1.08	1.11	1.05	—	—	—
	(2.04)	(1.61)	(2.07)	(3.99)	(4.10)	(3.77)			
Contagion index	0.70	0.53	0.50	0.75	3.86	1.17	0.35	0.35	0.34
	(1.79)	(1.32)	(1.50)	(1.96)	(1.74)	(2.02)	(0.84)	(1.20)	(1.15)
Contagion ×	0.57	−0.58	−0.51	3.60	−2.26	1.73	4.38	3.95	2.47
Short-term debt	(0.27)	(−0.40)	(−0.33)	(1.71)	(−1.68)	(0.79)	(2.91)	(4.14)	(2.32)
R^2	0.51	0.60	0.56	0.51	0.52	0.48	0.57	0.72	0.70
F-statistic p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Observations	60	60	60	60	60	60	60	60	60

Note: Data are for 20 countries for three crisis periods. See table A-1 for countries and text for crisis periods. Values are from OLS regressions with constants and dummy variables in the three ground-zero countries (not reported). White's robust t-tests are in parentheses. External similarity combines current account and RER overvaluation similarity.

Source: Authors' calculations.

TABLE A-1. Country List

Country	Neighborhood code
Argentina	1
Brazil	1
Chile	1
Colombia	1
Ecuador	1
Mexico	1
Peru	1
Venezuela	1
Indonesia	2
Korea	2
Malaysia	2
Philippines	2
Singapore	2
Thailand	2
Sweden	3
Finland	3
Portugal	3
Australia	3
New Zealand	3
South Africa	3

the context of contagion, this implies that a crisis could occur just because of contagion. In this article, we show that, although the crisis indicators are affected by contagion, fundamentals explain a large fraction of the crises. In particular, the current account deficit, exchange rate overvaluation, and credit boom affect our market pressure indicators. Given the sample size, the results change in some specifications and some caveats could be added, but we can conclude that fundamentals matter and it is not just what is going on elsewhere that causes crisis to happen.

At a 12-month horizon, fundamentals matter and both trade links and initial macroeconomic conditions explain which countries suffer stronger contagion. We find that the cross-country variation of a 12-month real exchange rate depreciation depends on growth and external similarities (overvaluation and current account deficit) and direct trade links. At this horizon, neighborhood (regional) effects are still important. Common shocks seem to explain cross-country correlation of a 12-month change in a foreign exchange market pressure index. For the other indicators of crisis we use—the 3-month change in foreign exchange market pressure index, the 12-month real exchange rate depreciation, and the change in the credit rating—we find that co-movements explained by specific forms of contagion are more important. To this end, we conclude that although crises may be triggered by common shocks, transmission across countries depends on regional, trade, and macroeconomic characteristics of the countries.

A policy issue that has been in the middle of the discussion on contagion is the way in which links across countries could be limited during crisis periods. The issue of the optimality of contagion should be addressed first, but at this stage we have taken a practical view in analyzing whether there may be policies that could curb contagion. To this end, we analyze the impact of capital controls, exchange rate flexibility, and debt composition. We find that capital controls do not affect contagion. Exchange rate flexibility and the structure of external debt have effects on some of our crisis indicators, affecting the country credit rating. Exchange rate flexibility also affects the real depreciation after 12 months.

References

Berg, Andrew, and Catherine Pattillo. 1998. "Are Currency Crises Predictable? A Test." IMF Working Paper WP/98/154.

Bank of International Settlements (BIS). Various years. *The Maturity, Sectorial and Nationality Distribution of International Bank Lending*. Basle: BIS.

De Gregorio, José, Sebastian Edwards, and Rodrigo O. Valdés. 2000. "Capital Controls: Do They Work?" *Journal of Development Economics* 63(1):59–83.

Edwards, Sebastian. 1999. "How Effective Are Capital Controls?" *Journal of Economic Perspectives* 14(4):65–84.

Eichengreen, Barry, Andrew K. Rose, and Charles Wyplosz. 1997. "Contagious Currency Crises." Mimeo, University of California, Berkeley, July. Available online at http://haas.berkeley.edu/~arose.

Forbes, Kristin, and Roberto Rigobón. 1999. "Measuring Contagion: Conceptual and Empirical Issues." Mimeo, MIT.

Glick, Reuven, and Andrew K. Rose. 1998. "Contagion and Trade: Why Are Currency Crises Regional?" Mimeo, University of California, Berkeley, August. Available online at http://haas.berkeley.edu/~arose.

Gourinchas, Pierre Olivier, Oscar Landerretche, and Rodrigo O. Valdés. 2001. "Credit Booms: Is Latin America Different?" *Economía* 1(2):47–99.

Goldfajn, Ilan, and Rodrigo O. Valdés. 1999. "The Aftermath of Appreciations." *Quarterly Journal of Economics* 114(1):229–62.

International Monetary Fund (IMF). Various years. *Exchange Arrangements and Exchange Restrictions*. Washington, D.C.: IMF.

———. Various years. International Financial Statistics. Washington, D.C.: IMF.

Kaminsky, Graciela L., and Carmen M. Reinhart. 1999. "The Twin Crises: The Causes of Banking and Balance of Payments Problems." *American Economic Review* 89(3):473–500.

———. 1998. "On Crises, Contagion, and Confusion." Mimeo, University of Maryland, November.

Kaminsky, Graciela L., Saul Lizondo, and Carmen M. Reinhart. 1998. "Leading Indicators of Currency Crises." *IMF Staff Papers*, March.

Rigobón, Roberto. 1999. "On the Measurement of the International Propagation of Shocks." Mimeo, MIT.

Sachs, Jeffrey, Aaron Tornell, and Andres Velasco. 1996. "Financial Crises in Emerging Markets: The Lessons from 1995." *Brookings Papers on Economic Activity* 1.

THE WORLD BANK ECONOMIC REVIEW, VOL. 15, NO. 2 315–340

Mutual Fund Investment in Emerging Markets: An Overview

Graciela L. Kaminsky, Richard K. Lyons, and Sergio L. Schmukler

International mutual funds are key contributors to the globalization of financial markets and one of the main sources of capital flows to emerging economies. Despite their importance in emerging markets, little is known about their investment allocation and strategies. This article provides an overview of mutual fund activity in emerging markets. It describes their size, asset allocation, and country allocation and then focuses on their behavior during crises in emerging markets in the 1990s. It analyzes data at both the fund-manager and fund-investor levels. Due to large redemptions and injections, funds' flows are not stable. Withdrawals from emerging markets during recent crises were large, which is consistent with the evidence on financial contagion.

One of the most remarkable characteristics of the financial crises of the 1990s is the speed at which they spread to other countries. The Mexican crisis in December 1994 prompted speculative attacks in Argentina and Brazil during the first quarter of 1995. The 1997 Thai crisis reached Malaysia, Indonesia, and the Philippines within days. Unlike these earlier crises, the 1998 Russian crisis was not confined to regional borders; it spread quickly to countries as distant as Brazil and Pakistan. Even developed countries were affected, with the default and devaluation reverberating in financial markets in Germany, the United States, and the United Kingdom.

The time clustering of crises in different countries has generated a vast literature on contagion (a term broadly understood as the cross-country spillover of

Graciela L. Kaminsky is with George Washington University. Her e-mail address is *graciela@gwu.edu*. Richard K. Lyons is with University of California, Berkeley, and the National Bureau of Economic Research. His e-mail address is *lyons@haas.berkeley.edu*. Sergio L. Schmukler is with the World Bank. His e-mail address is *sschmukler@worldbank.org*. The authors thank François Bourguignon and two anonymous referees for numerous helpful suggestions. They also benefited from comments from Stijn Claessens, Kristin Forbes, Jonathan Garner, Andrew Karolyi, Subir Lall, and participants at the International Monetary Fund (IMF)/World Bank conference on "Contagion," IMF/World Bank Brown Bag Lunch, and a workshop at the Institute of Development Studies, University of Sussex. For help with data the authors thank Erik Sirri of the U.S. Securities and Exchange Commission, Konstantinos Tsatsaronis from the BIS, and Ian Wilson from Emerging Market Funds Research. For excellent research assistance we thank Cicilia Harun, Sergio Kurlat, and Jon Tong. For financial support they thank the National Science Foundation, the World Bank (Latin American Regional Studies Program and Research Support Budget), and the World Bank Research Advisory Unit.

crises).[1] Many of these studies focus on the role of financial links. There is evidence, for example, that banks were important in spreading the 1997 crisis, due to the "common-lender channel" (Kaminsky and Reinhart 2000; Van Rijckeghem and Weder 2000).[2] The role of portfolio investors (foreign and domestic) during crises has also been under scrutiny,[3] with some researchers finding evidence of institutional panic and herding. This type of behavior might have helped spread crises even to countries with strong fundamentals. Kaminsky, Lyons, and Schmukler (2000b) note that individuals, too, can contribute to institutional panic by fleeing from funds—particularly mutual funds—forcing fund managers to sell when fundamentals do not warrant selling.

Although research on portfolio flows and the role of institutional investors has expanded dramatically in the late 1990s, information on the importance and evolution of institutional investors in emerging markets is still fragmented. Moreover, the role of mutual funds in capital-flow reversals during crises has not yet been documented. This article complements previous research in these two areas. First, it provides an overview of the importance and behavior of international mutual funds in emerging markets.[4] Second, it examines whether mutual fund investment tends to be stable over time and during crises.

There are two key advantages—beyond growing importance—to studying mutual funds rather than other institutional investors. The first is data quality. U.S.-based mutual funds report holdings to the U.S. Securities and Exchange Commission (U.S. SEC) semi-annually. In addition, private companies compile mutual fund data at higher frequencies, typically quarterly, through surveys. These data enable both cross-sectional and time-series analysis. In contrast, other institutional investors, like pension funds and hedge funds, are not required to disclose holdings. (Nor do there seem to be sources that compile data for these investor types from voluntary disclosures.[5]) The second key advantage to studying mutual funds is that their allocations to emerging markets have grown considerably in scope and size. There are now specialized subcategories within the broader mutual fund category. Some funds specialize in a particular country, some within a region, and some specifically in emerging markets, whereas some invest in emerging markets as part of a global strategy.

1. Many of the papers are available at www.worldbank.org/contagion.

2. The common-lender channel refers to cases in which common international banks lend to different countries, which consequently become linked. When a crisis hits the common lenders, all countries tend to be affected by the crisis.

3. See, for example, Cumby and Glen (1990); Bekaert and Urias (1996); Brown, Goetzmann, and Park (1998); Eichengreen and Mathieson (1998); Frankel and Schmukler (1996, 1998, 2000); Levy Yeyati and Ubide (1998); Bowe and Domuta (1999); Borensztein and Gelos (1999); Kaminsky, Lyons, and Schmukler (2000a, 2000b); and Pan, Cham, and Wright (2001).

4. Mutual funds from developing countries are also becoming important in some countries, helping develop local capital markets. Those funds are not covered in this study, however.

5. To study the behavior of pension or hedge funds one would need estimates of portfolio changes. Brown, Goetzmann, and Park (1998) provide such estimates for hedge funds during the Asian crisis.

I. Brief History of Capital Flows

Private capital flows have become the main source of external financing for developing countries, far surpassing public funds and accounting for some 80 percent of all flows to developing countries.[6] The first increase in capital flows occurred in the 1970s (see figure 1), triggered by the 1973–74 oil shock and amplified by the growth of the Eurodollar market and a spurt in bank lending during 1979–81. Latin America was the main recipient, with net flows peaking at $41 billion in 1981. Flows in this episode took the form mainly of syndicated bank loans (figure 2). The pace of international lending came to an abrupt halt in 1982 with the increase in world real interest rates to levels not seen since the 1930s.

By the late 1980s, there was a revival of international lending, with capital flows to Latin America making a tremendous comeback. Capital flows to Asia also surged, increasing tenfold from their averages in the late 1980s. The composition of capital flows changed dramatically, with bank lending replaced by foreign direct investment and portfolio investment. Bank lending to both East Asia and Latin America declined from 70 percent of net private capital flows in the 1970s to about 20 percent in the 1990s (see figure 2). While foreign direct investment in East Asia and Latin America constitutes the largest share of capital flows, portfolio investment (bonds and equity) has also increased substantially, accounting for about 30 percent of capital flows in the 1990s. In absolute values, bond and equity flows to each region—excluding those counted as foreign direct investment—increased from $1 billion in 1990 to $40 billion in 1996, with bond flows exceeding equity flows in Latin America since 1994. (Reported equity flows are underestimated: Any equity flow meant to acquire more than 10 percent of a company's outstanding shares is recorded as foreign direct investment, which accounts for around 50 percent of total capital flows.)

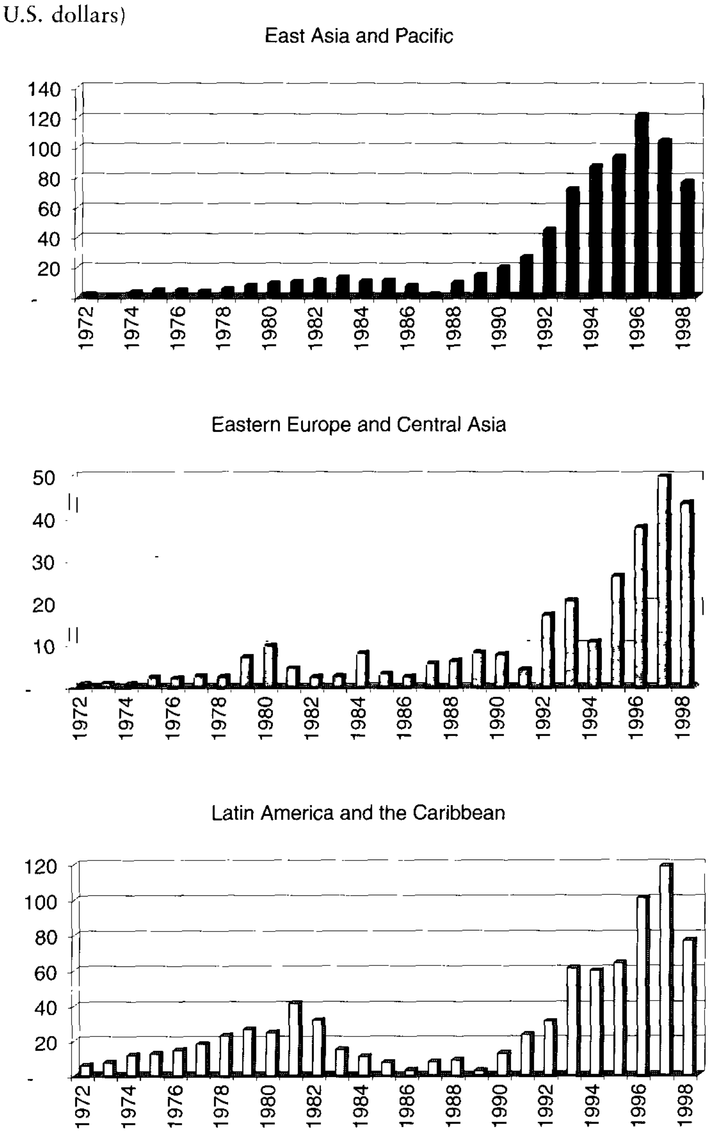
In the 1990s, as in the 1980s, booms were followed by a slowdown of capital inflows.[7] The first episode occurred in the immediate aftermath of Mexico's currency crisis in December 1994. Capital inflows resumed for most countries within six months and returned to their peak values soon thereafter. The crisis was confined to a small number of Latin American countries. Capital flows to Asian economies were largely unaffected. The second, more severe slowdown came in 1997, during the Asian crisis. The Russian default in August 1998 accentuated this slowdown, as capital flows collapsed. The change in inflows was similar in magnitude to that after the 1982 debt crisis, with total capital inflows declining about 35 percent to both Latin America and Asia.[8]

6. The data on capital flows come from World Bank databases. For more detailed description of capital flows, see World Bank (1997, 2000).

7. The term *reversal* is used in the literature in various ways. For some, a reversal is a shift from inflows to outflows. For others, a reversal is a reduction in inflows relative to what is expected.

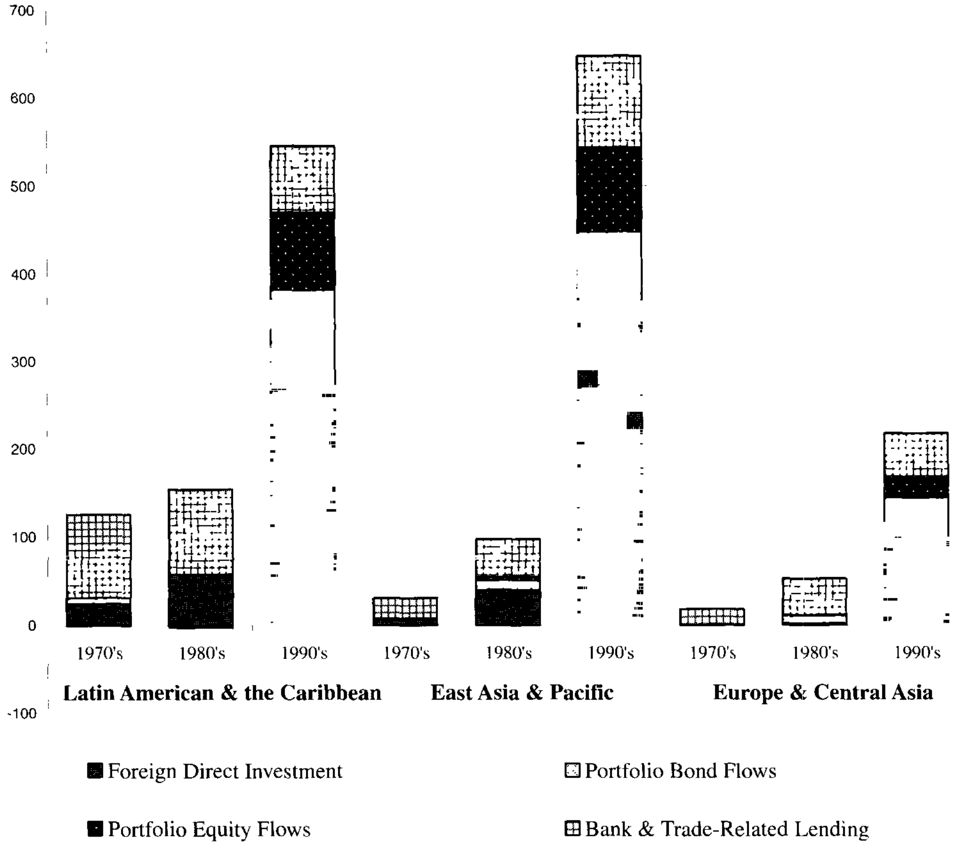
8. During the debt crisis, capital inflows declined about 24 percent in the first year of the crisis and 53 percent in the second year.

FIGURE 1. Total Net Private Capital Flows to Developing Countries, by Region 1972–98
(billions of U.S. dollars)



Notes: Net capital flows to developing countries include bank and trade-related lending, portfolio equity and bond flows, and foreign direct investment. The countries comprising Latin America and the Caribbean are Antigua and Barbuda, Argentina, Barbados, Belize, Bolivia, Brazil, Chile, Columbia, Costa Rica, Cuba, Dominica, Dominican Rebublic, Ecuador, El Salvador, Grenada, Guadeloupe, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, St. Kitts and Nevis, St. Lucia, St. Vincent and the Grenadines, Suriname, Trinidad and Tobago, Uruguay, and Venezuela. The countries comprising East Asia and Pacific are American Samoa, Cambodia, China, Fiji, Indonesia, Kiribati, Korea, Dem. Rep., Lao PDR, Malaysia, Marshall Islands, Micronesia, Fed. Sts, Mongolia, Myanmar, Palau, Papua, New Guinea, Philippines, Samoa, Soloman Islands, Thailand, Tonga, Vanuatu, Vietnam. The countries comprising Europe and Central Asia are Albania, Armenia, Azerbaijan, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Estonia, Georgia, Hungary, Isle of Man, Kazakstan. Kyrgyz Republic, Latvia, Lithuania, Macedonia FYR, Moldova, Poland, Romania, Russian Federation, Slovak Republic, Tajikistan, Turkey, Turkmenistan, Ukraine, Uzbekistan, and Yugoslavia FR (Serbia/Montenegro).

FIGURE 2. Type of Net Private Capital Flows to Developing Countries, 1970s to 1990
(billions of U.S. dollars)



Notes: See Figure 1 for countries included in each region.
Source: World Bank data.

The decline of short-term portfolio flows (bonds, equities, and bank lending) was even more pronounced, with flows falling about 60 percent in Latin America in 1998. Overall, bond and equity flows to Latin America declined from about $44 billion in 1996 to about $15 billion in 1998. Bond and equity flows to Asia collapsed in 1998 to $9 billion, from their peak in 1996 of $38 billion.

In sum, portfolio flows have become an important source of external financing in emerging markets. These flows have been unstable, with booms followed by pronounced reversals, and they have been channeled mainly through institutional investors, particularly mutual funds.

II. MUTUAL FUND INVESTMENT

Different data sources are needed to study the role of institutional investors. Unlike data on capital flows, which the World Bank collects on a regular basis, no agency has full detailed information on institutional investors. Institutions and companies like the Organisation for Economic Co-operation and Development (OECD), the U.S. SEC, the Investment Company Institute, Morningstar, Emerging Market Funds Research, Frank Russell, AMG Data Services, Lipper Analytical Services, and State Street Bank have partial information on institutional investors. The International Finance Corporation (IFC) has data on total market capitalization by country. Emerging Market Funds Research compiles data on dedicated emerging market funds. Morningstar and the U.S. SEC collect data on U.S. mutual funds. Data from the World Bank and the Bank for International Settlements (BIS) can be found elsewhere in the published literature in a similar format.

Getting an overall picture requires analyzing and combining data from various sources. This article contributes to the literature by compiling information from different sources and displaying it systematically and by presenting new evidence, though parts of the data are displayed elsewhere in a different format. The appendix summarizes the data sets used in this study and their sources.

Size of Mutual Funds and Institutional Investors

Institutional investors—including mutual funds, pension funds, hedge funds, and insurance companies—are a growing force in developed markets. Institutional investors held almost $11 trillion in the United States alone in 1995 (table 1). U.S. institutional investors accounted for more than half the assets held by institutions across the world.

When individual investors choose to allocate part of their portfolios to emerging markets, they typically make their purchase through mutual funds. In actively managed funds, it is the fund manager who ultimately determines the portfolio allocation by choosing how the fund invests its assets (within the limits of the fund's defined scope). In index funds, the manager's role is passive, aimed at replicating a predetermined index.[9]

Mutual funds based in developed countries have become one of the main instruments for investing in emerging markets.[10] The first funds, in the 1980s, were closed-end funds, which are well suited for investing in illiquid markets because their shares cannot be redeemed. As liquidity increased in emerging markets, the most widely used instrument became open-end funds. Mutual fund investors include other institutional investors as well as individual investors. For example, more than half of pension funds invest in emerging markets through existing mutual funds, for both liquidity and cost reasons (less expensive than giving specific mandates to fund managers). Therefore, in examining mutual funds, much of pension fund investment in emerging markets is covered as well. A World Bank (1997) survey estimates that pension funds hold around 1.5 to 2 percent of their portfolios ($50–$70 billion) in assets from emerging markets.

9. In all cases, but particularly for of index funds, fund managers tend to be evaluated against some benchmark indices. As a consequence, the behavior of managers is likely affected by these evaluations.

10. See New York Stock Exchange (2000) on U.S. investors in emerging market shares.

TABLE 1. Share of Global Assets Held by U.S.- and European-Based International Institutional Investors, 1995
(percent)

Institutional investor	U.S.-based	European-based
Pension funds	66	24
Insurance companies	37	37
Life insurance	35	36
Non-life insurance	45	37
Mutual funds	59	33
Open-end	65	34
Closed-end	57	41
Aggregate	52	32
Assets (billions of U.S. dollars)	10,994	6,666

Source: BIS 1998.

Hedge funds are a newer type of institutional investor. Still small relative to other institutional investors, hedge funds held estimated total assets of $81 billion by year-end 1997, only a small fraction of which is invested in emerging markets.[11] Like other institutional investors, insurance companies likely invest only a small proportion of their assets in emerging markets. However, unlike hedge funds, their asset holdings are large. More evidence on the investment allocation of this industry is needed.[12]

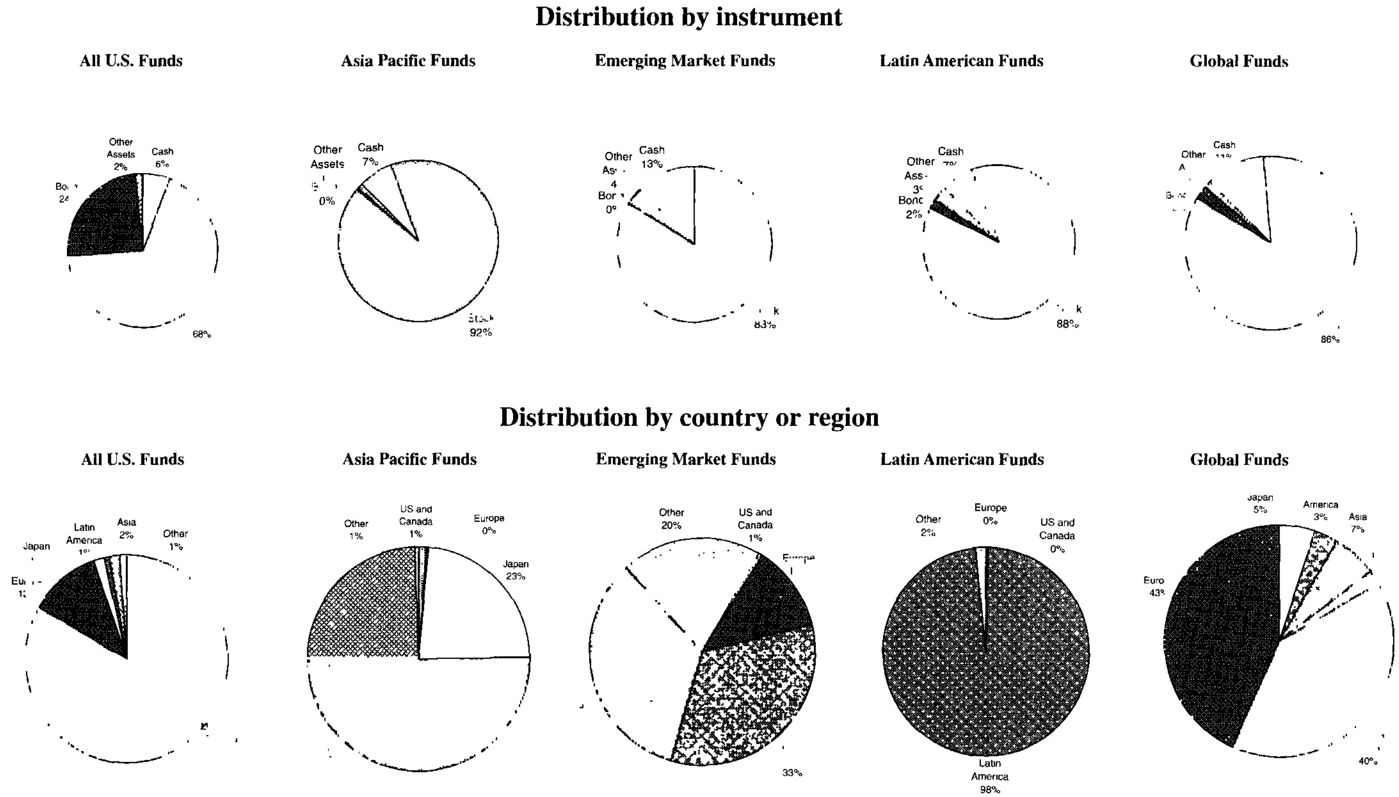
Of course, institutional investors in developed countries invest internationally not only in emerging markets but also in other developed economies. These broader, international portfolios are more concentrated in equities than in bonds (figure 3). Banks, for their part, tend to invest a bit more of their own assets and some of their clients' in foreign bonds. Despite the broader, international diversification of institutional investors, their portfolios still exhibit a strong home bias. For example, according to the World Bank (1997), U.S. equity pension funds held less than 9 percent of their assets in international instruments and around 2 percent in emerging markets (in 1994).

Even when international institutional investors hold only a small fraction of their portfolio in emerging markets, they have an important presence in these economies, given the relatively small size of their capital markets. Funds dedicated to emerging markets alone hold on average between 4 and 15 percent of the Asian, Latin American, and transition economies' market capitalization (table 2). By comparison, holdings of U.S. mutual funds accounted for 15 percent of the U.S. market capitalization in 1996 (see table 3). In Japan and the United Kingdom, domestic mutual funds held 4 and 8 percent of the local market capitalization that same year.

11. See Eichengreen and Mathieson (1998) for a detailed study of hedge funds.

12. Beyond institutional investors, it is difficult to determine the direct holdings of individual investors. No regulatory agencies (like the U.S. SEC or the BIS) or private companies (like Morningstar or Lipper Analytical Services) keep such records.

FIGURE 3. Distribution of U.S. Mutual Fund Assets by Fund Type as of December 31, 1998



Notes: Morningstar classifies the assets as being invested in one of six countries or regions: United States and Canada, Japan, Asia (excluding Japan), Europe, Latin America, or other. Holdings are classified in one of four asset classes: cash, stocks, bonds, or other. The Morningstar universe includes all types of mutual funds except money market funds. Funds that invest primarily outside the United States are mostly equity funds.

Source: Morningstar.

The estimates of the importance of mutual funds in emerging markets are quite conservative because they include only the holdings of dedicated emerging market equity funds.[13] Excluded are the holdings of global funds, which account for a substantially larger share of the stock market capitalization of emerging markets. Even though global funds hold only a small share of their assets in emerging markets, they are substantially larger than dedicated emerging market funds giving them a stronger presence.[14] Moreover, some of the outstanding equity in emerging markets—as well as in many developed countries—is not publicly traded because it belongs to families or corporations that control the companies. So international mutual funds hold a large and significant proportion of the publicly available equity, even though the total amount is not known.

The importance of mutual funds varied substantially during the 1990s (see table 2). Though net equity flows declined from their 1993 peak—about $27 billion to Latin America and $21 billion to Asia—the relative importance of mutual funds increased until 1997. For example, dedicated emerging market equity funds held $22 billion in Latin American stocks at the end of 1995 and nearly double that, $40 billion, by December 1997. Though mutual fund growth was less pronounced in Asia, mutual funds are still important in many countries. Overall, dedicated emerging market mutual funds held $77 billion in Asia at the onset of the crisis (December 1996). While the absolute amount of mutual fund investment in transition economies is not comparable to that in Asia and Latin America, fund growth in these transition economies has been remarkable. In market capitalization terms, mutual funds have become big players in these markets, with especially large positions in markets in Hungary and Poland.

The mutual fund industry specializing in emerging markets has a very concentrated portfolio by economies. At least half their total portfolio is invested in just six markets: Brazil, Hong Kong (China), Republic of Korea, Malaysia, Mexico, and Taiwan (China). Country shares have varied, sometimes substantially, in the 1990s. For example, Malaysia attracted about 12 percent of all the funds allocated to Asia in 1995 but only 4 percent after the crisis. In contrast, the share allocated to Indian assets increased from 7 percent to 14 percent. The proportion of assets allocated to countries in Latin America has been less volatile, with Brazil's share of the funds allocated to the region holding at about 40

13. Data on dedicated funds come from Emerging Market Funds Research, which collects aggregate data of emerging market mutual funds. They track the net cash flows of nearly 1,400 international emerging market equity funds, with an average position of about $120 billion in 1996. The data cover both U.S.-registered and offshore funds as well as funds registered in Luxembourg, the United Kingdom, Ireland, Cayman Islands, Canada, and Switzerland. It includes both open- and closed-end funds. The data set used in this article starts with the Mexican crisis of 1995 and ends in March 1999; it therefore includes observations on the major currency crises of the 1990s.

14. For example, estimates for the mutual fund industry more broadly suggest that international funds hold between 60 percent and 70 percent of the market capitalization in Hungary, in contrast to the estimates in table 3, which are all below 30 percent. We thank Jonathan Garner, from DLJ, for this information.

TABLE 2. Holdings of Dedicated Emerging Market Fund Assets and Their Share of Market Capitalization, by Country and Region, 1995–98

	1995		1996		1997		1998	
Economy	End-of-year holdings (billions of U.S. dollars)	Share of market capitalization[a] (percent)	End-of-year holdings (billions of U.S. dollars)	Share of market capitalization[a] (percent)	End-of-year holdings (billions of U.S. dollars)	Share of market capitalization[a] (percent)	End-of-year holdings (billions of U.S. dollars)	Share of market capitalization[a] (percent)
China	1.9	4	2.3	3	3.1	2	1.9	1
Hong Kong	12.6	n.a.	20.4	n.a.	13.2	n.a.	9.4	n.a.
India	4.5	3	6.1	4	7.4	5	5.6	5
Indonesia	4.5	9	5.5	7	1.9	2	1.3	7
Korea, Rep. of	10.3	6	7.7	5	2.5	2	7.3	11
Malaysia	8.2	4	12.0	4	2.4	1	1.5	2
Pakistan	0.6	6	0.5	5	0.8	7	0.2	3
Philippines	3.4	6	4.2	6	1.7	3	1.9	6
Singapore	5.1	n.a.	5.3	n.a.	3.0	n.a.	3.8	n.a.
Sri Lanka	0.2	9	0.1	5	0.2	10	0.1	7
Taiwan, China	4.6	2	7.2	3	5.9	2	5.7	2
Thailand	9.8	7	5.9	4	2.2	4	3.1	10
Total Asia	65.7	6	77.2	5	44.2	4	41.7	5
Argentina	3.1	9	3.3	8	4.6	9	3.1	6
Brazil	8.1	5	11.5	6	15.4	6	8.3	4
Chile	3.4	5	2.9	4	3.4	4	2.6	4
Colombia	0.4	2	0.6	4	0.6	3	0.3	2
Mexico	5.5	6	7.8	7	13.4	10	7.9	7
Peru	0.7	7	0.9	7	1.1	6	0.7	5
Venezuela	0.3	6	0.7	12	1.2	9	0.5	5
Total Latin American	21.5	6	27.7	7	39.7	7	23.2	5
Czech Republic	0.5	3	1.0	6	1.0	6	0.7	6
Hungary	0.4	25	1.2	29	2.3	26	2.2	16
Poland	0.7	18	1.5	20	1.9	17	2.2	14
Russia[b]	1.0	n.a.	2.6	10	7.5	7	1.7	3
Slovak Republic	0.1	n.a.	0.1	4	0.1	5	0.1	8
Total transition economies	2.7	15	6.4	14	12.8	12	7.0	10

n.a. is not available.

Note: Data cover only the holdings of the dedicated emerging market funds (based inside and outside the United States). Thus the importance of all foreign mutual funds in each country is significantly larger in most cases. The International Finance Corporation database does not contain market capitalization for some countries (shown as n.a. in the table).

[a]Share of country's stock market capitalization.

[b]Includes other members of the Commonwealth of Independent States.

Source: Emerging Market Funds Research and International Finance Corporation.

TABLE 3. Share of Total Mutual Fund Assets Held by Selected Developed Country-Based Funds, by Asset Type, 1996
(percent)

	United States	Japan	United Kingdom	France
Money market funds	25	29	0	45
Bond funds	22	45	5	29
Equity funds	49	24	88	11
Balanced funds	3	2	6	14
Share of total	76	9	4	11
As percent of GDP	46	9	16	34
As percent of market capitalization	15	4	8	18

Source: BIS 1998.

percent. Among transition economies, five countries account for most mutual fund investment: the Czech Republic, Hungary, Poland, Russia (and other members of the Commonwealth of Independent States), and the Slovak Republic. Again, the shares of crisis countries in the mutual fund portfolio swing substantially; Russian holdings varied from 25 percent to 59 percent of mutual funds' portfolios in transition economies.

Mutual funds also hold large positions in American and Global Depositary Receipts, (ADRs and GDRs), typically traded on the New York Stock Exchange, NASDAQ, and the American Stock Exchange. Therefore, mutual funds often do not trade in the local stock markets when investing abroad.[15]

Holdings of U.S.-Based Mutual Funds

U.S.-based mutual funds accounted for almost 60 percent of world mutual funds in 1995 (see table 1). The U.S. mutual fund industry expanded significantly during the 1990s (table 4). From 1991 to 1998 the number of bond and stock funds increased from 2,355 to 10,144 net assets rose from $705 billion to $3.6 trillion. The 20 largest U.S. mutual funds capture only a small proportion of all the assets of the U.S. mutual funds industry (not more than 4 percent).

The exposure of U.S. mutual funds to emerging markets increased substantially during the 1990s (see table 4). U.S.-based, open-end mutual funds (including Asia Pacific, Latin American, and emerging market funds) had around $35 billion by the end of 1996, up from about $1 billion at the end of 1991. As Asia Pacific funds grew from 11 funds in 1991 to 154 in 1998, their net assets rose from $1 billion in 1991 to $16.4 billion in 1996 and then fell to $6.5 billion in 1998 following the Asian crisis. Mutual funds specializing in emerging markets increased from 3 funds in 1991 to 165 funds in 1998, with total net assets rising

15. See Karolyi (1998) for a broad-based survey of global cross-listings. Also see Smith and Sofianos (1997) for a study of the effects of depositary-receipt listing in the New York Stock Exchange.

TABLE 4. Size of Mutual Fund Universe, as of December 31, 1991–98

	1991	1992	1993	1994	1995	1996	1997	1998
All U.S. funds								
Net assets (billions of U.S. dollars)	705	933	1,338	1,428	1,838	2,335	2,954	3,570
Number of funds	2,355	2,522	3,422	5,594	6,937	7,746	8,655	10,144
Net assets of 20 largest funds as percentage of all U.S. funds	2	2	3	4	3	4	4	3
Asia Pacific funds								
Net assets (billions of U.S. dollars)	1.1	1.4	8.4	11.9	12.1	16.4	9.0	6.5
Number of funds	11	14	27	59	79	106	127	154
Net assets of 20 largest funds as percentage of all Asia Pacific funds	100	100	97	94	94	93	90	82
Emerging market funds								
Net assets (billions of U.S. dollars)	0.1	0.5	3.7	8.7	8.5	15.6	16.9	13.5
Number of funds	3	7	10	32	64	94	119	165
Net assets of 20 largest funds as percentage of emerging market funds	100	100	100	92	89	72	71	67
Latin American funds								
Net assets (billions of U.S. dollars)	0.04	0.2	1.3	3.9	2.5	2.9	4.1	1.8
Number of funds	1	3	5	15	25	28	35	47
Net assets of 20 largest funds as percentage of Latin American funds	100	100	100	100	73	95	97	95
Global funds								
Net assets (billions of U.S. dollars)	16.1	18.3	28.1	45.4	58.1	82.0	108.1	125.4
Number of funds	52	56	78	143	180	198	223	273
Net assets of 20 largest funds as percentage of world funds	81	80	74	73	71	76	79	77

Source: Morningstar.

from $142 million in 1991 to $13.5 billion in 1998 (after peaking at $17 billion in late 1997). The number of Latin American funds increased from 1 to 47, and their net assets rose dramatically from $44 million to $1.8 billion. Global funds increased from 52 to 273, with total net assets rising from $16 billion to $125 billion. With the exception of mutual fund investment in U.S. assets, the mutual fund industry is highly concentrated, with the largest 20 funds holding about 80 percent or more of all assets.

Until 1993 bonds constituted the largest share of mutual fund portfolios. After that, equities began to predominate. By 1998, for mutual funds overall, about 68 percent of their portfolio was allocated to stocks; most of the rest (between 24 and 40 percent) was allocated to bonds (see figure 3). The proportion of assets held in stocks is substantially larger for mutual funds specializing in emerging markets (including Asia Pacific and Latin America), varying from 83 percent to 92 percent. Global funds also hold a large share of their assets in stocks (86 percent).

The country or regional composition of the total U.S. mutual fund portfolio in 1998 was 83 percent U.S. and Canadian stocks, 12 percent European stocks, 1 percent Japanese assets, 2 percent Asian assets, and 0.9 percent Latin American assets. Although the percentage dedicated to emerging markets is small, the large size of the U.S. mutual fund industry implies that the dollar amount held in assets from emerging countries is significant.

III. Behavior of Mutual Funds during Crises

The financial crises of the 1990s spread beyond the country and even the region of origin. Financial disruption spread to countries as far apart as Argentina, the Czech Republic, and South Africa. Crises before 1990 had also led to contagion—witness the debt crisis in 1982—but that contagion had tended to be regional until recently (with some exceptions). That changed in the 1990s. The East Asian crisis triggered financial disruption as far away as Argentina, Mexico, and Chile. The speculative attack on the Hong Kong dollar in October 1997 also spilled over into other markets. Even the U.S. stock market suffered sizable losses after the Hang Seng index fell 15 percent. The cross-country effects became more widespread following the Russian default in August 1998, with stock prices in all industrial countries declining between 20 percent and 50 percent. Contagion in these recent crises has been partly attributed to global financial links.

Studies have shown that the behavior of mutual funds and contagion may be linked, either because funds generate cross-country spillovers or because funds engage in feedback trading (trading in response to past returns, such as selling when past returns are low). International mutual funds can contribute to the spread of crises across countries through spillover effects if, for example, investors holding fund shares decide to sell their Asian funds when Russia devalues its currency. Or if managers of Latin American funds sell assets in Brazil when a crisis hits Mexico. These need not be irrational responses: New models of rational herding explain the transmission of crises through financial links. These models involve asymmetric information and cross-market hedging.[16] If mutual fund investors or managers also engage in feedback trading, their behavior can appear consistent with contagion even though mutual funds may not be the main force driving the spillovers.

Institutional investors, such as mutual funds, can also be a stabilizing force. If investors buy mutual fund shares for long-run gains, they might not withdraw their investments during a crisis. Marcis, West, and Leonard-Chambers (1995) and Rea (1996) find that shareholders did not redeem shares during crises periods. They find that net inflows to emerging markets are usually steady, and crisis-period outflows are small and short lived (at least during Mexico's crisis). Froot, O'Connell, and Seasholes (2001) present a related picture, but without focusing on institutional investors. Though net flows into individual emerging markets decreased during the Mexican and Asian crises, Froot, O'Connell, and Seasholes find little evidence of net outflows.[17]

This section provides evidence on the stability of mutual fund investment and the behavior of mutual funds following speculative attacks, distinguishing where possible between the behavior of mutual fund managers and underlying investors.[18]

Mutual Fund Flows

On balance, flows of dedicated emerging market mutual funds to Asia, Latin America, and transition economies (data from Emerging Market Funds Research) since 1995 have been muted, reaching about $20 billion, with booms in capital inflows followed by pronounced outflows (figure 4). Outflows from Latin America reached about $4 billion in 1995, but mutual funds increased their positions in Latin America by about $2 billion in the first half of 1996. The Mexico crisis did not spill over to Asia or the transition economies. In fact, flows to Asia ballooned to almost $11 billion, whereas flows to transition economies remained stable throughout 1995–96.

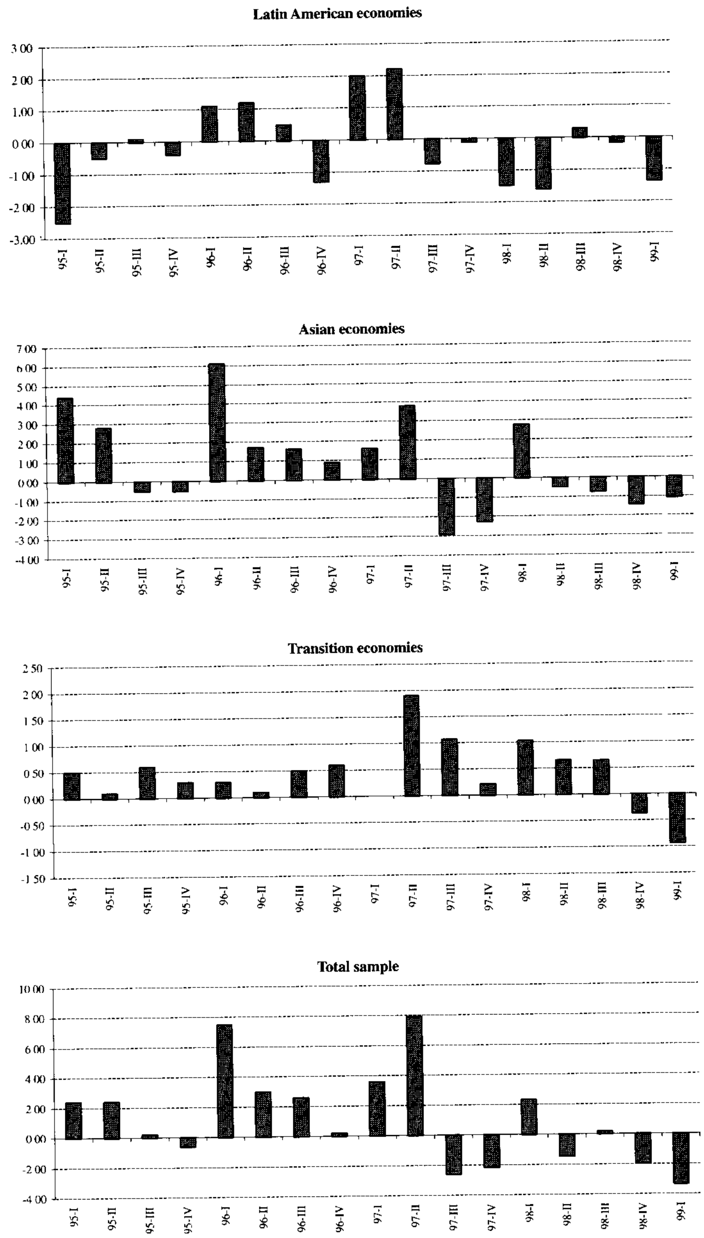
The picture changes during the currency turmoil in Asia in the second half of 1997. This time, mutual funds pulled out not only from Asia but from Latin

16. For example, in the Calvo and Mendoza (1998) model, the costs of gathering country-specific information induce rational investors to follow the herd. In the Calvo (1998) model, uninformed investors replicate selling by liquidity-squeezed informed investors, because the uninformed investors mistakenly (but rationally) believe that these sales are signaling worsening fundamentals. Kodres and Pritsker (1999) focus on investors who engage in cross-market hedging of macroeconomic risks. They find that international market co-movement can occur in the absence of any relevant information, even in the absence of direct common factors across countries. For example, a negative shock to one country can lead informed investors to sell that country's assets and buy assets of another country, increasing their exposure to the idiosyncratic factor of the second country. Investors then hedge this new position by selling the assets of a third country, completing the chain of contagion from the first country to the third.

17. Froot, O'Connell, and Seasholes (2001) and Choe, Kho, and Stulz (1999) are able to study the dynamics of capital flows during crises using higher-frequency data. But their data are aggregated across types of investors, so they cannot focus on the role of different kinds of institutional investors, as is done here.

18. This section examines data sets from various sources, including the Emerging Market Funds Research, Morningstar, the U.S. SEC, and the BIS.

FIGURE 4. Mutual Fund Quaterly Flows to Emerging Market Economies
(billions of U.S. dollars)



Notes: Latin America includes Argentina, Brazil, Chile, Colombia, Mexico, Peru, and Venezuela. Asia includes China, Hong Kong, India, Indonesia, Republic of Korea, Malaysia, Pakistan, Philippines, Singapore, Sri Lanka, Taiwan (China), and Thailand. Transition economies include Czech Republic, Hungary, Poland, Russia and other members of the Commonwealth of Independent States, and the Slovak Republic.

Source: Emerging Market Funds Research.

America as well, with net outflows from Latin America reaching about $1 billion in the six months following the collapse of the baht. In Asia, flows rebounded briefly in the first quarter of 1998 but declined thereafter. Overall in 1998, mutual fund withdrawals took a turn for the worse, reaching about $4 billion in Asia, with substantial outflows from Latin America and transition economies.

A closer look at the spillover effects surrounding the Mexican, Thai, and Russian crises shows how a crisis in one country triggers withdrawals in other countries. To isolate the behavior of mutual funds in crisis times, the average quarterly mean flow (as a percentage of funds' initial positions) during the entire sample period (1995–99) is subtracted from net buying or selling (see figure 5).[19] For example, following the Mexican devaluation, mutual funds sold about 5 percent of their Brazilian positions relative to their average quarterly buying and selling during 1995–99. Thus Brazil experienced unusual withdrawals of about 5 percent in the aftermath of the Mexican devaluation. Looking at country data according to the severity of the outflows conveys more clearly the extent of contagion across regions following the initial speculative attack. Thus, for example, Malaysia was most affected in the aftermath of the Russian crisis, with abnormal outflows of approximately 30 percent.

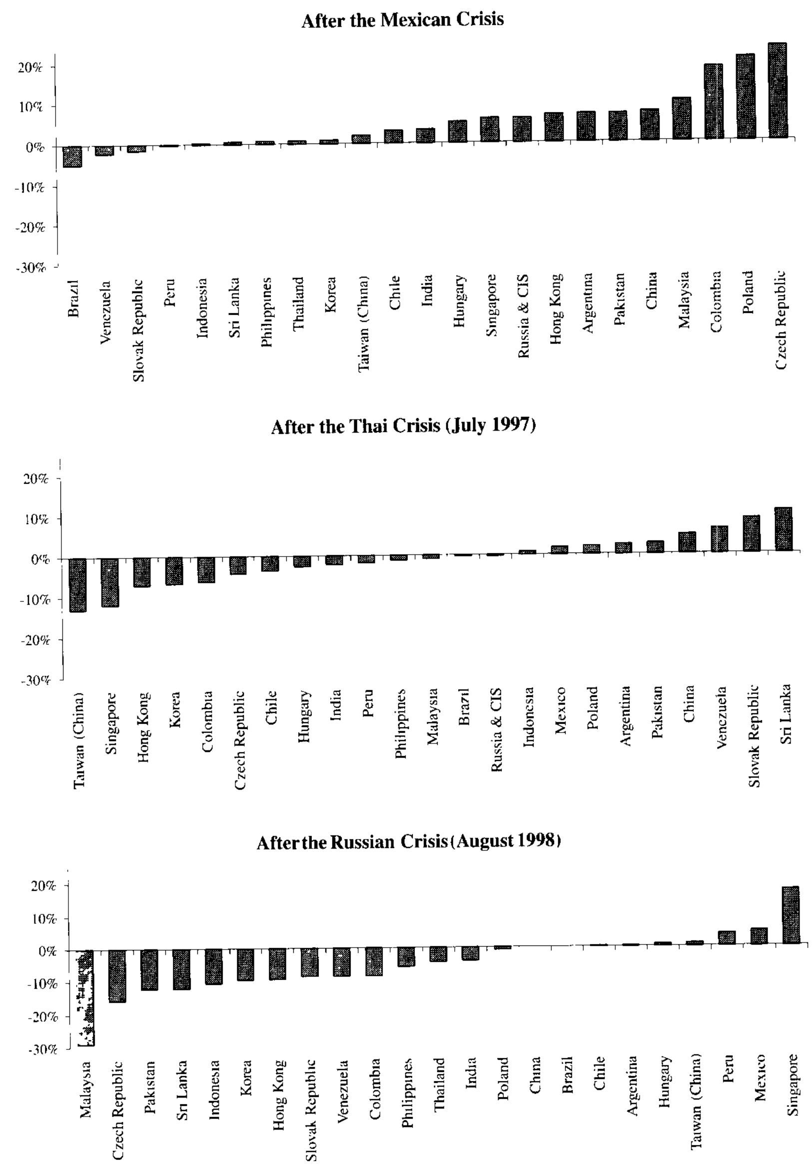
The repercussions of the three crisis episodes were dramatically different. The Mexican crisis was concentrated in Latin America and was confined to a handful of countries. Only Brazil and Venezuela—in addition to the crisis country, Mexico—suffered abnormal average withdrawals (of 5 percent and 2 percent) in the two quarters following the devaluation. Mutual funds increased their exposure to Asian countries and transition economies, with (above-trend) flows of around 4 percent for Asia and 11 percent for the transition economies.

The aftermath of the collapse of the Thai baht presents a different picture, with signs of a more general retrenchment of mutual funds in emerging markets. Mutual fund flows to Asian economies were basically all well below trend in the two quarters following the collapse of the Thai baht, except for flows to China, Pakistan, and Sri Lanka, which were above average. Withdrawals were also substantial from Hong Kong (12 percent), Singapore (7 percent), and Taiwan, China (12 percent). This time the retrenchment also reached Latin America and the transition economies, with average quarterly withdrawals reaching about 6 percent for Colombia and 4 percent for the Czech Republic during the two quarters following the outbreak of the Thai crisis. Colombia, the Czech Republic, Hungary, and Peru were most affected in this episode, with outflows of 3 percent or more above average.

Even more pronounced was the flight away from emerging during the Russian crisis. About half the countries in the sample experienced abnormal withdrawals of 10 percent or more. In some cases, withdrawals were massive: 30 percent in Malaysia and 16 percent in the Czech Republic. Some Latin American countries were also dramatically affected. Colombia and Venezuela suffered

19. Models of asset trade (such as microstructure finance models) provide a theoretical basis for focusing on changes in flow relative to what is expected, which here is proxied by average flow.

FIGURE 5. Mutual Fund Net Buying or Selling Following Recent Crises, by Country



Notes: The figures show the spillover of crises to other developing countries. The figures display average mutual fund flows (net buying or selling as a percentage of the end of the preceeding quarter holdings) in the two quarters following the outbreak of the crisis, after subtracting the sample average for the study period.

Source: Emerging Market Funds Research.

average quarterly outflows of about 8 percent. Only Mexico and Peru did not suffer following the worldwide financial disruption triggered by the Russian default. In fact, inflows to Mexico were 5 percent above average flows.

Several different factors seem to affect the varied pattern of responses in mutual fund flows across countries after a crisis erupts, including economic vulnerability and liquidity of financial markets.[20] Other risk factors—such as a possible change in political authority, severe new restrictions on purchases and sale of assets, or a debt moratorium—might also affect mutual fund investment behavior.

Not surprisingly, analysis shows that economic vulnerability matters. For example, the Czech Republic and Russia suffered severe outflows in 1997 and 1998, when both countries were in economic distress. Kaminsky, Lyons, and Schmukler (2000a) find that other factors also influence investors' withdrawals from emerging markets. In the case of China, for example, devaluation fears were widespread among investors and the vulnerability of its financial system was widely known, yet it did not suffer from the Asian crisis. In contrast, Singapore, Taiwan (China), and Hong Kong—the most liquid markets in Asia—suffered pronounced capital-flow reversals, even though their economies looked far healthier than China's. Though less important, risk factors, such as attacks against foreign investors and political instability, helped explain the outflows in Pakistan and Malaysia during the Russian crisis. Moreover, the crises that began in Mexico, Thailand, and Russia also showed increasing degrees of spillover, pointing to systemic factors in addition to country-specific factors.

Investors and Managers

Though mutual funds are commonly included among institutional investors, they differ from hedge funds, pension funds, and insurance companies in how much control underlying investors have over portfolio size. Fund behavior is thus determined by the decisions of both managers and investors.[21] This hybrid nature affects mutual fund flows to countries and regions.

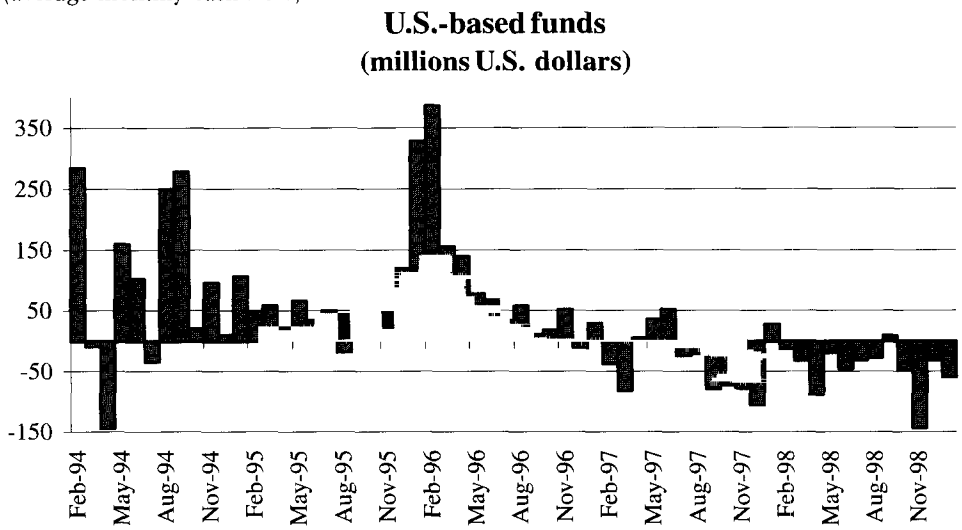
This characteristic provides an opportunity to study in detail the behavior of these two groups of agents. Kaminsky, Lyons, and Schmukler (2000b) focus on whether the trading strategies of these two groups are driven by current and past returns (positive-feedback trading—the buying of past winners and selling of past losers). This section provides additional evidence on the influence of each group, using detailed data from the BIS and the U.S. SEC, which help isolate the behavior of the two groups.

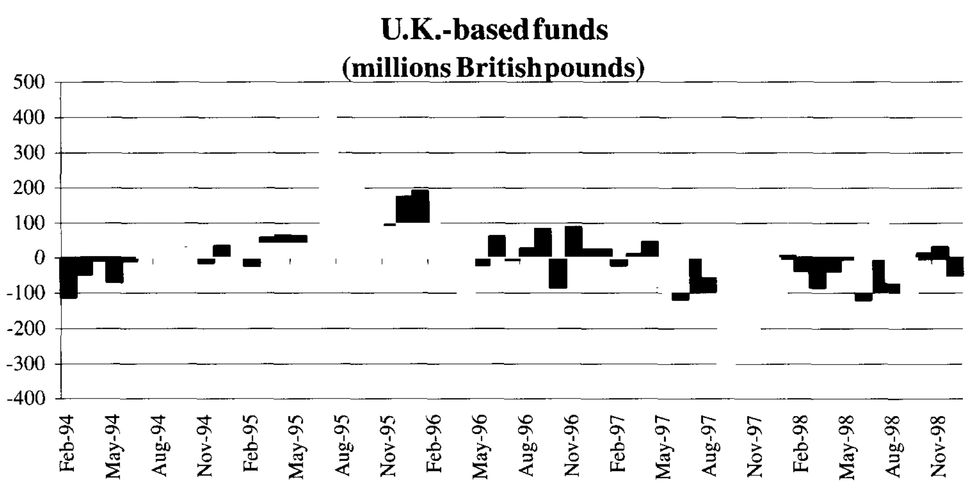
20. See Kaminsky, Lyons, and Schmukler (2000a) for more detail on the probable determinants of mutual fund behavior in crises.

21. *Mutual funds* here means open-end, nonindex funds, which account for most of the funds that invest in emerging markets. In closed-end funds, investors do not control portfolio size. In index funds, managers have little control over portfolio holdings.

INVESTORS. The behavior of underlying investors is described in figures 6 and 7. Cash flows to Asian mutual funds (from U.S.-and the U.K.-based funds), a decision belonging to investors, were high before the Asian crisis, particularly in 1995–96. After the Thai devaluation of 1997, large outflows began and continued in 1998. Outflows were particularly large for U.S.-based funds after the Russian crisis.

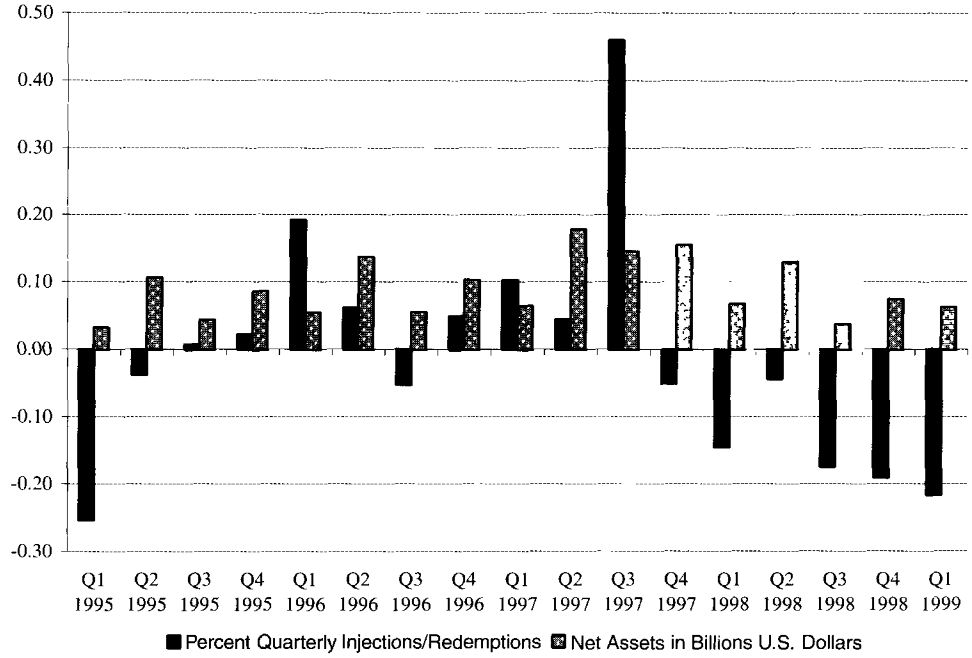
FIGURE 6. Injections and Redemptions in U.S.- and U.K.-Based Asian Equity Mutual Funds
(average monthly cash flow)





Note: Positive figures are injections; negative figures are redemptions.
Source: BIS 1998.

FIGURE 7. Net Assetsand Injections and Redemptions in U.S.-Based Latin American Funds



Note: All funds are open-end. The figure reports net assets and aggregate values (across funds) of quarterly injections or redemptions. Injections reflect percentage increases in the number of outstanding shares; redemptions reflect net decreases. For injections and redemptions, a value of 0.1 stands for 10 percent. For net assets (in billions of U.S. dollars) a value of 0.1 stands for $100 million.

Source: Morningstar and U.S. Securities and Exchange Commission.

Injections and redemptions for 13 Latin American mutual funds,[22] again, part of the investors' decision set, show a pattern of inflows and outflows consistent with the broad features of the recent crises reviewed above (see figure 7). Injections are measured by the percentage increase in the number of shares held by each mutual fund and redemptions by the percentage decrease, to control for fund-size changes due to capital gains and losses. Large redemptions from Latin American funds accounted for 25 percent of the outstanding shares in the first

22. The data come from Morningstar and the U.S. SEC. The sample here includes holdings of the largest 13 Latin-America equity funds (open-end) from April 1993 to January 1999 (24 quarters): Fidelity Latin America, Morgan Stanley Dean Witter Institutional Latin America, Van Kampen Latin America (formerly Morgan Stanley), BT Investment Latin America Equity, TCW Galileo Latin America Equity, TCW/Dean Witter Latin America Growth, Excelsior Latin America, Govett Latin America, Ivy South America, Scudder Latin America, T. Rowe Price Latin America, Merrill Lynch Latin America, and Templeton Latin America. These funds did not all exist from the beginning of our sample; on average we have about 10 quarters of data (out of a possible 24) per fund.

quarter of 1995 during the Mexican crisis. Thereafter, injections resumed to Latin American funds until the last quarter of 1997, during the Asian crisis. Redemptions continued during 1998, increasing during the Russian crisis, reaching 20 percent in late 1998 and early 1999. Fluctuations in injections and redemptions influence the funds' net assets, which are also determined by movements in underlying stock prices.

The patterns in figures 6 and 7 are closely associated with those in figure 4 on average quarterly flows. During the Mexican crisis, as investors pulled out of Latin American funds, there was a large outflow from Latin American countries. Then investors and flows returned to Latin American countries until the last quarter of 1997, when the Thai crisis expanded to other countries. In Asia, there are no signs of fund outflows or investor redemptions during the Mexican crisis, but there are large effects during the Asian crisis. This pattern suggests that investors' decisions influence fund flows.

MANAGERS. Managers cannot control the injections or redemptions of underlying investors. What they can control is the use of cash or short-term positions (for example, U.S. Treasury bills), which help buffer portfolios from the effects of redemptions. Holding assets that are highly liquid allows managers to meet redemptions without selling less liquid assets. In principle, this can mute the volatility caused by investment outflows. However, managers can also reinforce investors' actions and amplify volatility if they increase their liquidity positions in times of investor retrenchment. In multiple-country portfolios, managers make the decision about which country to withdraw from.[23]

Interestingly, managers' choices about short-term positions do not change as funds experience redemptions or injections (table 5). On average, the funds in the sample hold 5 percent of their assets in liquid positions. Examining short-term positions in more detail by size of the fund shows that large funds hold a larger share of their positions in liquid assets. This finding is somewhat unexpected because large funds are likely to have better access to bank credit lines than smaller funds and thus not to need to hold large liquid positions. Both large and small mutual funds hold smaller liquid positions in times of redemption, indicating that fund managers' behavior helps smooth the effects of investors' withdrawals on equity markets in Latin America. By contrast, medium-size funds hold more liquid assets in times of redemption, thus magnifying investors' withdrawals from emerging markets.

23. Investors obviously determine the withdrawal country in the case of single-country funds. There are two drawbacks to this data set. First, the data are only from Latin American funds. In the future, it will be interesting to study the behavior of managers by considering a broader set of mutual fund types. Second, the data do not provide a complete picture of managers' responses to liquidity squeezes because information on funds' credit lines with banks is lacking. Funds facing large redemptions may have resorted to using such credit lines.

TABLE 5. Average Short-Term Positions of U.S.-Based Latin American Funds, 1995–98
(percentage of total net assets)

	All times	Injection times	Redemption times
All funds	4.44	4.57	4.37
Large funds	6.97	8.40	5.22
Medium funds	3.81	2.24	4.40
Small funds	4.16	4.48	3.61

Note: Injection times are defined as periods when the number of the fund's outstanding shares increases; redemption times as periods when the number decreases. See appendix table for list of companies in each fund size category.
Source: U.S. SEC.

IV. CONCLUSIONS

The increasing globalization of financial markets and the crises of the 1990s have spawned a vigorous literature on financial integration, international financial architecture, and contagion. A central element of the debate is the behavior of financial markets. In particular, many have argued that financial markets are volatile and prone to contagion. Most of the literature has focused on market imperfections and how they lead to herding behavior and financial cycles that are unrelated to market fundamentals.

Though studies have covered several dimensions of foreign investors' role in emerging markets, this article provides an overview of a missing dimension—the behavior of international mutual funds. Institutional investors are the main channel of financial flows to emerging markets, and mutual funds are a large part of institutional investors. They are the only class of institutional investors for which reliable data are available on an ongoing basis.

Several general findings emerged. Equity investment in emerging markets has grown rapidly in the 1990s. A significant proportion of that equity flow is channeled through mutual funds. Collectively, these funds are large investors and hold a sizable share of market capitalization in emerging markets.

Among mutual funds, Asian and Latin American funds achieved the fastest growth. Their size remains small, however, compared with domestic U.S. funds and global funds.

When investing abroad, U.S. mutual funds hold mostly equity rather than bonds. Global funds invest mainly in developed nations (the United States, Canada, Europe, and Japan), with just 10 percent of their investment devoted to Asia and Latin America. Mutual funds generally invest in a subset of countries within each region. In Latin America, they invest primarily in Brazil and Mexico, then in Argentina and Chile. In Asia, the largest shares are in Hong Kong, India, Korea, Malaysia, Taiwan (China), and Thailand. In transition economies, mutual funds invest most of their assets in the Czech Republic, Hungary, Poland, and Russia (and other members of the Commonwealth of Independent States).

Mutual fund investment was very responsive during the crises of the 1990s. The Mexican crisis affected mostly Latin America, and the Asian and Russian crises had a large impact on Asian and Latin American funds. These findings are consistent with previous findings on contagion and with reports by industry analysts.

Investment by underlying investors of Asian and Latin American funds is volatile. Injections and redemptions are large relative to total funds under management. The cash held by managers during injections and redemptions does not fluctuate significantly, so investors' actions are typically reflected in emerging market inflows and outflows.

Many questions remain that could be addressed in future research. To test theories of financial crises, it would be valuable to examine the link between institutional investor behavior and country and market characteristics. It would also be useful to compare the behavior of different fund types—such as global, emerging market, and regional funds—to provide evidence for discussions of international financial architecture. These are areas that we are currently researching. Beyond studying institutional investors, it would also be interesting to analyze the behavior of banks' proprietary trading in emerging markets. This is an area where hard evidence is almost completely lacking.

APPENDIX. Data Sets and Sources

Data set	Source	Use	Description
Net private capital flows	World Bank	Figure 1 Figure 2	Net capital flows to developing countries, including the so-called emerging economies, typically middle-income developing countries. The amounts include bank and trade-related lending, portfolio equity and bond flows, and foreign direct investment. The countries included in each region are detailed in the figures.
International institutional investors	BIS, *68th Annual Report*	Table 1, Table 2, Figure 6	Distribution of institutional investors between the U.S. and Europe in 1995. Monthly averages of cash to and from Asian funds in the United States and the United Kingdom. Size of the mutual fund industry in the United States, Japan, the United Kingdom, and France.
Dedicated emerging market funds	Emerging Market Funds Research	Table 3, Figure 4, Figure 5	Country holdings of dedicated emerging market funds, based inside and outside the United States. The data are aggregate, tracking nearly 1,400 international emerging market equity funds, with an average position of about $120 billion in 1996. The data set covers both U.S. registered and offshore funds as well as funds registered in Luxembourg, United Kingdom, Ireland, Cayman Islands, Canada, and Switzerland. It includes both open- and closed-end funds.
Market capitalization	International Finance Corporation	Table 3	Total market capitalization by country.
U.S. mutual funds	Morningstar	Table 4, Figure 3	Net asset value and number of U.S.-based mutual funds. The funds are divided in five categories by investment allocation: all funds, Asia Pacific funds, emerging market funds, Latin America funds, and world funds.
Latin American mutual funds	Morningstar and U.S. SEC	Figure 7	Aggregate values (across funds) of quarterly injections / redemptions and the net asset values of U.S.-based Latin American mutual funds. Injections reflect percentage increases in the number of the funds' outstanding shares; redemptions reflect decreases.
Short-term positions of Latin American funds	U.S. SEC	Table 5	Average short-term positions (mostly in cash) held by Latin American mutual funds. Large mutual funds are Merrill Lynch Latin America, Fidelity Latin America and Scudder Latin America. Medium mutual fund is TCW/Dean Witter Latin America Growth. Small mutual funds are BT Investment Latin America Equity, Excelsior Latin America, Govett Latin America, Ivy South America, Morgan Stanley Dean Witter Institutional Latin America, TCW Galileo Latin America Equity.

References

BIS (Bank for International Settlements). 1998. *68th Annual Report*. Basel, Switzerland.

Bekaert, Geert, and Michael Urias. 1996. "Diversification, Integration, and Emerging Market Closed-End Funds." *Journal of Finance* 51:835–69.

Borensztein, Eduardo, and Gaston Gelos. 1999. "A Panic-Prone Pack? The Behavior of Emerging Market Mutual Funds." International Monetary Fund, Washington, D.C.

Bowe, Michael, and Daniel Domuta. 1999. "Foreign Investor Behaviour and the Asian Financial Crisis." Working Paper. University of Manchester.

Brown, S., W. Goetzmann, and J. Park. 1998. "Hedge Funds and the Asian Currency Crisis of 1997." NBER Working Paper 6427. National Bureau of Economic Research, Cambridge, Mass.

Calvo, Guillermo. 1998. "Capital Market Contagion and Recession: An Explanation of the Russian Virus." Working Paper. University of Maryland, College Park.

Calvo, Guillermo, and Enrique Mendoza. 2000. "Rational Herd Behavior and the Globalization of Securities Markets." *Journal of International Economics* 51(1):79–113.

Choe, Hyuk, Kho, Bong-Chan, and René Stulz. 1999. "Do Foreign Investors Destabilize Stock Markets? The Korean Experience in 1997." *Journal of Financial Economics* 54(2):227–64.

Cumby, Robert, and Jack Glen. 1990. "Evaluating the Performance of International Mutual Funds." *Journal of Finance* 45:497–521.

Eichengreen, Barry, and Donald Mathieson. 1998. "Hedge Funds and Financial Market Dynamics." IMF, Occasional Paper No. 166.

Frankel, Jeffrey, and Sergio Schmukler. 1996. "Country Fund Discounts and the Mexican Crisis of December 1994: Did Local Residents Turn Pessimistic Before International Investors?" *Open Economies Review* 7:511–534.

———. 1998. "Crisis, Contagion, and Country Funds." In R. Glick (ed.), *Managing Capital Flows and Exchange Rates*. New York: Cambridge University Press.

———. 2000. "Country Funds and Asymmetric Information." *International Journal of Finance and Economics* 5:177–95.

Froot, K., P. O'Connell, and M. Seasholes. 2001. "The Portfolio Flows of International Investors, I." *Journal of Financial Economics* 59:151–93.

Kaminsky, Graciela, and Carmen Reinhart. 2000. "On Crises, Contagion, and Confusion." *Journal of International Economics* 51(1):145–68.

Kaminsky, Graciela, Richard Lyons, and Sergio Schmukler. 2000a, "Liquidity, Fragility, and Risk: The Behavior of Mutual Funds during Crises." Available online at www.worldbank.org/research/bios/schmukler.htm.

———. 2000b, "Managers, Investors, and Crises: Mutual Fund Strategies in Emerging Markets." World Bank Policy Research Working Paper 2399, Washington, D.C., and NBER Working Paper 7855, National Bureau of Economic Research, Cambridge, Mass.

Karolyi, Andrew. 1998. "Why Do Companies List Their Shares Abroad? A Survey on the Evidence and Its Managerial Implications." New York University Salomon Bros. Center Monograph, vol. 7, no. 1.

Kodres, L., and M. Pritsker. 1999. "A Rational Expectations Model of Financial Contagion." International Monetary Fund, Washington, D.C.

Levy Yeyati, Eduardo, and Angel Ubide. 1998. "Crises, Contagion, and the Closed-End Country Fund Puzzle." IMF Working Paper 98-143. Washington, D.C.

Marcis, R., S. West, and V. Leonard-Chambers. 1995. "Mutual Fund Shareholder Response to Market Disruptions." *Perspective* 1(1).

New York Stock Exchange. 2000. *Shareownership Study*. Available online at www.nyse.com.

Pan, Ming-Shiun, Kam Cham, and David Wright. 2001. "Divergent Expectations and the Asian Financial Crisis of 1997." *Journal of Financial Research*. Forthcoming.

Rea, J., 1996. "U.S. Emerging Market Funds: Hot Money or Stable Source of Investment Capital?" *Perspective* 2(6).

Smith, Katherine, and George Sofianos. 1997. "The Impact of an NYSE Listing on the Global Trading of Non-U.S. Stocks." NYSE Working Paper No. 97–02.

Van Rijckeghem, Caroline, and Beatrice Weder. 2000. "Financial Contagion: Spillovers through Banking Centers." International Monetary Fund, Washington, D.C.

World Bank. 2000. *Global Development Finance: Analysis and Summary Tables*. Washington, D.C.

———. 1997. *Private Capital Flows to Developing Countries*. World Bank Policy Research Series. Washington, D.C.

Introduce Your Library to

THE WORLD BANK ECONOMIC REVIEW

OXFORD
UNIVERSITY PRESS

EDITOR: François Bourguignon, *The World Bank, Paris*

WBER is the most widely read scholarly economic journal in the world and specializes in quantitative development policy analysis. For more information, visit our website at:

www.wber.oupjournals.org

If you feel that a subscription would benefit students and colleagues, photocopy this page, fill in the name and address, and Oxford University Press will send a **FREE** sample copy of ***The World Bank Economic Review*** without obligation!

Institution: ______________________________

Librarian: ______________________________

Address: ______________________________

City: __________ **State:** __________ **Zip:** __________

Country: ______________________________

0104intlib/inwber

In the Americas:
Oxford University Press
2001 Evans Road, Cary, NC 27513
Tel: (919) 677-0977 Fax: (919) 677-1714
E-mail: jnlorders@oup-usa.org

Elsewhere:
Oxford University Press
Great Clarendon Street, Oxford OX2 6DP, UK
Tel: +44 (0) 1865 267907 Fax: +44 (0) 1865 267485
E-mail: jnl.orders@oup.co.uk

Coming in the next issue of

THE WORLD BANK ECONOMIC REVIEW

Volume 15, Number 3, 2001

- Capital Account Liberalization: What Do Cross-Country Studies Tell Us?
  Barry Eichengreen
- Where Has All the Education Gone?
  Lant Pritchett
- Measuring the Dynamic Gains from Trade
  Romain Wacziarg
- Ownership and Growth
  Thorvaldur Gylfason, Tryggvi Thor Herbertsson, and Gylfi Zoega
- Infrastructure, Geographical Disadvantage, Transport Costs, and Trade
  Nuno Limão and Anthony J. Venables
- Deposit Insurance around the World: A Database
  Asli Demirgüç-Kunt and Tolga Sobaci

THE WORLD BANK
1818 H Street, NW
Washington, DC 20433, USA
World Wide Web: http://www.worldbank.org/
E-mail: wber@worldbank.org

UPC
ISBN 0-1985-0907-3