The Urbanization Process and Economic Growth: The So-What Question

Vernon Henderson

Brown University

June 2002

Abstract: There is an extensive literature on the urbanization process looking at both urbanization and urban concentration, asking whether and when there is under or over-urbanization or under or over urban concentration. Writers argue that national government policies and non-democratic institutions promote excessive concentration -- the extent to which the urban population of a country is concentrated in one or two major metropolitan areas -- except in former planned economies where migration restrictions are enforced. These literatures assume that there is an optimal level of urbanization or an optimal level of urban concentration, but no research to date has quantitatively examined the assumption and asked the basic "so-what" question-- how great are the economics losses from significant deviations from any optimal degrees of urban concentration or rates of urbanization? This paper shows that (1) there is a best degree of urban concentration, in terms of maximizing productivity growth (2) that best degree varies with the level of development and country size, and (3) over or under-concentration can be very costly in terms of productivity growth. The paper shows also that productivity growth is not strongly affected by urbanization per se. Rapid urbanization has often occurred in the face of low or negative economic growth over some decades. Moreover, urbanization is a transitory phenomenon where many countries are now fully urbanized.

June 2002

The Urbanization Process and Economic Growth: The So-What Question[1]

Vernon Henderson
Brown University

There is an enormous literature on the urbanization process that occurs with development (see Davis and Henderson (2002) for a review). There are two key aspects to the process. One is urbanization itself and the other is urban concentration, or the degree to which urban resources are concentrated in one or two large cities, as opposed to spread over many cities. Part of the interest in the urbanization process arises because urbanization and growth seem so interconnected. In any year, the simple correlation coefficient across countries between the percent urbanized in a country and, say, GDP per capita (in logs) is about 0.85. The reason is clear. Usually economic development involves the transformation of a country from a rural agricultural based economy to an industrial-service based economy (as well as releasing labor from agriculture, as labor-saving technologies are introduced). That transformation involves urbanization, as firms and workers cluster in cities to take advantage of Marshall's (1890) localized external economies of scale in manufacturing and services (Henderson (1974), Fujita and Ogawa (1982), Helsley and Strange (1990), Duranton and Puga (2002)).

Economists have tended to focus on the issue of urban concentration, rather than urbanization per se. The literature that does exist on urbanization examines rural versus urban bias in the transformation process. Governments may favor the urban-industrial sector with trade protection policies, infrastructure investments, or capital market subsidies or they may discriminate against the rural sector with agricultural price controls (Renaud (1981), O (1993)), both leading workers to migrate to cities. But there can be a bias towards inhibiting urbanization. For example, former planned economies tend to exhibit a rural bias, in the sense of discouraging rural- urban migration, but not necessarily industrial development (Ofer (1977), Fallenbuchl (1977)).

[1] Support of the World Bank through a research grant is gratefully acknowledged, as well as support of the National Science Foundation. The opinions and conclusions in this paper are those of the author and do not necessarily represent those of the World Bank or of NSF. I thank Zmarak Shalizi for many helpful discussions. James Davis excelled as a RA on the project. I benefited from advice from David Weil and Andrew Foster on aspects of the paper and in conceptualizing the problem from early discussions with Shaid Yusuf. I benefited from comments on an earlier version of the paper by Duncan Black, Jim Davis, Yannis Ioannides, Diego Puga, and William Strange. Early versions of the paper were titled "How [Urbanization and] Urban Concentration Affect Economic [Productivity] Growth". Finally two anonymous referees provided very helpful comments used in revising the paper.

The more extensive literature on the degree of urban concentration and changes in that degree which occurs as urbanization and growth proceed has a variety of strands. Countries and international policy officials worry about whether key cities are too big or too small (Renaud (1981), UN (1993) WDR (2000)) and over the years various countries such as Egypt, Brazil, Korea, Mexico, and China have pursued medium size city programs designed to forestall the growth of larger cities (Henderson (1988) and Ades and Glaeser (1995)). International agencies presume that many of the world's mega-cities are over-populated, at considerable cost to those economies. The UN (1993) asks how bad "the negative factors associated with very large cities" need to get "before [it is in the] self-interest of those in control to encourage development of alternative centers." The same report warns of "unbalanced urban hierarchies" and the crime, congestion and social inequality in mega-cities. The World Development Report (2000) has a chapter (7) on the grim life of people in mega-cities in developing countries. And the Economist in one of its special surveys has posed the question directly (7-29-95): Do the splendors of large cities outweigh their dark side?

How do we start to analyze these issues? In the economic development literature, there is the Williamson (1965) hypothesis, as adapted to an urban context (Hansen (1990)), which states that a high degree of spatial or urban concentration in the early stages of economic development is helpful. By spatially concentrating industrialization, often in coastal cities, the economy conserves on "economic infrastructure" -- physical infrastructure capital (transport and telecommunications) and managerial resources. Such spatial concentration also enhances information spillovers and knowledge accumulation at a time when the economy is "information deficient". As development proceeds, eventually deconcentration occurs for two reasons. The economy can afford to spread economic infrastructure and knowledge resources to hinterland areas. Second, the cities of initial high concentration become high cost, congested locations that are less efficient locations for producers and consumers. Indeed a number of authors find the pattern of first increasing and then decreasing urban concentration across countries as income rises (El-Shakhs (1972), Alonso (1980), Wheaton and Shishido (1981), Junius (1999), Davis and Henderson (2001)). Growth rates of the very largest cities tend to slow, while those of medium and large size cities continue unabated (WDR 2000, Chapter 6, Table 6). There is a related literature on regional convergence within countries over time (Barro and Sala-I-Martin (1991, 1992)), which makes a similar point.

Another strand of the literature argues that political institutions and policies in countries may encourage over-concentration (Renaud (1981), Henderson (1988), Ades and Glaeser (1995), Moomaw and Shatter (1996), Henderson and Becker (2000), Davis and Henderson (2001)). The idea is that, in many countries, there is a lack of a level playing field across cities with the national government favoring one or more cities over others. Such cities may be national capitals (Bangkok, Mexico City, Jakarta, or Seoul, not to mention Paris) or the seat of national elites (São Paulo). The national government may underinvest in interregional transport and telecommunications, which favors producers and investors (who may include national politicians) in the national capital over those in hinterland cities (Fujita, Krugman and Venables (1999)). Favoritism, as in Indonesia (Henderson and Kuncoro (1996) and Kaiser (1999)), can involve restrictions in capital markets, export/import markets, and licensing of production rights, all favoring firms that locate in the national capital. This allows central bureaucrats and politicians to extract rents in the allocation of loans and licenses, without competition from lower ranked bureaucrats in other locations. Favoritism can involve the allocation of local public services in favor of national capitals, where decision-makers live. These forms of favoritism disproportionately draw migrants into the national capital. While the hypothesized favoritism may apply in many settings, over-concentration doesn't always rise. In former planned economies, migration restrictions work to limit urban agglomeration (see Au and Henderson (2002) for the case of China).

Both the urbanization and urban concentration literatures describe a geographic-migration process as affected by policy. Much of the discussion in those literatures is predicated on the idea that, for a given level of urban development, there is an optimal level of urbanization or an optimal level of urban concentration, with economic losses from deviations from these levels. Or as authors such as Gallup, Sacks and Mellinger (1998) imply, urbanization per se may promote economic growth. However no research to date has attempted to directly quantitatively examine whether urbanization promotes growth or whether there are optimal degrees of urbanization or urban concentration. This paper does so. And then it examines the basic “so-what” question--

how great are the economics losses from significant deviations from any optimal degrees of urban concentration or urbanization?

In a cross-country panel context, the paper estimates the effects of urbanization and urban concentration on productivity growth. In the growth literature there is a debate about the formulation and relevance of traditional cross-country growth regressions. As articulated by Temple (1999) and discussed towards the end of this paper, differences in technology growth rates across countries and within countries over time that may depend in part on the level of each country's economic development do not readily fit in the structure of a traditional cross-country growth framework. Given this debate, I adopt the "cleaner" approach of directly estimating the determinants of national productivity growth (e.g., Behabib and Spiegel (1994), Hall and Jones (1999)).

The paper shows that productivity growth is not strongly affected by urbanization per se, but it is strongly affected by the degree of urban concentration. On concentration, the findings are (1) there is a best degree of urban concentration, in terms of maximizing productivity growth, (2) that best degree varies with the level of development and country size, and (3) over or under-concentration can be very costly in terms of productivity growth. The strength and consistency of the findings are of some surprise-the priors were that the data are too poor and the issues too subtle to show up in cross-country productivity studies. But that is not the case.

The paper finds little support for the idea that urbanization per se drives growth. Urbanization is a "by-product" of the move out of agriculture and the effective development of a modern manufacturing sector, as economic development proceeds, rather than a growth stimulus per se. Nevertheless it may be true that, for countries still in the urbanizing stage, for any income there may arguably be a best degree of urbanization which promotes productivity growth. However, quantifying such a relationship proves difficult econometrically. Part of the problem lies with the facts: in the data, rapid urbanization has often occurred in the face of low or negative economic growth over some decades (sub-Saharan Africa). Another part of the problem is that the definition of who is urban varies substantially from region to region. In addition, the exercise is of increasingly limited relevance. Urbanization is a transitory phenomenon so that, in 1990, only about 50% of my base sample of countries are still actively urbanizing. On the other hand, I will argue that urban concentration (a ratio) is fairly well measured across countries and its role persists even as medium income countries become fully urbanized.

The Effects of Urban Concentration on Growth Development

In this section I examine the effect of urban concentration on productivity growth. I start with urban concentration, or primacy, because that examination yields the key results. The examination also develops the methodology that is then applied to the examination of the effect of urbanization on growth. The first issue is how to measure urban concentration. There are three measures that people use. First, Wheaton and Shishido (1981) and Henderson (1988) use the standard Hirschman-Herfindahl index of concentration which in an urban context is the sum of squared shares of every city in a country in national urban population. Second Rosen and Resnick (1981) use the Pareto parameter looking at the distribution of city sizes within a country, which measures how quickly size declines as we move from top to bottom in the size distribution, or the overall degree of disparity in the size distribution. In these papers, both measures were constructed for just one year for a limited sample of mostly larger countries in the world; they are not available for a larger group of countries over the time span that we look at, 1960-1995.

The third measure is what is available over the years for many countries and what is commonly utilized in the literature -- urban primacy (e.g., Mutlu (1989), Ades and Glaeser (1995), (Junius (1999)). Primacy is measured typically by the share of the largest metro area in national urban population. Is urban primacy as measured by the share of the largest city in national urban population a reasonable measure? Because such shares are typically very large, primary measures tend to be closely correlated with Hirschman-Herfindahl indices. Since Hirschman-Herfindahl indices contain squared shares, they are dominated by the largest share if that is a high number (e.g., over 0.25). Average primacy in our sample, over countries and years is .31. Similarly Rosen and Resnick (1981) show a high degree of (negative) correlation between primacy and estimated Pareto parameters. The idea of close correlation is supported conceptually, given evidence on Zipf's Law (Gabaix

(1999)) in estimation of the Pareto parameter where the coefficient tends to -1. Then, within countries when we rank cities from largest (rank 1) to smallest, rank times population size is approximately the same constant for all cities, at least among cities above a certain minimum size criterion. If this is the case, the size of the largest city in the country delineates all other city sizes and is sufficient information to calculate any comparative index of national urban concentration.

It is important to note that, although primacy represents the share of the largest city in national urban population, I am not trying to talk about optimal city size per se (even though taken literally, later, optimal primacy results for a country could be translated into results on the optimal size (with appropriate error bands) of the largest city). The issue in this paper is national urban concentration or the extent typically to which national urban resources are relatively spatially concentrated in mega-coastal cities versus spread across a variety of cities stretching into the hinterlands. So the primacy measure is treated as sufficient information to measure overall national urban concentration, or the relative resource allocation between big and small cities. For that reason concentration is not measured by the absolute size of the largest city; we want a relative measure. Questions of optimal city size per se are better analyzed using individual city data, where optimal city size varies with local industrial-service composition and effective level of technology in a complex fashion (see Au and Henderson (2002) on China).

The next issue is how to estimate the effect of urbanization processes on growth. We look directly at the effect on growth and levels of real GDP per worker. Output in an economy is produced according to aggregate Cobb-Douglas function of the form

$$Y_i(t) = K_i(t)^a \, (A_i(t) \, N_i(t))^{1-a} \tag{1}$$

where $K_i(t)$ is capital in country i in time t and $N_i(t)$ is employment. Rearranging (1), taking logarithms, and differencing we have

$$\begin{aligned} &\ln (Y_i(t)/N_i(t)) - \ln (Y_i(t-1)/N_i(t-1)) = a \, [\ln (K_i(t)/N_i(t)) \\ &- \ln (K_i(t-1)/N_i(t-1))] + (1-a) \, [\ln A_i(t) - \ln A_i(t-1)]. \end{aligned} \tag{2}$$

Eq. (2) states that worker productivity growth is a function of changes in capital per worker and in technology. I will estimate this equation as structured, directly estimating a, after specifying the arguments determining

$\ln A_i(t) - \ln A_i(t-1)$. Given the Cobb-Douglas form in (1), worker productivity growth due to changes in technology is also equivalent to total factor productivity growth. Changes in technology, $\ln A_i(t) - \ln A_i(t-1)$, are modeled in eq. (3) below as a function of base period characteristics such as (i) education of the labor force, which affects the ability to adopt new technologies (Nelson and Phelps (1966), Grossman and Helpman (1991), and Behabib and Spiegel (1994), Barro and Sala-i-Martin (1995), Durlauf and Quah (1998)), and (ii) internal country considerations affecting growth in efficiency, such as openness (again, the introduction of new technologies) or urbanization or primacy. Glaeser, Kallal, Scheinkman, and Shleifer (1992) formulate productivity growth within cities exactly as in (2) and (3). We will also look at a formulation where these variables affect the level of technology $\ln A_i(t)$, as opposed to its growth. Of course if there are growth effects, in some sense there must be level effects since levels are an accumulation of growth. However econometrically, as we will see, growth effects are much easier to quantify than level effects.

The key question is why urban concentration affects productivity growth. Losses from excessive or deficient primacy in static urban models come from GDP losses from resource misallocation, where, for example, under excessive primacy where urban development is concentrated in just one or two primate cities, these cities are subject to exhausted scale economies, excessive congestion, and excessive per capita infrastructure costs, while smaller cities have unexploited scale economies and often deficient capital investment (e.g., Tolley, Gardner, and Graves (1979), Fujita (1989), Henderson and Becker (2000), and Au and Henderson (2000)). In a dynamic context, the issue is rephrased and the influences somewhat differently articulated. In Black and Henderson (1999) building on Lucas (1988), in an endogenous growth model of a system of cities, city size affects positively the degree of local information spillovers, which interactively affects local knowledge accumulation, promoting productivity growth. However cities of excessive size draw resources away from investment and innovation in productive activity to try to maintain quality of life in a congested local environment. For four developing countries, Richardson (1987) documents the exceedingly high per capita infrastructure and housing investment costs of people residing mega cities, drawing investment out of productive and innovative uses. He finds a minimum of a threefold differential in per capita investment costs of locating a family in a mega versus small city.

From the urban literature, there are promising micro-foundations for these ideas in Duranton and Puga (2002). In that paper, primate cities are urban areas of experimentation, in deriving appropriate product designs. Relatively under-sized primate cities result in environments that have too little experimentation, affecting productivity nationally. Relatively over-sized primate cities have people devoting excessive amounts of time to commuting and other "wasteful" activities, drawing resources away from experimental activity. In principle, one could adapt the Duranton and Puga dynamic model to a growth context where under-concentration in an economy results in lower knowledge accumulation due to lack of experimentation and over-concentration siphons resources away from experimental activity, similarly inhibiting productivity growth. Then primacy affects growth in a continuous non-linear fashion. But in this context, given the Williamson (1965) hypothesis, we would expect the effect of urban concentration to depend on a country's level of development, representing national scarcity of knowledge accumulation and economic infrastructure. All these statements cry out for a comprehensive growth model that captures these specific considerations, but that is simply beyond the scope of this paper. Whatever the precise model, the empirics with cross-country data are going to come down to asking the so-what question -- to what extent does primacy affect growth?

To determine this, I hypothesize that productivity growth in eq. (2) is a function of primacy, national scale variables, education, and base period output per worker, representing the level of development, or

$$\begin{aligned} &\ln A_i(t) - \ln A_i(t-1) = f\ (primacy_i(t-1), national\ scale_i(t-1), \\ &\ln\ (Y(t-1)/N(t-1)), education_i(t-1)) + d_t + m_i + e_{it} \end{aligned} \quad (3)$$

The error structure discussed below consists of d_t common shocks across all nations, m_i a country fixed effect where geography, culture, and institutions affect productivity growth, and e_{it} time varying error term. Here I discuss the variables in (3) and specific functional forms for $f(\cdot)$. Education is always included as a basic determinant of productivity growth in the literature as noted above. The key is to establish that primacy affects productivity growth, and that, in any context, there is a best degree of primacy and deviations from that degree significantly reduce productivity growth. For the growth formulation in eq. (3), the “best degree” of primacy will be defined as one that maximizes productivity growth, other things being equal. Uncovering a best degree of primacy indicates that underlying policies that promote too little or too much primacy detract from productivity growth. To have a best degree of primacy, apart from the education term, the $f(\cdot)$ function in (3) will be specified as a quadratic in primacy. For $a_0\ primacy + b_0\ primacy^2$ where $a_0 > 0$ and $b_0 < 0$, the best degree of primacy is the peak point $-a_0/(2b_0)$.

The literature suggests that optimal primacy ought to also vary with the level of development and with national scale variables. First, we would expect the best level of primacy to decrease with both national population and land. Any city's share of national urban population should decline as national urban population grows and more cities form. For that reason we measure country population scale by urban population rather than total population. It should also decline as national geographic size increases and national resources are more spatially dispersed. Second, there is the issue of whether this best level of spatial concentration varies with output per worker, as suggested by Williamson (1965). Under Williamson high spatial concentration at the earliest stages of development is important but as development proceeds deconcentration occurs. We thus hypothesize that optimal primacy declines as output per worker rises. In summary we give the $f(\cdot)$ function in (3) the form

$$(a_0 + a_1 \ln(\textit{national scale}) + a_2 \ln(Y/N))\ primacy + b_0\ primacy^2 + b_1 \text{ education} \tag{4}$$

The working hypotheses would be that (i) the collection of terms multiplying *primacy* is positive, while $b_0 < 0$; and (ii) $a_1 < 0$ and $a_2 < 0$ so that the best degree of primacy declines as output per worker or national scale increase, where best primacy is given by

$$primacy^* = -(a_0 + a_1 \ln(\textit{national scale}) + a_2 \ln(Y/N))/(2b_0) \tag{5}$$

National scale includes both urban population and land size measures, but it will turn out that scale effects are not always statistically significant. The paper will focus on the specification in (4), first without national scale variables (just primacy and $\ln(Y/N)$ variables) and then with scale variables. In either case, however eq. (4) doesn't test a more subtle version of Williamson where best primacy might first increase from very low output per worker levels, peak, and then decline as output per worker rises. That would require some quadratic form to the output per worker term in (4) interacted with primacy variables. I will explore such a specification later.

Turning to empirical implementation, estimation of a an output growth equation as in (2) requires information on capital stock; I use the Dhareshwar and Nehru (1993) data for capital and their corresponding output per worker measure, as discussed in the Appendix. While sample coverage is limited in instrumental variable models to 69 or 70 countries, this is the cleanest formulation that examines productivity growth. In the formulation, researchers often hesitate to estimate the capital share coefficient and instead assign it a value and take the term to the LHS, so the LHS is TFP or TFP growth in a differenced form (e.g., Hall and Jones (1998). While the Dhareshwar and Nehru capital stock measures are carefully done, they presumably suffer from measurement error. Under instrumental variables estimation (see later) which can deal with (time uncorrelated) measurement error I estimate the production relationship in eq. (2) directly getting good results on the capital coefficient. But I check in a section on robustness and show that primacy results are not affected by doing it the other way. In the robustness section, I will also show that omitting capital stock measures entirely and proxying changes in capital by investment rates and GDP levels in eq. (2) will yield similar results on primacy effects (see especially fn. 10 below).

Data and Error Structure. The data come from a variety of sources listed in the Appendix including the Penn World Tables, Barro and Lee, Dhareshwar and Nehru, World Bank and the United Nations. Generally they cover the period 1960-1995 in five-year intervals, although the Dareshwar-Nehru data only go to 1990 and cover about 70 countries for which I have also primacy and education measures. The five-year intervals centered around the census and mid-census time periods in most countries represent the best times to get accurate metro area population measures. The Dareshwar-Nehru capital stock measure is based on perpetual inventory methods and is in local currency units. I also use their output per worker measure in constant local currency units. Given (see next) the estimating eq. (2) will be differenced to remove the country fixed effect we are examining internal country productivity growth with consistent LHS and RHS numbers. To have results in common units, all figures are converted to PPP numbers for the 1987 exchange rate and PPP conversion.

In estimation of eq. (2) and (3), the error structure is critical. While variables such as primacy, output per worker, and urbanization are correlated, we want to identify the "causal" effect of primacy on productivity growth. In eq. (3), as noted earlier, the d_t are time shocks/trends across countries. The m_i are country fixed effects representing unobserved country time invariant factors such as geography and culture. These will affect both growth and covariates, a basic problem in identifying primacy effects. The contemporaneous e_{it} shocks that affect growth from $(t-1)$ to t, such as internal country innovations and changes in political or legal regimes, I assume are exogenous to predetermined values of covariates in $(t-2)$. But e_{it} do affect covariates in t and even potentially $(t-1)$, such as primacy as determined by migration to the dominant metro area in the country; that is, covariates are not strictly exogenous, which becomes relevant in trying to account for fixed effects. To deal with the fixed effect m_i, I first difference eq. (2) and (3) to eliminate it, yielding a set of equation-years, where

$[\ln (Y_i(t)/N_i(t))$ - $\ln (Y_i(t-1)/N_i(t-1))]$ - $[\ln(Y_i(t-1)/N_i(t-1))$ - $\ln(Y_i(t-2)/N_i(t-2))]$ is a function of covariates $(X_i(t-1)$ - $X_i(t-2))$ and error terms, $(e_{it} - e_{it-1})$ and $(d_t - d_{t-1})$. I estimate these equation-years jointly constraining slope coefficients to be equal across years for an unbalanced panel of countries. For an equation year t minus $t-1$, I then instrument with level values of covariates from $t-3$ and $t-4$ because covariates are not exogenous.

One issue concerns the viability of instruments: why should past levels of variables be good instruments for current changes? Part of the answer lies in the underlying national economic growth process, where for example past GDP per worker is a predictor of future output per worker changes through the growth process. Another part lies in frictions in domestic capital and labor markets (Rappaport (2000)). Migration frictions relate current primacy changes to past primacy, and capital market frictions and accumulation processes relate current changes in capital stock or investment rates to past levels or rates. The set of instruments is strong in terms of first stage correlations and F-tests; and, the models readily pass specification tests on the over-identifying, as will be noted. Generally two periods of predetermined values of variables are used as instruments on the basis that these underlying complex dynamic adjustments processes are better represented with two, rather than just one, period of predetermined values of covariates. I will note basic results if just one period of predetermined values are used as instruments (see fn. 4 below).

A concern might be whether, with this differencing to eliminate m_i in estimation, there is sufficient variation in certain (differenced) covariates to identify effects. For example, for those unfamiliar with metro area population data, it might seem that primacy wouldn't vary much within a country over time. In fact this is not the case. Primacy tends to change fairly quickly, as the rural-urban allocation of resources changes with development, driven by changing relative factor returns in different sectors (where eq. (2) is a meta-production function subsuming these sector allocations). As reported in Table A2 in the Appendix, for the estimating sample, the average absolute relative change in primacy $(|\,primacy(t) - primacy(t-1)|/\,primacy(t-1))$ every five years is .18. The only variables with a low percentage relative change are ln (national urban population) and $\ln\,(Y/N)$. These enter only interactively with primacy in the estimating equation, as does (unchanging) national land area. To some extent then, heuristically, identification of interactive effects relies on the time variation in primacy, which may place practical limits on to the extent to which complex interactive effects can be identified.

The final issue concerns how to implement IV estimation; there are several possibilities. There is basic 2SLS, using two periods of lagged covariates as instruments.[2] There is 3SLS which accounts for serial correlation across equation-years (which exists since by construction $e_{it} - e_{it-1}$ in one equation year is correlated with $e_{it-1} - e_{it2}$ in the previous equation year). Finally there is what is termed GMM estimation which adjusts the 3SLS estimates to account for within year heteroskedasticity in coefficient estimation. For the last two procedures to add an extra equation-year

$(\{\ln\,[Y/N\ (1975)] - \ln\,[Y/N\ (1970)]\} - \{\ln\,[Y/N\ (1970)] - \ln\,[Y/N\ (1965)]\})$ I instrument with covariates from just $t-3$ (1960) for that one equation year, and $t-3$ and $t-4$ for all other years. In general I rely on the GMM formulation, using the DPD98 Gauss program for estimation (Arellano and Bond (1991), Caselli Esquire, and Lefort (1996)). However for key sets of results I will present all three IV results.[3] With differencing

[2] For 2SLS, I allow the functional form for the first stage regression to vary by equation year, to improve efficiency. First stage regressions are a reduced form approximation for how, in underlying dynamic adjustment processes, past level variables predict future changes. There is no reason to expect that in each period we are at the same overall point in the dynamic adjustment process or even operating under the same process, and hence want the same approximation.

[3] Technically for 3SLS results I use first step estimates in GMM from the DPD98 program. Computationally there is a slight difference between 3SLS and first step DPD98 estimates, although the two estimates are symptomatically identical.

and instrumenting there is a loss of observations. For example, if instruments are covariates from $t-3$ and $t-4$, for a country to be in an equation-year there must be five periods of consecutive data $(t$ to $t-4)$. Some countries fail to meet that standard, so country coverage under this estimation procedure is less than for OLS. In all formulations I also show baseline OLS results.

Results

I start by presenting results on the basic productivity model and then turn to the results from incorporating primacy. For the productivity model the data span 1960-1990 in five-year intervals. Given eq. (2) and (3) are differenced, in instrumental variables estimation, there are four equation years, with 1960 values needed as instruments. Country coverage is 69-81 countries depending on the variables included.

Table 1 presents baseline results. The dependent variable as in eq. (2) is listed as $\ln\,(Y(t)/N(t)) - \ln\,(Y(t-1)/N(t-1))$ which is the case under OLS. For IV estimation eq. (2) is differenced so the dependent variable as noted above is $\{\ln\,[Y(t)/N(t)]\text{ - }\ln\,[Y(t-1)/N(t-1)]\}\text{ - }\{\ln\,[Y(t-1)/N(t-1)]\text{ - }\ln\,[Y(t-2)/N(t-2)]\}$. OLS, fixed effect, and instrumental variables results are given for the case where the only argument in the $d\ln A$ function is education, as measured by the average years of secondary and above schooling in the adult population. 2SLS, 3SLS, and GMM results for instrumental variables are presented. The point estimate of the coefficient on capital, the Cobb-Douglas share parameter, starts at .54 under OLS and declines to .405 in the instrumental variables estimation under GMM, which incorporates a within period heteroskedastic structure. While that is higher than the capital coefficient assumed in the literature such as Hall and Jones' (1999) .33, the GMM estimate is similar to micro work in developing countries such as Henderson, Lee, and Lee (2001) or Korea and Jefferson and Singhe (1999) on China, where both studies find a coefficient of .37-.39 for the typical industry.

For education, the effect under GMM instrumental variable estimation is large. A one-standard deviation increase (see Appendix) in base period national education leads to a 8% growth in productivity. However this effect is not robust; other instrumental variable estimates are lower and insignificant. In the literature, education typically does not have robust effects (Temple (1999)). Authors such as Caselli et al. (1996) enter male and female education separately in cross country growth regressions getting respectively negative and positive coefficients. I also get that same result if male and female education are entered separately although both coefficients are insignificant. I choose to rely on the combined measure, recognizing that its effect is not robust.

As in Table 1, all GMM estimations in the paper pass Sargan tests, with results improving as the formulation incorporates primacy effects in later tables. Sargan tests, for a correctly specified model, examine the validity of the over-identifying restrictions that presume error terms are orthogonal to instruments. Second all models show first degree serial correlation (given $(e_{it} - e_{it-1})$ and $(e_{it-1} - e_{it-2})$ are correlated by construction) but no second degree serial correlation indicating that the e_{it}'s in eq. (3) themselves are not serially correlated. Third, standard errors on coefficients under GMM are always lower than under 3SLS; an issue in Arellano and Bond (1991) from Monte Carlo studies is whether GMM estimated standard errors in small samples are consistently too low. When 3SLS results for a model are not reported in a Table, I will always footnote them, so a reader may both compare coefficients and have a different estimate of standard errors.

Basic Primacy Results

With these results in mind, I now turn to the primacy variable. The raw data don't tell us much. There is a modest negative correlation between either $d\ln(Y/N)$ and $primacy$ or $dd\ln(Y/N)$ and $dprimacy$. Controls and a non-linear specification to the effect of primacy are needed to sort out what is going on.

The basic econometric results are in Table 2, columns (1) - (4) where there is a quadratic form to primacy and it is interacted with output per worker to allow best primacy to vary with output per worker. Before analyzing those results, I note that a simple linear primacy term has a negative coefficient. Second in columns (5) and (6) of Table 2, I report on a simple quadratic, to make the point that there is a best degree of primacy. In columns (5) and (6) and in all other reported results in the paper, OLS and instrumental variable results on primacy do differ. OLS tends to give lower best primacy values with less curvature to the $f(\cdot)$ function. In column 5, under OLS the best primacy value is .20, while under instrumental variables estimation (GMM) the best degree primacy has a high point estimate of .46, with strong and significant coefficients. From the best level of primacy, a one-standard deviation (.13) increase in primacy leads productivity growth to be .12 less over five years, a huge effect, albeit for a large change in primacy. However the best degree of primacy should vary with level of development, under the Williamson hypothesis.

Table 2, columns (1)-(4) contain the key results of the paper along with results in Table 3. In this specification in Table 2 there is a quadratic to primacy but now the primacy term is also interacted with output per worker; or I am estimating a function $[a_0 + a_2 \ln(Y/N)]\ primacy + b_0\ primacy^2$ where we expect $a_0 > 0$ $b_0 < 0$ and $a_2 < 0$ but $a_0 + a_2 \ln(Y/N) > 0$ for relevant Y/N. For the record I note that adding in (base period) $\ln(Y/N)$ on its own in this formulation results in a completely insignificant coefficient, with little impact on other coefficients or standard errors. More complex formulations are discussed below. Under the current formulation, we hypothesize that best primacy declines as output per worker rises.

In Table 2, columns (2)-(4), the instrumental variable results under different estimation methods on primacy variables do differ, although 3SLS and GMM results are very similar. Also at lower output per worker levels, best primacy results are quite similar. For 2SLS, 3SLS, and GMM, best primacy at $\ln(Y/N) = 8$ ($3000) has respective values of .28, .40, and .40. The 2SLS results are inefficient, not only because they don't incorporate serial correlation and heteroskedasticity, but also because they use a smaller sample size. With two periods of instruments for all years, we lose one equation-year in 2SLS estimation relative to 3SLS or GMM (where for the earliest equation year I use only one period of predetermined values of variables as instruments). For efficiency reasons, my focus is on the GMM results in column (4).[4]

In the results of choice in column (4), best primacy declines linearly with output per worker. At a low output per worker of $1100, best primacy is .48, while at middle ($8100) and high ($36000) output per worker it is respectively .32 and .21. High concentration seems very important to growth at early stages when economic

[4] In column (4), if all equation-years are limited to having just one period of predetermined values of variables as instruments, coefficients are very close to those in column (4), taking values (with standard errors) in order of .487 (.098), .052 (.029), 6.13 (2.45), -2.94 (1.83) and -.435 (.230). With a weaker first stage formulation, standard errors are larger than in column (4); the Sargan test result is the same (p-value of .678).

infrastructure is scarce and domestic knowledge accumulation is low. But then its importance declines as growth progresses.[5]

Deviations from best primacy are very costly: 0.071 growth points over 5 years or about 1.41% a year, for a one standard deviation (.15) increase [decrease] in primacy above [below] its best value. In this linear formulation of best primacy, this growth loss amount doesn't change with output per worker; and it seems to be a not implausible value for such a large deviation in primacy. Note that mean primacy is .31 so a change of .15 is a 50% increase or decrease in "typical' primate city size, in contexts where the primate city dominates the urban landscape (has a high value of .31). In contexts where primacy is, say, .2, a change of .15 becomes difficult to envision. If we segment the sample into coarse output per worker or country size groups, the standard deviations of primacy decline to around .1. Then the growth loss from being .1 above best or below primacy falls to about .03 over five years, or about .6% a year. Still these losses are very large. They should answer the so-what question, indicating that urban concentration is an important issue in the growth process.

The next issue concerns how to assess whether particular countries have too little or too much primacy. Since country sizes vary and best primacy should vary substantially by country size for any output per worker level, we need to control for country scale, as in eq. (4) -- both land area and population. With country scale, results can be statistically weaker; and OLS estimation fits poorly. But the IV results under GMM for the basic model are reasonable and are presented in Table 3. There, in columns (1) and (2) for OLS and GMM, I add in terms interacting the primacy variable with output per worker and two measures of national scale -- national urban population and national land area. This gives a best primacy level as in equation (5), one that is a linear function of output per worker and national scale. For the GMM results, best primacy declines with output per worker and with national scale. At a national urban population of 22m (= 10 in natural logs, for population in thousands) in an average land area country (12.9), optimal primacy declines from .41 at output per worker of $1,100 to .19 at an output per worker of $36,300. The result is illustrated in Figure 1, where output per worker ranges from ln(1100) = 7 to ln(36300) = 10.5. As national urban population rises, best primacy changes very little, but the effect of land area is very large. For the same national urban population, the intercept at an output

[5] These results are immune to adding in variables that are commonly cited to affect growth, such as openness or government

per worker of $1100, shifts down from .41 to .17 as country size increases from the average land area (12.9) to 1 ½ standard deviations above the average (15.2), as illustrated in Figure 1.

In Table 4, results similar to Figure 1 are calculated, showing how best primacy declines with increases in national land area and output per worker. The negative point estimate at high output per worker in a large land area country (e.g., the USA) will be discussed below. The Table shows for each point estimate of best primacy, the associated standard error calculated by the delta method. In the middle ranges of size and output per worker, standard errors are very small and the fit is tight. Standard errors are much larger at the extremes such as very large land area countries with medium or higher income, or small low output per worker countries.

In terms of whether countries are above or below their best primacy levels, in 1990, 55% are above, 41% are below, and the rest next to the best level. What is more telling are the countries with large deviations. Countries in 1990 with highly excessive primacy are defined for illustration as having an actual primacy value that is (i) at least twice the best point estimate (for countries with positive point estimates) and (ii) outside the 95% error band. The list comprises usual suspects with traditionally highly centralized governments-- Argentina, Chile, Algeria, Mexico, Peru, and Thailand. In 1970 the list is those same countries plus Iran, Mozambique, and Venezuela, where all of the last three experienced sharp declines in primacy between 1970 and 1990.[6] Countries with very low primacy defined as under 50% of the point estimate of best primacy and also as outside the 95% confidence interval are Belgium, West Germany, Malaysia, Switzerland, and Netherlands. Note the first three countries are peculiar in the sense of respectively being a de facto split country, being a de jure split country in the data, and having the primate city (Singapore) defect to become its own country (which is not in the sample).

National Policies Affecting Primacy. Countries with excessive primacy as noted earlier may have excessive primacy due to institutional arrangements or due to national policies concerning trade or deficient investment in inter-regional transport infrastructure. In Davis and Henderson (2002), we examined determinants of primacy such as openness and transport infrastructure, with the latter having large effects. Both openness and road

---

consumption. Since everyone has their favorite additional consideration and I couldn't find any that strongly impact the effect of primacy. I stick with a sparse specification.

density in the productivity growth equations in Table 2 have negative and insignificant coefficients and have little impact on other coefficients. Their impact is indirect -- through their effect on primacy. For example in Table 4, consider a middle output per worker middle size country that is one standard deviation (.15) above its best primacy value of .28 (and hence also outside the 95% confidence interval for best primacy), with an actual primacy value of .43. If that country were to increase either its road density or its waterway density by one standard deviation, that would lower primacy by .02 in Davis and Henderson (2002). Ignoring the cost of implementing such a change, that would raise its productivity growth rate by.018 points over 5 years, or its percentage annual growth rate by about .35%, a noticeable effect.

I have two final notes on the results so far, before examining issues of robustness and model specification. First Figure 1 and Table 3 show the limits to econometrically modeling best points—at extreme values of variables, we can get degenerate results. At high output per worker levels best primacy becomes negative, although the 95% error bands stretch well into positive values. This problem holds in most formulations. For the specific model in column (2) in Table 3 the problem can be minimized by replacing *primacy*ln(land)* by *primacy/(ln (land))*, which switches the sign of the coefficient but forces the effect of increased land to diminish with greater scale. The results for the two models are very similar, except at the large land area extreme. Of the four countries in 1990 with negative point estimates for best primacy (Australia, Brazil, Canada, and USA), all but one (USA) become positive.[7] However the list of excessive and low primacy countries is unaffected.

Second the formulations in Table 2 impose a simple structure in terms of how best primacy varies with output per worker. This allows for the basic idea of the Williamson hypothesis that high concentration is important at low output per worker levels and then best primacy declines with development. However it doesn't allow best primacy to first increase from very low output per worker levels before peaking and then declining with further growth. I found no evidence of that but given the low time variation in $\ln(Y/N)$ (but not primacy)

[6] Missing from the list is Indonesia. Indonesia is the only country where I think the size of the metro area is seriously misrepresented. Jakarta is defined as DKI Jakarta, the jurisdictional metro area, rather than the greater metro area (Jabotabek). Today the difference is about twofold.

[7] The coefficients following the listing in Table 3 are .522, .063, -.032, -3.62, -.394, -.079, and 81.0, with the last now applying to the *primacy/ln(land)* term. All except the almost zero coefficient on the primacy term are significant.

there may be a limit to what interactive effects can be identified. First, a simple Taylor series in relevant variables doesn't yield any significant results. Second, just adding a quadratic income term in the primacy expression in equations (4) and (5) (that is, adding *primacy*ln(Y/N)*^{2}) to potentially represent a Williamson effect results in an insignificant relationship, in all specifications. Finally, I tried a quadratic formulation in output per worker within both the primacy and primacy squared terms. That is, the $f(\cdot)$ function takes the form

$$\begin{aligned}&\left(a_0 + a_1 \ln(\textit{national scale}) + a_2 \ln(Y/N) + a_3 \ln(Y/N)^2\right) \textit{primacy} \\ &+ \left(b_0 + b_1 \ln(Y/N) + b_2 \ln(Y/N)^2\right) \textit{primacy}^2\end{aligned} \qquad (6)$$

This formulation is quite general and allows for all kinds of patterns. Estimates of equation (6) are statistically weak, especially under 3SLS. [8] In the results, best primacy declines throughout any relevant output per worker range for most countries, especially medium and large size ones, just as in Figure 2. In the tails of country size and output per worker values, other patterns can hold, but not the Williamson effect in relevant output per worker ranges.

Robustness and Other Specifications

Table 2 indicates the robustness of results to estimation method. The other issue is robustness to other specifications and sample changes. The first problem concerns the capital measure. Given measurement error an alternative procedure used in the TFP literature is to assign capital a share coefficient of, say, .35 and estimate the determinants of TFP growth or

$\{\ln[Y(t)/N(t)] - .35\ \ln[K(t)/N(t)]\} - \{\ln[Y(t-1)/N(t-1)] - .35\ \ln[K(t-1)/N(t-1)]\}$, rearranging eq. (2). If we estimate that model, here, again differencing out the fixed effect and instrumenting, GMM estimates on primacy variables are given in Table 5 corresponding to the Table 2, column (4) model. (Education has a coefficient (and standard error) of .049 (.018).) The results are quite similar to those in Table 2. For low and middle output per worker ($1100 and $8100 respectively) optimal primacy under the TFP formulation is .52 and .24 respectively, compared to the production function formulation in Table 2 of .48 and .32.

---

[8] Coefficients for the function in eq. (6) are for a_0 -5.37, for a_{11} for *ln(nat. urb. pop)* -.164, for a_{12} for *ln(nat. land)* -.907, for a_2 5.58, and then in order -.381, 55.1, -14.3 and .886. All except the first are significant in step two estimates

Second we have hypothesized that primacy affects the rate of productivity growth rather than its level. A level production function approach estimates eq. (1) where the determinants of $A(t)$ are now level education, primacy and other variables in either t or $t-1$. Now it is less straightforward to interact $Y(t)/N(t)$ or even $Y(t-1)/N(t-1)$ with primacy in the $A(t)$ function. In the base specification I follow columns (5) and (6) in Table 2, where arguments of $A(t)$ are $educ(t-1)$, $primacy(t-1)$ and $primacy^2(t-1)$, to see in fact whether there is an optimal primacy value. In the model the capital and education coefficients are high (.662 and .168 respectively, both significant). In column (2) of Table 5, the quadratic term on primacy is insignificant; results for primacy variables for t vs. $t-1$ are almost identical. Point estimates give a best primacy value over 1.0 in column (2) formulation, which makes no sense. Adding in primacy $t-1$ interacted with $\ln(Y(t-1)/N(t-1))$ leaves the primacy quadratic term negative but near zero, and has optimal primacy rising with income.[9] A levels formulation doesn't give plausible results. One issue for such a formulation is that the time variation in the dependent variable $\ln[Y(t)/N(t)]$ from Table A.2 is extremely modest (unlike the time variation in $d\ln[Y(t)/N(t)]$).

Third, our productivity approach is limited in sample size to 69 countries under instrument variables estimation and the period 1965-1990 (for 270 observations) because of country coverage on capital stock measures in the Dhareshwar-Nehru data. An alternative approach used in Henderson (2000) is to specify a growth model, which expands the country coverage to 79 and the sample to 361 (adding in an equation year with growth from '90 to '95). In a growth model based on the production function in (1), we are converging to a steady state value of $Y_i/(A_i N_i)$. Growth in observed per capita income is given by (e.g., Durlauf and Quah (1998))

$$\ln(Y_i(t)/N(t) - \ln(Y_i(t-1)/N_i(t-1)) = (1-e^{-bt})[\ln(Y_i(t-1)/N_i(t-1)) - \frac{a}{1-a}\ln(\frac{s}{d+n+g}) - A_0] + g\,(t - e^{-bt}(t-1)) \qquad (7)$$

which adjust coefficients for heteroscedasticity; however their step-one t-statistics are all under one except for the *ln(nat. land)* term.

[9] The coefficients (and standard errors) on $primacy\ (t-1)$, $primacysq.\ (t-1)$ and $primacy\ (t-1) * \ln(Y(t-1)/N(t-1))$ are -3.77 (1.14), -.277 (.746) and .631 (.124). For the capital and education terms, we have .418 (.060) and .107 (.021).

where g, d, n, s and b are rates of technological change, depreciation, population growth, savings, and convergence. t is the time between t and $t-1$. The second term in the square brackets gives the steady-state value of $Y_i/(A_i N_i)$. In implementation in a panel (e.g., Caselli et al. (1996)), rather than having exogenous technological change g and population growth n, determinants of population change (fertility rate) and technological change are inserted, moving us beyond the simple neo-classical framework, as Temple (1999) and Durlauf and Quah (1998) emphasize. Technological change is now related to endogenous envolving variables (education, primacy); any steady state to which an economy might be converging keeps shifting; and the equation takes a reduced form. Having $Y_i(t-1)/N(t-1)$ also potentially affect the rate of technological change as say interacted with primacy, further removes us from the standard framework. For that reason, I utilize the conceptually clean direct productivity approach in this paper.

Nevertheless in Henderson (2000), I estimated a growth specification. Growth in output per capita is postulated to be a reduced form function of base period output per capita, fertility rate, investment rate, average years of high school and college, and primacy and national scale variables. It is also possible to interpret this formulation as another version of eq. (2) where changes in capital stock are proxied for by the investment rate and base period output per capita. For this growth formulation, I also have significant optimal primacy findings and a more subtle William effect from estimating eq. (6). Optimal primacy appears to first rise from low per capita income levels into middle income levels before declining. In general optimal primacy levels are lower than in this paper and many more countries appear to have excessive primacy. However the results are specific to using growth in income per capita, rather than in output per worker as the outcome measure. The more subtle Williamson effects in eq. (6) go away with output per worker and optimal primacy levels are more in line with this paper.[10] For a simple primacy formulation as in Table 2 column (4) the primacy results are in column (3) in

[10] For a medium size country, for values of $\ln(Y/N)$ of 7, 7.5, 8, 9, 9.5, and 10, optimal primacy takes values of .21, .33, .29, .31, .35 and .30, not a compelling pattern. The primacy terms are (-66.35 -.28 ln national urban pop.) -.71 ln (land) + 18.58 ln (Y/N) - 1.07 (ln $(Y/N))^2$) primacy + (152.9 - 36.38 $\ln(Y/N)$ +2.11 ln $(Y/N))^2$ primacy^2. All coefficients are significant in both GMM and 3SLS. This formulation also best approximates eq. (2), when the investment rate and base per output per worker are substituted in for changes in capital per worker, to avoid using capital stock data.

Table 5.[11] Interestingly the results are statistically weaker and by $\ln(Y/N)=9$, any primacy effects are negative. The direct productivity formulation in this paper has more compelling and robust results.

2. The Effect of Urbanization on Growth

Examining the effect of urbanization on productivity is difficult, in the sense of the ability to isolate meaningful results. I start by discussing three reasons for this difficulty. First, rapid urbanization in African countries in particular over the last 30 years has occurred in the face of negative and low-income growth. This in itself suggests urbanization is a result of a variety of factors related to changes in national output composition and social conditions, not a force promoting growth per se. Second urbanization is a transitory process, where with economic growth all countries eventually "fully urbanize". At some middle-income level, urbanization tops off or ceases when a country is in the 65-85% urbanized category; and almost 50% of our countries fall into a fully urbanized category by 1990. Finally, urbanization definitions vary widely across countries, making it very difficult to quantify any best degree of urbanization, since that would depend on how the country counts urban.

Focusing on the definition of urbanization for a moment, fully urbanized for Switzerland, Austria and Finland means 60-65% urbanized; for the USA it is just over 70% (with miniscule full-time employment in agriculture); and for countries like Argentina, Chile, and Brazil, fully urbanized is 80-85% urbanized. A lot of these differences depend on how low density non-agricultural populations are treated in defining urban, especially around the fringes, or ex/peri-urban areas of metropolitan areas. For example, while China is officially 30% urbanized, about 70% of its population live within "municipal" boundaries (jurisdiction of the city). The debate about whether China is de facto 40- 50% urbanized revolves in part around whom to count as urban among the lower density non-agricultural populations in the ex/peri-urban areas outside the "city proper" (China's official urbanized area of the city), but within municipal boundaries. For primacy measures, definitional issues of urban may not be such a problem since it is a ratio of the urban population of one metro area relative to the urban populations of all metro, city, town and village locations within a country, all

[11] The coefficients (and standard errors) on lagged output per worker, investment rate, fertility and education are -.158 (.046), .0069

consistently defined within a country.[12] So, for China, the proportions of ex-urban populations of its municipalities don't vary consistently by size.

With these problems in mind, I econometrically explore the relationship between growth and urbanization. As with primacy, we hypothesize that for any income level, there is a best degree of urbanization. Even if "urbanization promotes growth", presumably no one would argue that low-income countries, with high degrees of semi-subsistence farming and high illiteracy rates, should switch to being fully urbanized over night.

To examine the urbanization-growth, I use an $f(\cdot)$ function of the form

$$(a_0 + a_1 \ln(Y/N) + a_2 \ln(national\ scale))\ share\ urban + b_0\ shareurban^2 ,$$

corresponding to eq. (4) and columns (2) of Table 3, although the national scale variable is unimportant. In this formulation, or any other, there are no significant results to the $f(\cdot)$ function for the whole sample. To get any results with an optimal degree of urbanization, it is necessary to restrict the sample to potentially urbanizing countries. Here I define that as the set of countries period by period that are less than 70% urbanized; an alternative restriction is to eliminate all countries that are high-income in 1965. For this restricted sample, OLS and instrumental variable (GMM) results are reported in Table 6, columns (1) and (2). The instrumental variable results suggest (1) country size is not a factor in determining an optimal degree of urbanization and (2) that there is potentially an optimal degree of urbanization. But beyond that, the results are perverse, in the sense that the effect of output per worker growth is to reduce the "best degree" of urbanization -- that is $a_1 < 0$. (This result applies whether national scale is controlled for or not.) That is the best degree of urbanization declines, as output per worker rises, a completely implausible result.

Moreover results deteriorate when I put urbanization and primacy in the same estimating equation. For that estimation, I interact national scale with primacy but not urbanization given results in column (2) of Table 6 and column (2) of Table 3. Results are in column (3) of Table 6. While the basic best primacy patterns persist, the notion that there is a best degree of urbanization evaporates, albeit in a much more limited sample size in

(.0016), .157 (.056), and .045 (.016). Variables are defined in the Appendix.

[12] That is, the definition of who is urban or not urban is applied consistently to the primate city (numerator) and all potentially urban locations (denominator). Heuristically, if roughly the same proportion is excluded across the national population locations within a country, then even if that proportion varies across countries it won't affect the primacy measure.

estimation (requiring for any period-country urbanization< .7 and for primacy data to be available). In those instrumental variable results in column (3) of Table 6, in fact, urbanization would appear to have a negative effect on growth over most output per worker ranges. In summary, these results suggest that urbanization per se, at least as measured across countries, does not directly affect productivity growth.

Conclusion

This paper argues that urbanization represents sectoral shifts within an economy as development proceeds, but is not a growth stimulus per se. However the form that urbanization takes, or the degree of urban concentration, strongly affects productivity growth. Urban concentration is affected by national policies and institutions, reflecting the extent to which a particular city (e.g., a national capital such as Bangkok or Mexico City) is favored. For any country size and level of development, there is a best degree of urban concentration, which balances the gains from enhanced concentration such as local knowledge accumulation against the losses such as resources diverted to shoring up the quality of life in crowded mega-cities. That best degree of concentration declines with country size and level of development.

References

Ades, A.F. and E.L. Glaeser (1995), "Trade and Circuses: Explaining Urban Giants," Quarterly Journal of Economics, 110, 195-227.

Alonso, W. (1980), "Five Bell Shapes in Development," Papers of the Regional Science Association, 45, 5-16.

Arellano, M. and S.R. Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," Review of Economics Studies, 58, 227-297.

Au C.C. and V. Henderson (2002), "How Migration Restrictions Limit Agglomeration and Productivity in China", NBER Working Paper No. 8707.

Barro, R. and X. Sala-i-Martin (1991), "Convergence Across States and Regions", Brookings Papers on Economic Activity, (1), 107-182.

Barro, R. and X. Sala-i-Martin (1992), "Regional Growth and Migration: A Japan-United States Comparison", Journal of the Japanese and International Economies, 6, 312-346.

Barro, R. and X. Sara-i-Martin (1995), Economic Growth, McGraw-Hill, New York.

Benhabib, J. and M.M. Spiegel (1994), "The Role of Human Capital in Economic Development: Evidence from Cross-Country Data," Journal of Monetary Economics, 34, 143-173.

Black, D. and J.V. Henderson (1999), "A Theory of Urban Growth," Journal of Political Economy, 107, 252-284.

Caselli, F., E. Esquire, and F. Lefort (1996), "Re-opening the Convergence Debate: A New Look at Cross-Country Growth Empirics," Journal of Economic Growth, 1, 363-389.

Ciccone, P. and R. Hall (1995), "Productivity and Density of Economic Activity," American Journal Review, 86, 54-70.

Davis, J. and J.V. Henderson (2002), "Evidence on the Political Economy of the Urbanization Process," Journal of Urban Economics, forthcoming.

Dhareshwar, A. and V. Nehru (1993), "A New Database on Physical Capital Stock: Sources, Methodology and Results," Revista de Analisis Economico, 8, 37-59.

Duranton, G, and D. Puga (2002), "Nursery Cities", American Economic Review, forthcoming

Durlauf, S. and D. Quah (1998), "The New Empirics of Economic Growth," NBER Working Paper No. 6422.

El-Shakhs, S. (1972), "Development, Primacy, and Systems of Cities," Journal of Developing Areas, October, 7, 11-36.

Fallenbuchl, Z. (1977), "Internal Migration and Economic Development under Socialism: The Case of Poland," in Internal Migration, A Comparative Perspective, A. Brown and E. Neuberger (eds), Academic Press, 277-303.

Fujita, M. (1989), Urban Economic Theory, Cambridge University Press.

Fujita, M. and H. Ogawa (1982), "Multiple Equilibria and Structural Transition of Non-Monocentric Urban Configurations," Regional Science and Urban Economics, 12, 161-196.

Fujita, M., P. Krugman and A. Venables (1999), The Spatial Economy, Cambridge: MIT Press.

Gabaix, X. (1999), "Zipf's Law for Cities: An Explanation," Quarterly Journal of Economics.

Gallup, J., J. Sachs, and A. Mellinger (1999), "Geography and Economic Development," International Regional Science Review, 22, 179-232

Glaeser, E., H.D. Kallal, J. Scheinkman and A. Shleifer (1992), "Growth in Cities," Journal of Political Economy, 100, 1126-52.

Grossman, G. and E. Helpman (1991), Innovation and Growth in the Global Economy, MIT Press., Cambridge.

Hall, R.E. and C.I. Jones (1999), "Why Do Some Countries Produce So Much More Output Per Worker than Others?," Quarterly Journal of Economics, 83-116.

Hansen, N. (1990), "Impacts of Small and Intermediate-Sized Cities on Population Distribution: Issues and Responses," Regional Development Dialogue, Spring, 11, 60-76.

Helsley, R. and W. Strange (1990), "Matching and Agglomeration Economics in a System of Cities," Journal of Urban Economics, 20, 189-212.

Henderson, J.V. (1974), "The Sizes and Types of Cities," American Economic Review, 64, 640-56.

Henderson, J.V. (1988), Urban Development: Theory, Fact and Illusion, Oxford University Press.

Henderson, J.V. (2000), "The Effects of Concentration on Economic Growth," NBER Working Paper #7503.

Henderson, J.V. and A. Kuncoro (1996), "Industrial Centralization in Indonesia," World Bank Economic, 10, 513-540.

Henderson, J.V. and R. Becker (2000), "Political Economy of City Sizes and Formation," Journal of Urban Economics, 48, 453-484.

Henderson, J.V., T. Lee, and Y-J. Lee (2001), "Scale Externalities in Korea," Journal of Urban Economics, in press.

Jefferson, G. and I.J. Singh (1999), Enterprise Reform in China: Ownership, Transition and Performance, Oxford University Press.

Junius, K. (1999), “Primacy and Economic Development: Bell Shaped or Parallel Growth of Cities?” Journal of Economic Development, 24(1), 1-22.

Kaiser, K. (1999), "Pre- and Post-Liberalization Manufacturing Location in Indonesia (1975-96)," 5-27-00 mimeo, LSE.

Lucas, R.E. Jr. (1988), "On the Mechanics of Economic Development," Journal of Monetary Economics, 22, 3-42.

Moomaw, R. and A. Shatter (1996), “Urbanization and Economic Development: A Bias toward Large Cities?” Journal of Urban Economics, 40, 13-37.

Mutlu, S. (1989), “Urban Concentration and Primacy Revisited: An Analysis and Some Policy Conclusions,” Economic Development and Cultural Change, 37, 611-639.

Nelson, R. and E. Phelps (1966), "Investment in Human, Technological Diffusion, and Economic Growth," American Economic Review, 56, 69-75.

O, J.C., (1993), “Reform and Urban Bias in China” Journal of Development Economics, 29, 4, 129-148.

Ofer, G. (1977), “Economizing on Urbanization in Socialist Countries: Historical Necessity or Socialist Strategy,” in Internal Migration, A Comparative Perspective, A. Brown and E. Neuberger (eds), Academic Press, 277-303.

Rappaport, J., (2000), “Why are Population Flows so Persistent?” Federal Reserve Bank of Kansas City, mimeo

Renaud, B. (1981), National Urbanization Policy in Developing Countries, Oxford University Press.

Richardson, H. (1987), "The Costs of Urbanization: A Four-Country Comparison," Economic Development and Cultural Change, 33, 561-580.

Rosen, K. and M. Resnick (1980), “The Size Distribution of Cities: An Examination of the Pareto Law and Primacy,” Journal of Urban Economics, 8, 165-186.

Temple, J. (1999), "The New Growth Evidence," Journal of Economic Literature, 37, 112-156.

Tolley, G., J. Gardner and P. Graves (1979), Urban Growth Policy in a Market Economy, New York: Academic Press.

UN (1993), World Urbanization Prospects: The 1992 Revision, New York: United Nations.

Wheaton, W. and H. Shishido (1981), "Urban Concentration, Agglomeration Economies, and the Level of Economic Development," Economic Development and Cultural Change, 30, 17-30.

Williamson, J. (1965), "Regional Inequality and the Process of National Development," Economic Development and Cultural Change, June, 3-45.

World Bank (2000), Entering the 21st Century World Development Report 1999/2000, Oxford University Press.

Table 1. A Basic Productivity Growth Equation
(dependent variable: $\ln[Y(t)/N(t)] - \ln[Y(t-1)/N(t-1)]$)

	(1) OLS	(2) fixed effects	(3) 2SLS	(4) 3SLS	(5) GMM
$\Delta\ln$ (capital/ labor)	.538** (.048)	.510** (.054)	.470** (.232)	.460** (.121)	.405** (.069)
avg, yrs. hs & college of adults $(t-1)$	.012** (.0052)	.0038 (.019)	.025 (.089)	.056 (.038)	.071** (.019)
year effects	yes	yes	yes	yes	yes
N [countries]	482 [82]	482 [82]	231	313 [81]	313 [81]
adj R^2 (within)	.365	.362	.144		
Sargan p-value					.386

** Significant at 5% level.

Table 2. Productivity Growth With Primacy[13]
(dependent variable:
$\ln[Y(t)/N(t)] - \ln[Y(t-1)/N(t-1)]$)

	Basic Model				Primacy Alone	
	(1) **OLS**	(2) **2SLS**	(3) **3SLS**	(4) **GMM**	(5) **OLS**	(6) **GMM**
$\Delta\ln$ (capital/ labor)	.563** (.051)	.493** (.248)	.447** (.133)	.454** (.071)	.556** (.052)	.514** (.044)
avg. yrs. hs & college $(t-1)$	.022** (.0066)	.081 (.068)	.049 (.039)	.047** (.017)	.0082 (.0053)	.097** (.017)
primacy $(t-1)$	.682** (.194)	12.5** (4.52)	5.99** (3.01)	6.45** (.669)	.172* (.089)	5.09** (.798)
primacy sq. $(t-1)$	-.184 (.187)	-3.90 (3.10)	-3.67* (2.15)	-3.15** (.109)	-.431** (.109)	-5.51** (.808)
primacy $(t-1)$ * $\ln[Y(t-1)/N(t-1)]$	-.075** (.028)	-1.29** (.394)	-.386 (.254)	-.488** (.110)		
year effects	yes	yes	yes	yes	yes	yes
N [countries]	417 [69]	199	270 [70]	270 [70]	417 [70]	270 [70]
adj. R^2	.447	.411			.432	
Sargan p-value				.663		.415

** Significant at 5% level; * Significant at 10% level.

**Significant at 5% level; * Significant at 10% level.

[13] 3SLS coefficients (and standard errors) for column (6) are in order of variables listed .580 (.127), .080 (.040), 4.70 (2.22), and -5.16 (2.26).

Table 3. Productivity Growth: Primacy Effects With Country Scale[14]
(dependent variable:
$\ln[Y(t)/N(t)] - \ln[Y(t-1)/N(t-1)]$)

	(1) OLS	(2) Inst. Vars.
Δ ln (capital/ labor)	.523** (.057)	.570** (.027)
avg. years h.s. & college of adults (t-1)	.020** (.0066	.072** (.0081)
primacy (t-1)	.557* (.303)	14.9** (1.05)
primacy sq. (t-1)	.065 (.187)	-3.26** (.571)
primacy (t-1)* $\ln[Y(t-1)/N(t-1)]$	-.094** (.031)	-.420** (.039)
primacy (t-1)* ln (nat. urb. pop (t-1))	.052** (.022)	-.064** (.028)
primacy (t-1)* ln (nat. land area)	-.020 (.017)	-.075** (.070)
year effects	yes	yes
N [countries]	411 [69]	266 [69]
adj. R^2 [within]	.457	
Sargan p-value	.818	

[14] 3SLS coefficients (and standard errors) for column (2) variables are respectively .572 (.106), .069 (.045), 13.9 (7.11), -3.77 (2.01), -.387 (.174), -.060 (.179), and -.586 (.455). Note scale effects are statistically weak.

<u>Table 4. Best Primacy Points [and Their Standard Errors]</u>

(medium population country of 22m urban residents)

		Geographic Area		
		small (ln (land) = 10.5)	**medium** (ln (land) = 12.8)	**large** (ln (land) = 15.2)
Output per Worker	**low** ($1100 PPP in 1987)	.65 [.054]	.41 [.014]	.16 [.048]
	medium ($8100)	.52 [.036]	.28 [.029]	.034 [.077]
	high ($36000)	.42 [.033]	.19 [.052]	-.062 [.099]

Table 5. Primacy Effects Under Different Formulations: IV (GMM) Results[15]

	(1) TFP	(2) Levels Formulation	(3) Cross-Country Growth Model
	$(\{\ln[Y(t)/N(t)] - .35\ln[K(t)/N(t)]\} - \{\ln[Y(t-1)/N(t-1)] - .35\ln[K(t-1)/N(t-1)]\})$	$(\ln[Y(t)/N(t)])$	$(\ln[Y(t)/N(t)] - \ln[Y(t-1)/N(t-1)])$
primacy $(t-1)$	7.49**	2.24**	3.15**
	(1.09)	(1.17)	(1.04)
primacy squared $(t-1)$	-2.88**	-1.06	-.701
	(1.42)	(1.25)	(.649)
primacy $(t-1)$* $\ln[Y(t-1)/N(t-1)]$	-.676**		-.350**
	(.138)		(.091)
other covariates and year effects	yes	yes	yes
Sargan p-value	.676	.552	.211

** Significant at 5% level.

[15] For column (1), 3SLS coefficients (and standard errors) for variables listed are 6.69 (2.95), -3.52 (2.24), and -.493 (.241). For column (2), they are 2.83 (1.80) and -2.07 (1.83). For column (3), they are 3.15 (1.04), -.701 (.649), and -.350 (.091). Note in column (3), although the notation N is used, N refers to population rather than number of workers.

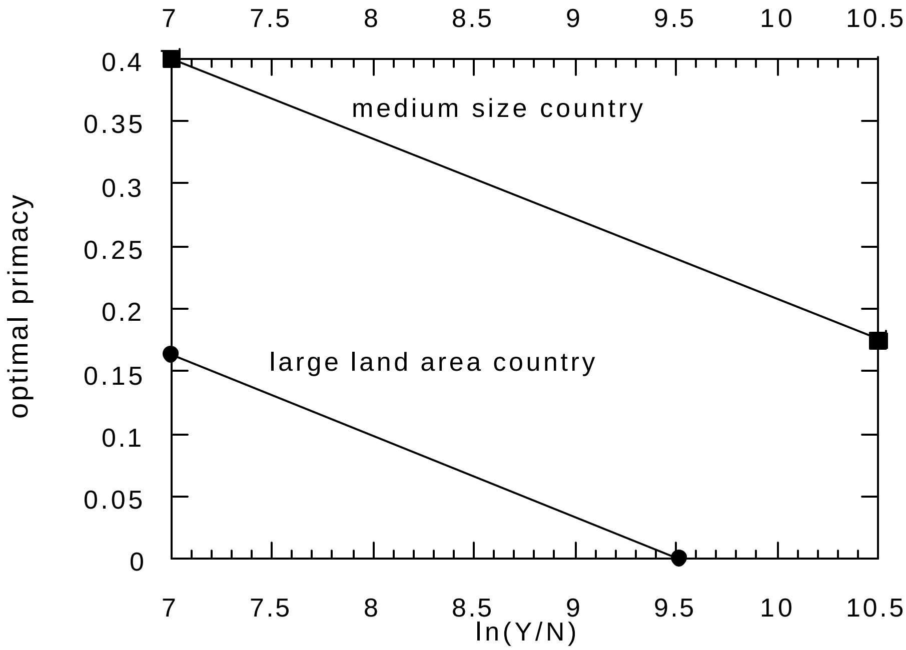
Table 6. Urbanization and Growth[16]
(dependent variable: $\ln[Y(t)/N(t)] - \ln[Y(t-1)/N(t-1)]$**)**

	OLS	GMM	GMM
Δ ln (capital/labor)	.502**	.626**	.648**
	(.610)	(.043)	(.043)
avg. years hs and college $(t-1)$	.073**	.185**	.202**
	(.585)	(.033)	(.034)
urban share $(t-1)$	.585**	5.57**	6.66**
	(.417)	(1.16)	(1.45)
* ln $[Y(t-1)/N(t-1)]$	-.042	-.554**	-.737**
	(.047)	(.084)	(.173)
urban share squared $(t-1)$	-.527*	-.922**	-.073**
	(.297)	(.367)	(.813)
urban share $(t-1)$ * ln (nat. pop $(t-1)$)	.035**	.022	
	(.018)	(.102)	
* ln (nat. land area)	-.012	.028	
	(.013)	(.081)	
primacy $(t-1)$			13.8**
			(3.14)
*ln $[Y(t-1)/N(t-1)]$			-.328*
			(.183)
primacy squared $(t-1)$			-4.42**
			(1.89)
primacy $(t-1)$ *ln (nat. pop $(t-1)$)			-.627**
			(.140)
*ln (nat. land area)			-.333
			(.249)
N [countries]	369	233 [62]	195 [52]
adj. R^2	.427		
Sargan p-value		.969	1.00

* Sample is restricted to countries in any year where urban share ≤ .70.

[16] Sample is restricted to countries in any year where urban share ≤ .70. 3 SLS coefficients (and standard errors) for column (2) variables are .624 (.106), .165 (.066), 7.67 (2.31), -.592 (.242), -1.42 (1.20), .105 (.232), and -.135 (.210). For column (3) variables are .618 (.101), .170 (.053), 8.14 (2.18), -.837 (.288). -.403 (1.22), 18.6 (5.95), -.262 (.211), -3.17 (1.91), -.676 (.201), and -.779 (.345).

Figure 1. Optimal Primacy by Output per Worker



Appendix

Data

The data are from a variety of sources.[17] The data cover 1960-95 in five-year intervals. Data on constant dollar income per capita (Chain index), output per worker, and investment share of GDP are from the Penn World Tables Mark 5.6. Data on the total fertility rate (children per women) are from the World Bank's World Development Indicators [WDI]. Openness ((exports plus imports)/GNP) is from the World Bank WDI. Data on average years of high school and college education of the adult (over 25) population are from Barro and Lee (1996). Population data on total population, urban population and primacy (population of the largest metro area/national urban population) are from the UN World Urbanization Prospects, Tables A12, A.5 and A.3.

The Nehru capital stock variable is based on Dhareshwar and Nehru (1993). They estimate constant local currency values of physical capital stocks, using perpetual inventory methods with a 4% rate of depreciation. In models using the Nehru capital stock variable, I use their output per worker measure, which is also in constant local currency units. Since this is a productivity growth model and the basic results rely on a differenced version of that equation, we are examining internal real productivity growth within a country with a consistent set of left hand side and right hand side numbers, which is appropriate. In estimation, to present graphs and results in common units, all figures were converted (i.e., scaled) to PPP numbers for the 1987 exchange rate and PPP conversion, so they are in 1987 constant PPP values. In principle, this should have no impact on estimates (although scaling does in practice affect instrumental variables estimates as in GMM).

For transport, (time invariant) kilometers of navigable waterways are from the CIA World Factbook and time varying kilometers of roads (motorways, autobahns, highways, and main national, secondary and regional roads) are from the International Road Federation supplemented by CIA data, for 1967 and 1970-1995. Both

[17] Barro, R.J.. and J.W. Lee, International Measures of Schooling Years and Schooling Quality online data, World Bank Economic Growth Research Group, Washington D.C.: World Bank, 1996;
Central Intelligence Agency (CIA), World Factbook, Washington D.C.: US Government Printing Office, various years;
Freedom House, Freedom in the World, New York: Freedom House, various years;
International Road Federation (IRF), World Road Statistics, Washington D.C.: International Road Federation, various years;
Summers, R., and A. Heston, Penn World Table Mark 5.6 version of Summers and Heston (1991) online data, Computing in the Humanities and Social Sciences (CHASS), Toronto: University of Toronto, 1995;
United Nations, World Urbanization Prospects: The 1996 Revision, United Nations Population Division, Department for Economic and Social Information and Policy Analysis, New York, 1996;
and World Bank, World Development Indicators (WDI) on CD-Rom, Washington D.C.: World Bank, 1998.

measures are divided by national land area. Given I control for national urban population and given per person road investments are much higher in rural areas, I am presuming the variation in national road densities should capture investments in interregional road systems. I also measured transport infrastructure by highway density, but the definition of this variable is much less consistent across countries.

Means and standard deviations of all variables are given in Table A below. Given the five-year intervals, for any year (e.g., 1990), for investment share, openness, and fertility rate, the X_{it_1} are the annual average rates over $t_2 - 1$ to t_1 (e.g., 1990-94 for 1995). The Penn World Tables only go to 1992. Missing data to 1994 or 1995 (including 1995 income) are filled in using the WDI numbers. For example, 1995 income per capita is projected by growing the Penn World Tables constant 1992 income by WDI numbers on annual income growth from 1992-1995. Amongst missing observations are data on Czechoslovakia, Yugoslavia, USSR, and West Germany for 1995, with 1960-1990 data defined for these countries as they existed in 1990.

Table A1. Descriptive Statistics

	Mean	**Standard deviation**
ln (output per worker, Nehru)	8.89	.994
ln (capital/labor, Nehru)	9.71	1.16
Avg. years of high school and college, pop. over 25	1.17	1.06
ln (PPP GDP per capita)	7.97	.962
PPP GDP_{pc} growth rate (5 yr.)	.110	.150
Avg. annual investment rate	18.2	8.96
ln (avg. fertility rate)	1.34	.527
Primacy	.305	.154
Urban share	.483	.242
Log (nat. urban pop.)	8.77	1.45
Openness	62.7	36.8
Log (nat. land area)	12.9	1.54

Table A1. Descriptive Statistics (Continued)

Log (nat. waterway density)	.0079	.018
Log (nat. road density)	-1.75	1.46

Table A2. Average Absolute Relative Change of Differenced Covariates

$$\left(\frac{\sum_{t=2}^{T} \sum_{i=1}^{n_t} \left| \frac{X_i(t) - X_i(t-1)}{X_i(t-1)} \right|}{\sum_{t=2}^{T} n_t} \right)$$

d ln (capital per worker)	3.0
avg. years of h.s. and college	.45
primacy	.18
urban share	.15
ln (urban population)	.053
ln (output per worker)	.030
d ln (output per worker)	3.4