Forecasting City Population Growth in Developing Countries: Illustrative Examples

Mark R. Montgomery, Deborah Balk,
Thomas Buettner, and Donghwan Kim

Paper presented to the PRIPODE workshop on

Urban Population, Development and Environment Dynamics in Developing Countries

Jointly organized by CICRED, PERN and CIESIN

With support from the APHRC, Nairobi

11-13 June 2007

Nairobi, Kenya

Forecasting City Population Growth in Developing Countries: Illustrative Examples

Mark R. Montgomery, Deborah Balk,
Thomas Buettner, and Donghwan Kim

Draft, May 2007

1 Summary of findings

This note explores techniques for estimating econometric models of city population growth rates and developing forecasts of these rates. To illustrate, we make use of the most recent available version of the United Nations Population Division's cities database, and link its city growth series to estimates and projections of total fertility rates and child mortality rates. Limiting attention to cities in developing countries, we proceed to estimate a variety of random-effect and fixed-effect models of city growth and develop growth forecasts.

The empirical results we obtain prove to be strongly supportive of this approach. This is especially true of the fixed-effect city growth models, which introduce a great number of city-specific dummy variables and yet display large and statistically significant total fertility rate and city size coefficients. The city growth rate forecasts generated by these econometric models are demographically reasonable, suggesting that over the span of the forecast from 2000 to 2045, median city growth rates in the developing world will fall from about 2.5 percent to 1.6 percent. But we would hasten to add that these are preliminary results, based on city size and growth data that are undergoing substantial revision.

Our city growth forecasts are consistent with, and indeed largely based upon, the United Nations forecasts for fertility and mortality rate decline at the national level. Yet the United Nations has never previously linked its fertility and mortality projections to its city growth rate projections. It has developed the city projections autonomously, using methods that the UN acknowledges are simplistic and in need of revision. Our approach thus unites two of the large programs of population projection in which the UN engages, and does so in a way that permits the city growth forecasts to be expressed in probabilistic terms, as strongly recommended by the U.S. National Research Council's extensive review of population projection methods (Panel on Population Projections, 2000).

As just mentioned, these are preliminary results. Much remains to be done to assess the nature and extent of measurement error in the basic city population data series. Until the city data have been thoroughly cleaned and validated, no one can make definitive pronouncements about which forecasting methods yield the best results. In addition, further work is in order on the explanatory variables that enter these models. The results developed in this note are based not on the preferred urban estimates of the total fertility and child mortality rates, but rather on national estimates of fertility and mortality. We intend to revisit questions of the robustness of the city growth estimates and their vulnerability to measurement error when the cities database has been cleaned, probable errors of measurement have been uncovered and flagged, and the full complement of urban fertility and mortality data have been made ready to be linked to the city growth series. At that point the rigorous comparison of forecast results from alternative methods can begin.

2 Overview of methods

The basic city growth model is set out as equation (1),

$$g_{i,t} = \alpha + \beta \mathrm{TFR}_t + \delta q_t + v_{i,t}. \tag{1}$$

In this equation the i subscript denotes a particular city and t is a point in time; $g_{i,t}$ is the estimated city population growth rate at that time; and the fertility and mortality components of growth are represented by the total fertility rate TFR_t and q_t, the child mortality rate.

At first glance, equation (1) might not appear to provide a useful starting-point for city growth rate estimation and projection—after all, no observable city-specific explanatory variables appear on its right-hand side. How, then, could such an equation possibly supply city-specific growth estimates? To understand our approach, recall that the United Nations city database provides a short time-series of growth observations for the cities in the database. When the disturbance term $v_{i,t}$ of equation (1) is appropriately specified, and econometric techniques for time-series, cross-sectional data are applied, informative city-specific growth estimates can be extracted from the equation even in the absence of city-specific explanatory variables.

Of course, growth models including observed city-specific explanatory variables will generally be preferred to those without such variables, provided that the city-specific observables are either fixed over time or can be forecast with reasonable confidence. To show how our approach generalizes to include observed city-specific explanatory variables, we will develop below an expanded model of city growth in which city i's population size exerts an influence on its growth rate. As we will demonstrate, the inclusion of city size in the econometric models brings our growth rate forecasts closer into line with the UN's current forecasts.

In what follows, we explore three specifications of $v_{i,t}$, the regression disturbance term. The first is a *random effects* specification in which the disturbance term is represented as a composite $v_{i,t} = u_i + \varepsilon_{i,t}$, containing one component, u_i, that is specific to city i and whose value can be estimated as $\hat{u}_i$. In this approach, u_i is assumed to be uncorrelated with the other right-hand side explanatory variables (TFR_t and q_t). Our second specification is a *fixed effect* specification in which the disturbance term also takes the composite form $v_{i,t} = u_i + \varepsilon_{i,t}$, but in which u_i is allowed to be correlated with other right-hand side variables. As in the random-effects approach, the value of u_i can be estimated (using techniques similar though not necessarily identical to those applied in the random-effects method). This specification will prove useful when city-specific endogenous explanatory variables are introduced in the model.

Our third specification is a *random-effects first-order autoregressive* specification in which the disturbance term again takes the composite form $v_{i,t} = u_i + \varepsilon_{i,t}$, but with $\varepsilon_{i,t} = \rho\varepsilon_{i,t-1} + w_{i,t}$. In this approach, the city-specific growth forecast for a future period $t+s$, given data up to period t, involves $\hat{u}_i + \hat{\rho}^s \cdot \hat{\varepsilon}_{i,t}$. In the simple random-effects and fixed-effects models, city i's growth rate is forecast to be relatively high ($\hat{u}_i > 0$) or low ($\hat{u}_i < 0$) indefinitely, whereas the autoregressive approach allows a portion of the projected city-specific growth difference to fade away with time.

The model with lagged city population size as a covariate is specified as in equation (2),

$$g_{i,t} = \alpha_0 + \alpha_1 \mathrm{P}_{i,t-1} + \beta \mathrm{TFR}_t + \delta q_t + v_{i,t}. \tag{2}$$

with $\mathrm{P}_{i,t-1}$ being the lagged value of city i's population.[1] A random-effects specification is not appropriate here, because $\mathrm{P}_{i,t-1}$ is akin to a lagged dependent variable and is therefore correlated with the u_i component of the disturbance term. Variants of the fixed-effect estimation strategy are required in this case.

Please note that whereas we hope to employ *urban* total fertility rates and child mortality rates in our future analyses of city growth, in this illustrative note we make use of *national* total fertility and child mortality rates. Urban fertility and child mortality rates are available from the World Fertility Surveys, the Demographic Health Surveys, and numerous other sources. We are in the process of

[1]The notation conveys the essential features of the model we use, but over-simplifies the situation that faces us. Annual city population data are not generally available in the UN cities database, and city population counts are recorded at unequally-spaced intervals.

assembling these urban rates, critically reviewing them, and linking them to the city population database. For the purposes of this illustrative note, we will rely on the *national-level* fertility and child mortality estimates and forecasts already available from the Population Division.

3 Overview of the UN cities database

The current version of the UN Population Division cities database supplies population counts for over 2500 cities in the developing world. The database includes records of each city's population size as reported in a census or another official estimate, together with the year of the report. In general, a city appears in this database if it is a capital or if its population has exceeded 100,000 residents.[2] In Africa, there are on average 3.5 records available on the population of each city in the database; in Asia, there are 3.2 such records; and in Latin America and the Caribbean, 5.7 records. These records refer to population counts taken as long ago as the 1940s, in a few instances, and as recently as 2003–04. City populations are not necessarily recorded at regular intervals even in one country, and the intervals between measurements vary a good deal across countries. As we will discuss below, the uneven spacing in the time dimension that is a feature of these data makes it difficult to apply conventional time-series estimation techniques.

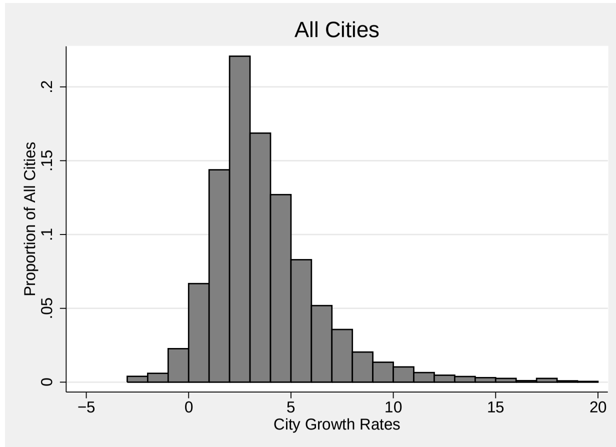
For each city, we have converted the available population data into measures of city growth rates g_{i,t_0}, with growth over the period t_0 to t_1 defined in continuous terms and estimated as $g_{i,t_0} = (\ln P_{i,t_1} - \ln P_{i,t_0})/(t_1 - t_0)$. The conversion from population counts to growth rates yields some 8,000 observations on city growth.

Figure 1 depicts the distribution of city growth rates for all cities (and time periods) in the database, and separately for the broad developing regions of Africa, Asia, and Latin America. The median growth rate recorded here is 3.20 percent and the mean is 3.76 percent. As the figure shows, there are instances of city population decline evident in these data as well as cases of rapid growth at rates of 10 percent and above.

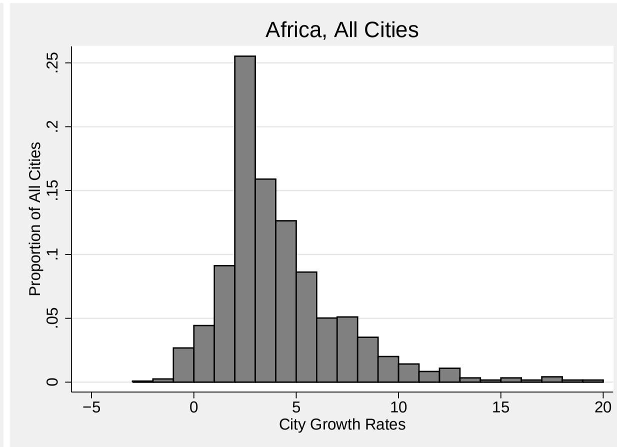
In formulating its urban projections, the UN Population Division has made use of an equation that forces city growth rates to decline as city size increases. The empirical basis for this relationship can be seen in Figure 2, represented via box plots. (These plots indicate the 25th percentile, the median, and the 75th percentile; the 'whiskers' show lower and higher percentiles. To aid in inspection of the central

[2]Once a city passes this size threshold, the Population Division endeavors to reconstruct its population trajectory in earlier years. Hence, the database contains many records of cities with populations under 100,000. However, these records have not generally been subjected to the critical scrutiny that the Population Division applies to larger cities.

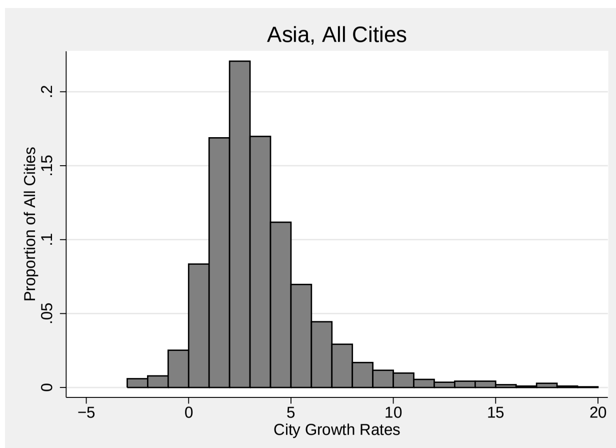
Figure 1: Distribution of city population growth rates, all developing countries and by region.



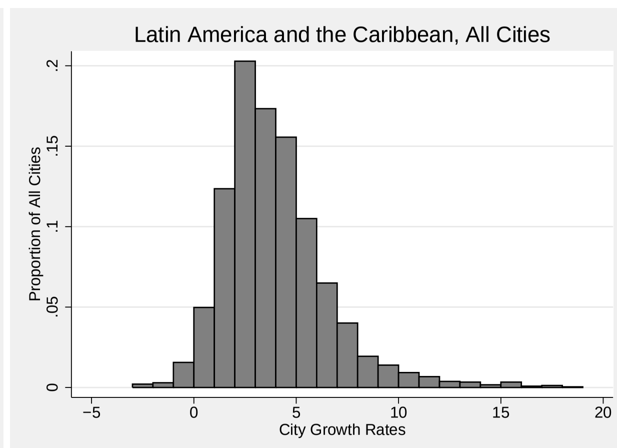
(a) All cities.



(b) African cities.



(c) Asian cities.



(d) Latin American cities.

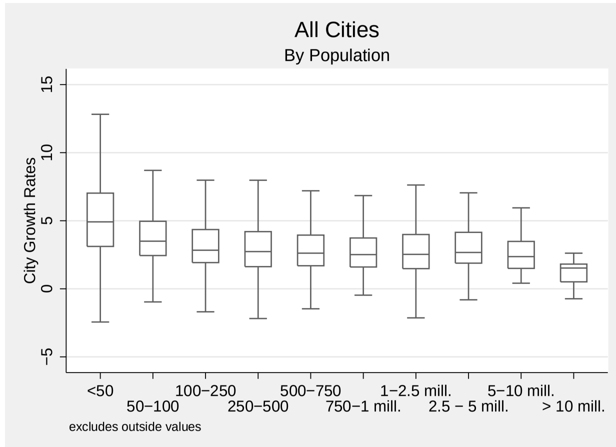
tendencies and to reduce visual clutter, the plots omit a handful of lower and upper growth outliers; these were displayed in Figure 1 above.) As the UN has emphasized in its discussion of the method, the relationship between city size and growth is weak in terms of variance explained but is, nevertheless, highly significant in statistical terms and sufficiently general to warrant consideration in the forecasts. In Africa, as the figure shows, the relationship between city size and growth is a bit irregular, being mainly apparent in the growth rate difference between the smallest and largest cities of the region. In other regions, however, the relationship appears to be more robust. The negative association between size and the rate of growth is evident among the smaller cities (under half a million population) as well as the larger (3). The growth rate differences depicted in these figures amount to a few percentage points above and below the median growth rate, and clearly there is substantial residual variation that cannot be attributed to city size as such—but there is sufficient regularity in evidence here to justify further examination.[3]

Substantial differentials emerge when we consider how city growth rates vary with national total fertility rates (TFRs). As we have noted, this linkage—shown in Figure 4 for all cities—has not previously been featured in the UN city projection methods. To judge from our descriptive figures, there is ample reason to incorporate fertility rates in city growth forecasts. Note in particular that in Latin America, the TFR–city growth gradient is especially steep. In this region, which is the most highly urbanized in the developing world, national TFRs are probably better proxies for urban TFRs than is the case in the other regions. We would expect that when we are able to link urban fertility data to the cities database, steeper TFR–city growth gradients will emerge in Africa and Asia as well.

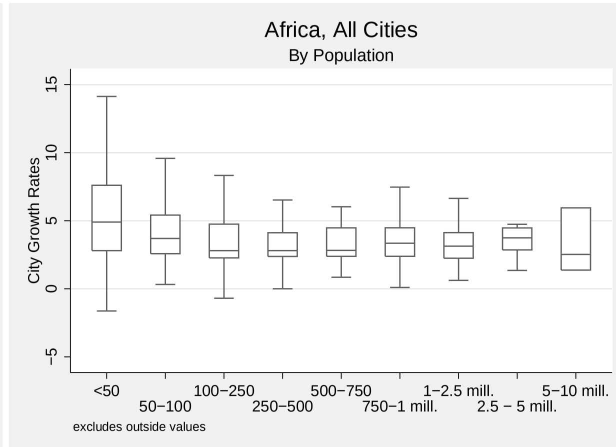
The city growth–child mortality gradient (not shown) is generally positive, owing (we believe) to the positive association between the total fertility rate and the child mortality rate. As will be seen in the next section, evidence of the expected negative association between city growth and mortality emerges when statistical controls are put in place for the level of the total fertility rate.

[3]Although we do not pursue the point further in this note, there is a suggestion of heteroskedasticity in the plots of growth rates by city size, with greater variance in growth apparent for smaller cities.

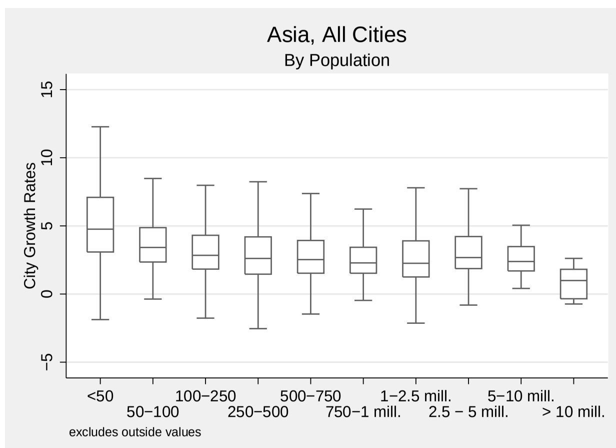
Figure 2: City population growth rates by city size, all developing countries and by region.



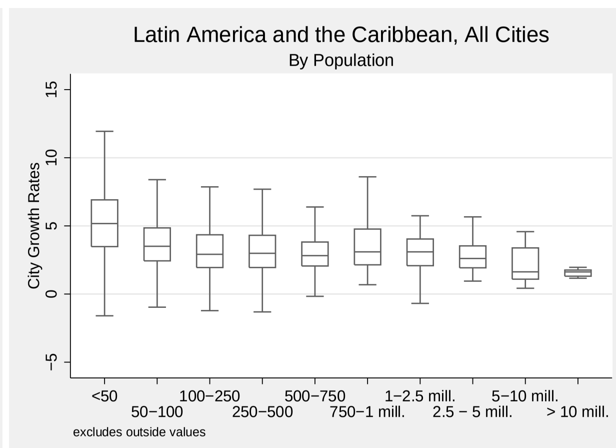
(a) All cities.



(b) African cities.

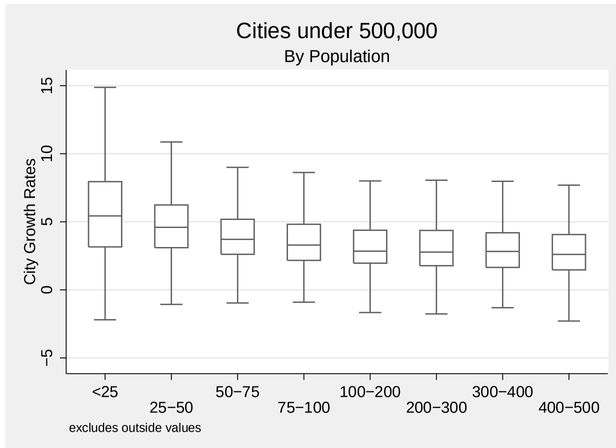


(c) Asian cities.

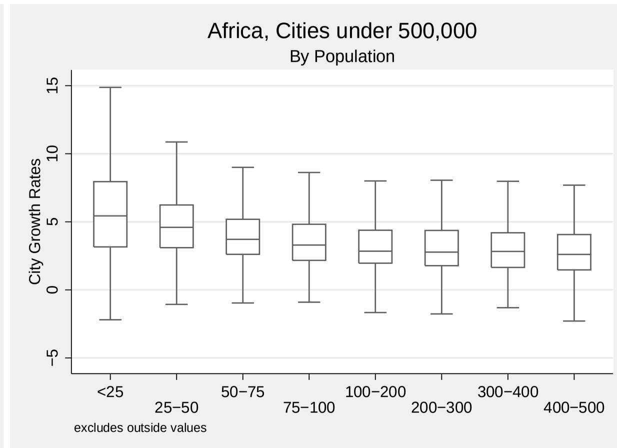


(d) Latin American cities.

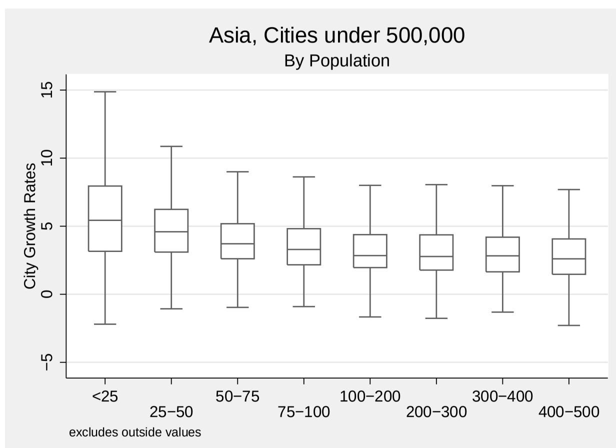
Figure 3: City population growth rates by city size for cities under 500,000 population, all developing countries and by region.



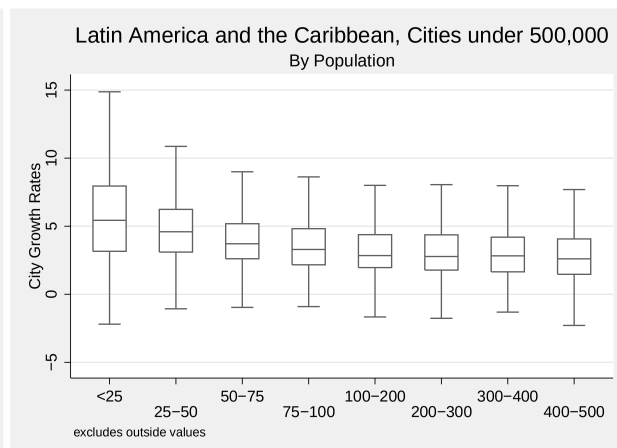
(a) All small cities.



(b) Small African cities.

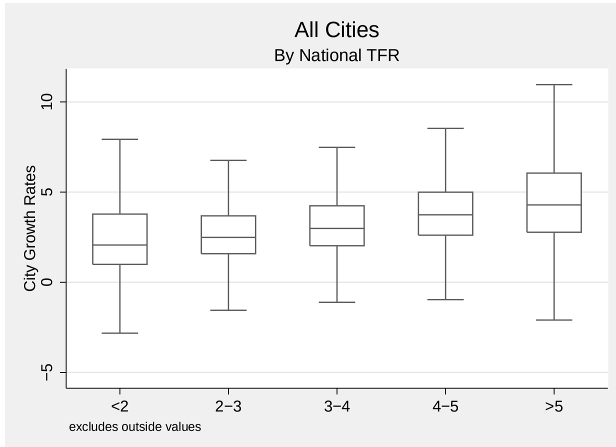


(c) Small Asian cities.

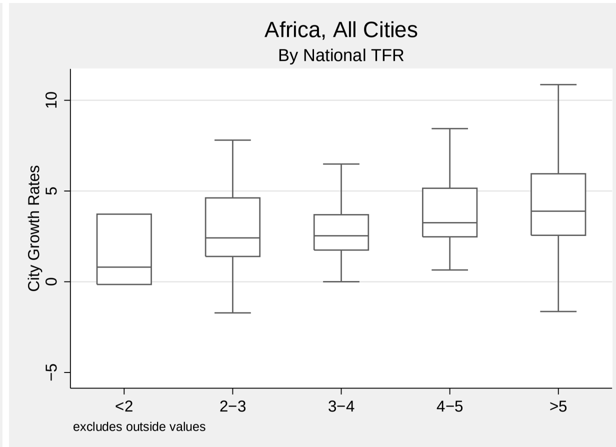


(d) Small Latin American cities.

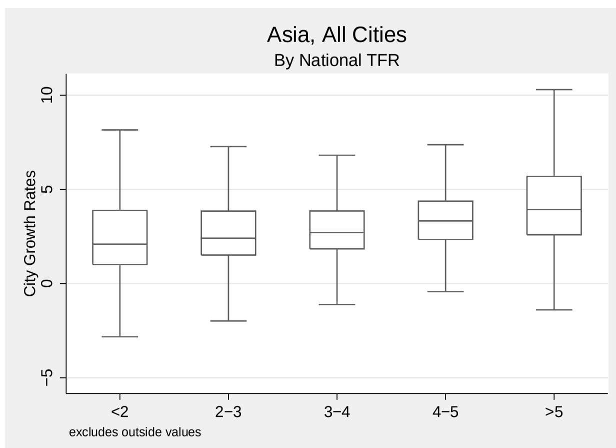
Figure 4: City population growth rates by national total fertility rate, all developing countries and by region.



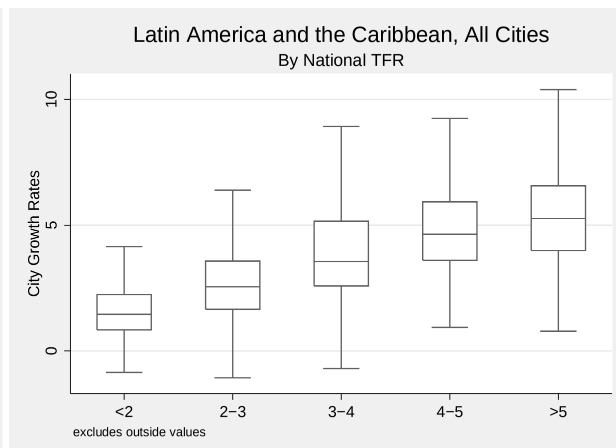
(a) All cities.



(b) African cities.



(c) Asian cities.



(d) Latin American cities.

4 Estimating city growth rate equations

In what follows we present illustrative estimates of the basic city growth rate models, using the *random-effects* and *fixed-effects* specifications described above. In each case, the disturbance term $v_{i,t}$ of the regression takes the composite form $v_{i,t} = u_i + \varepsilon_{i,t}$. In the random-effects specification, the city-specific component u_i is assumed to be uncorrelated with the right-hand side explanatory variables, whereas in the fixed-effect specification a correlation between u_i and the explanatory variables is permitted. We also present estimates of the *random effects, first-order autoregressive* model in which $v_{i,t} = u_i + \varepsilon_{i,t}$ with $\varepsilon_{i,t} = \rho\varepsilon_{i,t-1} + w_{i,t}$.

Given data to period t, forecasts of city growth rates in period $t+s$ are made as follows. For the purposes of this note, we will take the United Nations point forecasts of national total fertility rates and child mortality rates as given. (This assumption can be relaxed to allow for forecast errors in future fertility and mortality.) Then for city growth in period $t+s$, we have

$$\tilde{g}_{i,t+s} = \hat{\alpha} + \hat{\beta}TFR_{t+s} + \hat{\delta}q_{t+s} + \tilde{v}_{i,t+s},$$

in which the symbol '$\tilde{}$' denotes a forecast value and the symbol '$\hat{}$' denotes an estimated quantity based on data up to period t. In the simple random-effects and fixed-effects models, $\tilde{v}_{i,t+s} = \hat{u}_i$ although the way in which $\hat{u}_i$ is calculated generally differs between the two methods. In the random-effects autoregressive case, $\tilde{v}_{i,t+s} = \hat{u}_i + \hat{\rho}^s \cdot \hat{\varepsilon}_{i,t}$. The models we describe below were estimated in STATA, which offers the Baltagi–Wu (1999) routine to estimate random-effects autoregressive models when the data are unequally spaced in the time dimension.

Table 1 presents the basic regression models, with ordinary least squares estimates shown in the first two columns, followed by the simple random-effects and fixed-effects models. As can be seen, the coefficient on the total fertility rate is highly significant, with an increase of 1 child in the TFR implying increases in city growth rates ranging from 0.466 to 0.887 percentage points, depending on the model. Interestingly, the fixed-effects estimate of the TFR coefficient is the largest in this set of estimates. The child mortality rate (the variable is coded in terms of deaths per 1000 children) has a smaller effect on city growth, but the coefficient attains statistical significance.

Table 2 presents region-specific estimates of the random-effects and fixed-effects models. The major differences between these estimates and the estimates based on pooled data are, first, that the child mortality coefficient changes sign in one region (Africa) and loses statistical significance in another (Latin America). Second, the coefficient on the total fertility rate is evidently smaller in Africa than in the other two regions. However, the TFR coefficient is highly significant in all three regions.

Table 1: City growth regression models, developing countries, all cities. (Asymptotic Z-statistics in parentheses.)

		Model 2		
	Model 1	OLS	Random Effects	Fixed Effects
Total Fertility Rate	0.466	0.602	0.685	0.887
(Z statistic)	(26.83)	(19.97)	(20.34)	(17.68)
Child Mortality Rate		-0.004	-0.005	-0.007
		(-5.53)	(-5.54)	(-4.49)
Constant	1.878	1.757	1.464	0.802
	(24.43)	(22.01)	(16.54)	(7.25)
σ_u			1.184	1.907
			(27.71)	
σ_ε	2.667	2.662	2.394	2.381
			(107.08)	
log-likelihood[a]	-18640	-18624	-18446	-16568

[a] Likelihood calculation assumes that disturbances are normally distributed.

Table 2: Random and fixed-effect city growth models, by region. (Asymptotic Z-statistics in parentheses.)

	Africa		Asia		Latin America	
	RE	FE	RE	FE	RE	FE
Total Fertility Rate	0.375	0.297	0.646	1.082	0.675	0.944
(Z statistic)	(3.89)	(2.40)	(14.12)	(13.63)	(9.10)	(10.17)
Child Mortality Rate	0.004	0.011	-0.008	-0.014	0.003	-0.006
	(1.94)	(3.37)	(-6.57)	(-6.99)	(1.29)	(-1.87)
Constant	1.519	0.829	1.797	0.773	1.185	0.845
	(4.17)	(1.66)	(15.47)	(4.55)	(7.89)	(5.50)
σ_u	0.963	2.028	1.060	1.948	1.237	1.694
	(6.93)		(16.71)		(18.73)	
σ_ε	2.753	2.756	2.510	2.494	1.963	1.947
	(40.88)		(77.88)		(60.64)	
log-likelihood[a]	-2944	-2652	-10111	-9099	-5129	-4605

[a] Likelihood calculated on the assumption that disturbances are normally distributed.

Table 3: Random-effects city growth models with autoregressive disturbances. (Asymptotic Z-statistics in parentheses.)

	Assumed value of ρ				
	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
Total Fertility Rate	0.682	0.680	0.674	0.661	0.670
(Z statistic)	(20.31)	(20.10)	(19.63)	(18.64)	(16.30)
Child Mortality Rate	-0.005	-0.005	-0.005	-0.005	-0.005
	(-5.60)	(-5.58)	(-5.53)	(-5.38)	(-4.82)
Constant	1.488	1.500	1.523	1.572	1.536
	(16.93)	(16.84)	(16.66)	(16.31)	(13.39)
σ_u	1.138	1.112	1.044	0.794	0.000
σ_ε	2.087	1.942	1.767	1.569	1.375

Table 3 presents estimates of the random-effects autoregressive model. Here we fix the value of ρ, the autoregressive coefficient of the disturbance term, and show how the other model coefficients are influenced by its value.[4] In this case, the value chosen for ρ has little apparent effect on the other parameter estimates, leaving virtually unchanged the estimates of the TFR and child mortality coefficients, and exerting only minor influence on the estimated constant term. Nevertheless, although the *estimates* are similar, *forecasts* based on the random-effects, first-order autoregressive model would behave differently from forecasts of the simple random-effects and fixed-effects models. With $|\rho| < 1$, the forecast of the composite $\tilde{v}_{i,t+s}$ disturbance given data to period t is

$$\tilde{v}_{i,t+s} = \hat{u}_i + \rho^s \cdot \hat{\varepsilon}_{i,t}$$

in which the estimated $\hat{\varepsilon}_{i,t}$ disturbance component for period t exerts a persistent but gradually waning influence on forecasts of city growth for future periods. The higher the value of ρ, the larger this effect will be in any given future period. As s increases, however, the autoregressive effect steadily diminishes in size, leaving $\hat{u}_i$ in place as the principal city-specific effect on future growth.

In Figures 6 and 7 we compare the United Nations city growth forecasts (with an endpoint of 2015) with the forecasts derived from the simple random-effects and

[4]With unequally-spaced data, it is difficult to obtain credible estimates of ρ without recourse to special-purpose programming. The routine we have used in STATA calculates an estimate of ρ using the observations that happen to be one period (that is, one year) apart—such pairs of observations are rather rare in the cities database. Estimates of ρ for our dataset will require additional programming routines to be coded in Fortran or Matlab. Also, the last column of the table shows that highly persistent autoregressive disturbances (produced by high values of ρ) are difficult to distinguish from fully persistent u_i values.

Figure 5: United Nations city growth rate forecasts to 2015. Median growth rates shown.

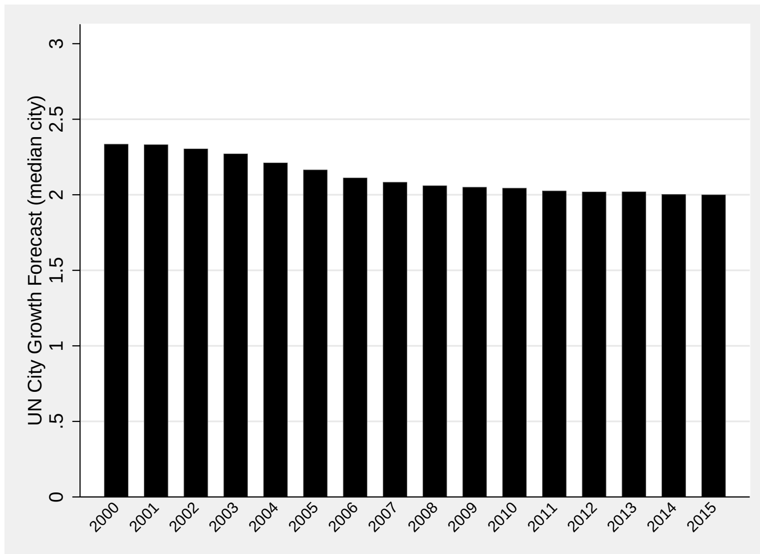
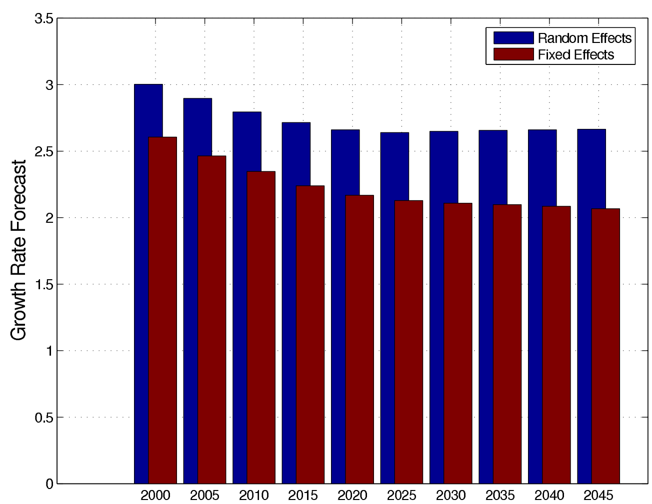


Figure 6: City growth rate forecasts from the random-effects and fixed-effects regressions, to 2045. Median growth rates shown.



fixed-effects models of Table 1, which extend to 2045 in keeping with the UN's forecasts of total fertility and child mortality rates. The UN forecasts suggest a modest decline of half a percentage point in median city growth rates from 2000 to 2015. This is about the same as the median projected growth rate decline suggested by the regression models, although the regression estimates suggest more rapid growth overall.

The difference in the level of growth rates between the UN forecasts and ours merits discussion. The key to understanding the difference is this: The UN forecasts build in a negative relationship between city size and city growth rates, whereas in the regression models of Table 1, the forecasted decline in city growth is wholly attributable to declines in future fertility and mortality. (The mortality effect by itself would imply rising rates of city growth as child death rates fall, but in our models these mortality effects are overwhelmed by the effects of falling fertility.) In the next section, we will show that the addition of a city size variable to our econometric models draws our growth rate forecasts closer into line with those of the UN.

5 Models and forecasts with lagged city population size

Having shown what can be achieved in a stripped-down growth model containing no city-specific explanatory variables, we now introduce such a variable—city population size, lagged—into the city growth rate specification and explore the implications of this expanded model. Recall that the model with city size as a covariate is specified as

$$g_{i,t} = \alpha_0 + \alpha_1 \mathrm{P}_{i,t-1} + \beta \mathrm{TFR}_t + \delta q_t + v_{i,t}.$$

with $\mathrm{P}_{i,t-1}$ being the lagged value of city i's population. We again write the disturbance term as $v_{i,t} = u_i + \varepsilon_{i,t}$, but note that a random-effects specification is not appropriate in this case. The reason is that $\mathrm{P}_{i,t-1}$ is in part the product of the growth rates for city i that were in force in earlier years, and because of this $\mathrm{P}_{i,t-1}$ can be likened to a lagged dependent variable that is correlated with the u_i component of the disturbance. A large and growing literature in econometrics explores the estimation techniques appropriate to this situation, which include variants of the fixed-effect estimation strategy and the use of instrumental variables (e.g., Arellano and Bond, 1991).

To indicate the role of city size in determining the growth rate, Table 4 presents estimates of an ordinary least squares model and a fixed-effects model with lagged city size included as a covariate along with the total fertility rate and the child mortality rate. These results reconfirm what was seen before, that the TFR has a

Table 4: Growth Regression Models with Lagged City Size, All Cities. (Asymptotic Z-statistics in parentheses)

	OLS	Fixed Effect
Total Fertility Rate	0.578	0.839
(Z statistic)	(19.05)	(16.62)
Child Mortality Rate	-0.004	-0.007
	(-5.35)	(-5.02)
Lagged City Size[a]	-0.000200	-0.000477
	(-6.24)	(-6.76)
Constant	1.917	1.256
	(22.92)	(9.73)
σ_u		1.906
σ_ε	2.656	2.371
log-likelihood	-18605	-16537

[a] Measured in thousands of residents.

strong association with city growth and the child mortality rate a weaker but (in this pooled regression) statistically significant influence. Lagged city size (measured here in terms of thousands of residents) also achieves statistical significance in these regressions.

With lagged city size in the model, forecasts of city growth must be made recursively. The growth rate forecast $\tilde{g}_{i,t}$ for period t to $t+1$ implies a forecast for city i's population size as of time $t+1$, or $\tilde{P}_{i,t+1}$, which then goes on to influence the growth rate $\tilde{g}_{i,t+1}$ forecast for the period $t+1$ to $t+2$. Using the fixed-effects estimates from Table 4, together with the estimated effect $\hat{u}_i$ for each city, we have generated such recursive forecasts for the cities in our database.

The results are shown in Figure 9, with the UN forecasts re-displayed in Figure 8 to facilitate comparison. (Note again that the UN forecasts are made to 2015, whereas our growth forecasts extend to 2045.) These forecasts are much closer to the UN's forecasts in the level of the projected growth rate, and owing to the inclusion of a negative city size feedback effect, they decline more steeply with the passage of time than did our earlier forecasts, which were made without consideration of city size. Region-specific summaries of our forecasts (not shown) suggest decreases in the median city growth rate of about 2 percentage points in Africa over the full span of the projection, and declines on the order of 0.5 percentage points for the other regions.

In short, we have demonstrated that it is a simple matter to reconcile the main features of our city growth forecasts with those of the United Nations, by introduc-

Figure 7: United Nations city growth rate forecasts to 2015. Median growth rates shown.

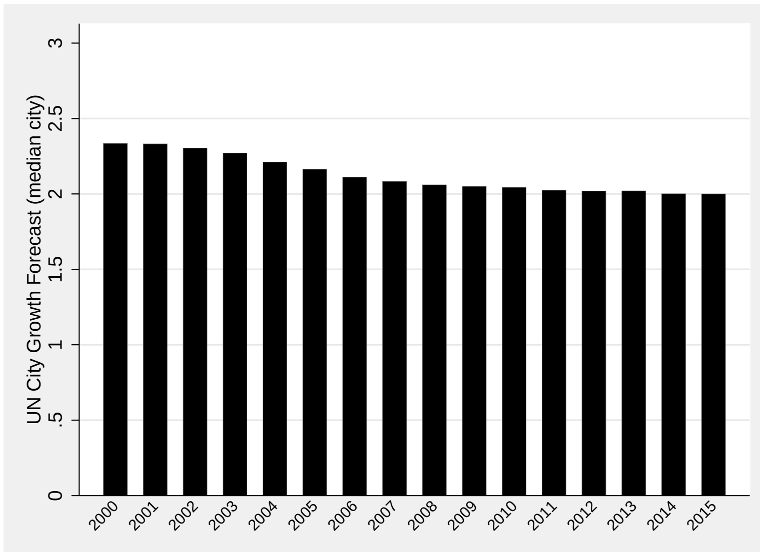
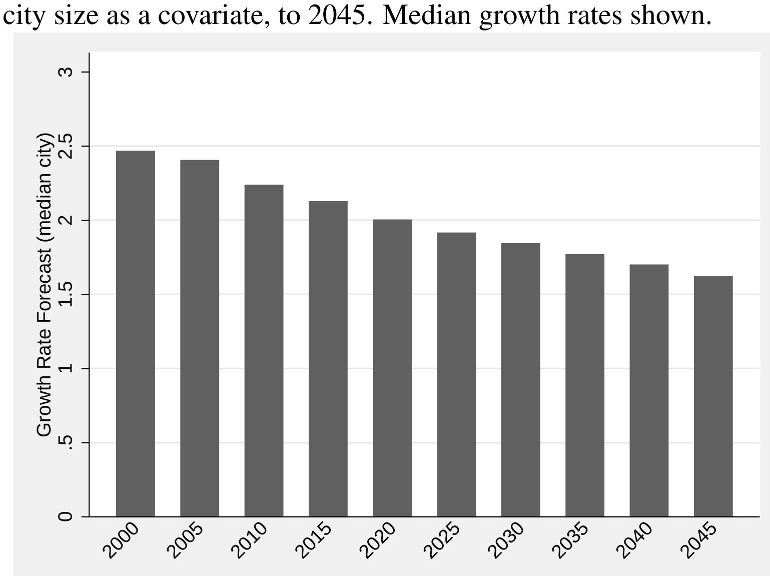


Figure 8: City growth rate forecasts from the fixed-effects regressions including lagged city size as a covariate, to 2045. Median growth rates shown.



ing lagged city size into the specifications. To be sure, it is not at all obvious that reconciliation of these forecasts should be our aim. Too much doubt has been cast on the validity of the UN forecasts to adopt them, uncritically, as the standard of comparison.

6 Next steps

The analyses and forecasts presented here are meant to clarify and illustrate some of the main methods described in the research proposal. Using national rather than the preferred urban data on total fertility rates and child mortality, we have uncovered strong evidence supporting the use of total fertility rates in econometric models of city growth and the forecasts based on these models. The findings on child mortality are less robust, but they also invite further analysis. As we have seen, statistical models incorporating lagged city size—which has been a key feature of the UN city projection methods—clearly merit consideration in our future work.

The result of the fixed-effect models are especially striking, given that such models include a great number of city-specific dummy variables (whose effects are expressed in the $\hat{u}_i$) and yet exhibit large and statistically significant TFR coefficients. But the results are, of course, highly preliminary. As already mentioned, the results shown here are based not on the preferred urban estimates of the total fertility rate and child mortality, but rather on national estimates of fertility and mortality. Much remains to be done to assess the nature and extent of measurement error in the city population series. More attention needs to be paid to regional and country differences in the coefficients. We intend to revisit questions of the robustness of the estimates, their vulnerability to measurement error, and possible region-specific variation in coefficient values when the cities database has been cleaned and vetted, probable errors of measurement have been identified, and a full set of urban fertility, mortality, and related data are ready to be linked to the city growth series.

As we have discussed elsewhere, Bayesian estimation and forecasting methods offer a means of incorporating measurement error via the specification of prior distributions. These methods hold great promise, but they too will require substantial preparatory work and additional programming. Additional work will be required to identify with confidence the spatial coordinates of the cities in the database, which are necessary inputs in the models of spatial correlation across cities. And as Voss et al. (2005) argue, when there is reason to suspect that spatial error correlation exists, models that do not take it into account will likely be biased in terms of coefficient standard errors, thus contaminating inference and causing forecast error variances to be calculated incorrectly. Even with these limits on our present efforts acknowledged, we believe that the results obtained so far lend a good deal of support to our general approach.

Literature Cited

Arellano, M. and S. Bond. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58: 277–297, 1991.

Baltagi, Badi H., and P. X. Wu. Unequally spaced panel data regressions with AR(1) disturbances. *Econometric Theory* 15: 814–823, 1999.

Panel on Population Projections (John Bongaarts and Rodolfo A. Bulatao, editors), *Beyond Six Billion: Forecasting the World's Population*. National Academy Press, Washington, DC, 2000.

Voss, Paul R., David D. Long, Roger B. Hammer, and Samantha Friedman. Child poverty rates in the U.S.: A spatial regression approach. Forthcoming, *Population Research and Policy Review*, 2005.