Bayesian Model Selection: An Application in Urban Economics

To be presented at the Southern Economics Association Meetings in New Orleans, LA – November, 1999

Lee C. Adkins

Ronald L. Moomaw

Jui-Chu Tien

Department of Economics and Legal Studies in Business
Oklahoma State University
Stillwater, OK 74078
Phone: 1-405-744-8637
e-mail: ladkins@okstate.edu

Version: November 17, 1999

Abstract

This paper examines the relationship between primacy and economic development for countries in Asia and the Americas. Varying explanations of primacy are found in demography, economic geography, political science, and sociology. To help sort out which of the theories are consistent with the data, a Bayesian procedure is used to determine the posterior probability with which each of the available variables are important determinants of urban primacy.

"It is now generally acknowledged that econometric models are 'false' and that there is no hope, or pretense, that through them 'truth' will be found."

Peter Kennedy (1998)

1 Introduction

Urban primacy refers to a country's largest one or two cities being "abnormally" large (using an adverb from Jefferson (1939)) seminal study) relative to the country's next largest cities. In discussing urbanization and development Bairoch (1988), considers both absolute and relative dimensions; he says that "Another direct consequence of the urban explosion is the great size of Third World cities. Today too great a proportion of the urban population lives in cities of excessive size . . . " (Bairoch, 1988, 511). He further argues that rapid urbanization and concentration in large cities are largely independent of economic forces and harmful to economic performance. Mills and Hamilton (1994) agree that excessive primacy and excessive urbanization can result if there are negative externalities associated with urban size. They caution, however, that positive externalities also exist, and that there is no presumption that primacy is excessive. Part of this controversy arises from disagreements about the source of urban primacy. If the concentration of population in large cities arises from the economic calculations of individual decision makers responding to economic incentives, this concentration is more likely to benefit the economy than if it arises from noneconomic forces.

Carroll (1982) finds three major classes of explanations of urban primacy in the literature–economic, political, and world systems (including international dependency and ecology approaches). The main purpose of this paper is to develop empirical evidence related to Carroll's three major classes of explanations using a Bayesian model selection technique to specify the likely independent variables in a model of urban primacy.

Classical model selection procedures use a sequence of hypothesis tests (e.g., stepwise regression) or an estimate of the model fit that weighs the tradeoff between lower sum-of-squared errors and a more parsimonious model (e.g., Akaike's AIC). Although these procedures are easy to do, the resulting estimator has statistical properties that are complicated and defy our abilities to carry out valid hypothesis tests or to construct confidence intervals. Classical procedures also handle any relevant nonsample information we may have in an awkward way (e.g., 'right signs' or bounds for one or more of the parameters).

Thanks to recent developments in the field of Bayesian computation, the ability to handle nonsample information in a statistically coherent way is improving rapidly. In this paper we use a Bayesian model selection procedure proposed by (Geweke 1994) to shed light of the specification of a model of urban primacy. The results from classical model selection and the Bayesian alternatives are compared and directions for future assessments of model adequacy or of economic hypotheses are suggested.

2 A Basic Model of Urban Primacy

The purpose in studying urban primacy is to isolate factors that, using Jefferson's (1939) adverb, cause a country's largest city to be "abnormally" large relative to other cities. No objective criterion exists, however, to measure what is "abnormally" large. One approach is to develop a measure of the deviation of the size distribution of cities from a norm, such as the rank-size rule or some other distribution derived from a stochastic model. Sheppard (1982), for instance, proposes an index that measures primacy as deviations from a rank-size relationship. Such measures are necessary in dealing with some size-distribution issues. For instance, Sheppard's purposes are to evaluate the use of a predetermined distribution as a norm and to evaluate various theories of urban size distribution. This paper, however, is not concerned with testing stochastic processes that generate particular size distributions or with central place theory.

It, instead, is concerned with the size of the largest city compared with other cities. From an economic perspective, a city is too large if it reduces economic welfare. Although a large literature exists that suggests that large cities in developing countries are too large in this sense, the evidence, except for Ades and Glaeser (1995), who find that large main cities may inhibit growth, is not systematic. Moomaw and Shatter (1993) find urban concentration works both ways: a greater share of a country's population in cities of 250,000 or more population increases growth, but the share in the largest city decreases growth. This result may suggest that it is the size of the largest city compared with other cities–primacy–rather than absolute size that adversely affects economic performance. If the relative size of large cities does not result from economic forces, it is plausible that they are too large from an economic perspective. Consequently, a failure to find economic determinants of urban primacy would place the burden of proof on people who contend that it is not excessive. Finding economic determinants, of course, does not demonstrate that primacy is not excessive; it may, however, change the presumption.

In this paper the dependent variable, primacy, is measured as the ratio of the population of the largest city to that of the second largest city, where cities are defined as urban agglomerations.

Economic theory suggests that the (economic) size of a country is negatively associated with urban primacy. This (testable) hypothesis is crucial to our approach. The tradeoff between benefits due to agglomeration (or concentration) and costs due to distance (transportation costs) implies that three components of a country's economic size–output, population, and land area–are relevant. Therefore, the model including GDP, population (POP) and arable land area (LAND), is

$$\ln(PRIMACY_t) = \beta_1 + \beta_2 \ln(POP_t) + \beta_3 \ln(LAND_t) + \beta_4 \ln(GDP_t) + \delta z_t + e_t \tag{1}$$

where δ is a vector of unknown parameters and z_t is a vector of additional explanatory variables.

An increase in size, by our definition, is a joint proportional increase in GDP, population, and land area. As the economic size of a nation increases, it enables several production sites, creating new urban centers, and, thus, reducing urban primacy. It is expected that these variables will enter any model of urban primacy with relatively high probability. The failure to find such a relationship would, we believe, discredit an economic approach to explaining urban primacy, suggesting that other explanations would be more powerful.

3 Alternative Explanations of Urban Primacy

Social scientists–economists, geographers, political scientists, and sociologists–have offered additional explanations of urban primacy, which focus on international economic relations, internal political factors, and demographic factors. Krugman (1996) has developed models that imply that primacy decreases with the openness of a national economy. In contrast, dependency theory implies that economies, particularly developing economies, that are more open to foreign trade will experience increased primacy because (dependent) trade concentrates production in the larger cities. According to Castells (1977), dependent urbanization, which implies that developing countries rely on industrialized countries for trade, investment, aid, and technology transfer, "causes a superconcentration in the urban areas" (i.e., primate cities–pp. 47-48). In the extended model, openness (dependency) is measured by the logarithm of the export/GDP ratio (LEXP).

Internal political forces are expected to result in urban primacy in developing countries, according to the theory of urban bias (Lipton 1977). A variable used by London (1987) (and others) to measure urban bias is the ratio of nonagricultural to agricultural productivity. Nonagricultural productivity is meausured using nonagricultural output as a percent of GDP divided by nonagricultural employment as a percent of total employment. Agricultural productivity is measured in the same way. This urban disparity, LURBDISP in logarithms, is assumed to result from urban bias. The assumption is that large investments (public and private) in large urban areas and neglected investment opportunities in rural areas cause increases in nonagricultural productivity relative to agricultural productivity and promote urban primacy.

Others have suggested that an urban political bias may increase primacy in some developing countries and not others, depending upon internal political forces. Benson and Faminow (1988) have suggested that rent seeking is more effective in the capital, which is often the largest city. Although Gilbert and Gugler do not use rent-seeking terminology, they argue that "In many countries . . . it is the location of government and the paraphernalia of modernization rather than industrial growth *per se* that is the principal source of urban and regional concentration" (Gilbert and Gugler (1992), p. 56).

As already mentioned, Bairoch states that excessive urbanization has resulted in an excessively large proportion of the urban population living in very large cities. Although this would suggest that the greater the percent urban the greater the primacy, Richardson (1988) **?** argue the opposite. They suggest that a higher urban percentage implies that cities other than the primate city can develop and attract population from it. They imply that primacy is largely a demographic phenomenon, resulting from small national populations and low degrees of urbanization. Because population is in the basic model, the only new variable needed to test this is LURBPOP, the logarithm of urbanized population.

Furthermore, a more educated population and large cities interact positively [Henderson (1988); Rauch (1993)]. More educated and skilled people may opt for jobs and services that are not available in smaller cities and towns. With a higher proportion of such people in the economy, the preference for living in large cities may be greater. As a result, any premium necessary to attract people to large cities will be less and large cities will be more profitable locations. Moreover, this concentration of human capital in large cities may have external effects (Rauch 1993) that increase the productivity and thus size of the largest cities. This variable is measured by LEDUC, which is the logarithm of the country's average years of education of people more than 25 years of age.

Another factor that may be related to primacy is the degree of political freedom. Strong, undemocratic political leaders may concentrate power in administrative centers, particularly the largest city, to serve the interests of the military, political, and economic elite, much as the colonial powers concentrated resources in the primate city. These undemocratic leaders have greater ability to ignore the wishes of the politically weak hinterland–smaller cities and rural areas (Ades and Glaeser 1995). We use the Gastil classification of countries in free, partially free, and not free to measure dictatorship (DICT); depending upon the category, countries are assigned the number 1, 2, or 3 (3 is not free).

The growth rate of GDP and the growth rate of GDP per capita are used to measure the effect of economic growth on urban primacy. Growth pole theory suggests that a rapidly growing economy will generate both backwash and spread effects. Consequently, other things equal, economic growth could promote primacy through a backwash effect or discourage it through a spread effect. Because we expect the spread effect to dominate, we expect that the coefficient of an economic growth variable would be negative.

A rapid population growth rate is generally associated with a rapid growth rate of the rural population, which in turn leads to rural to urban migration. Bairoch (1988) argues that rural population pressures have led to greater urban primacy. DeCola (**?**) finds a correlation between primacy and the population growth rate.

Finally, in many countries the largest city is the capital city. Public administration and government offices increase the employment in the largest city, and thus may increase the total population of the largest city. In addition, rent seeking activities may make the capital city larger than it would otherwise be (Benson and Faminow 1988). In Southeast Asia, the capital and primate city emerged in the 19th century both as the city through which resources were shipped from the interior to Europe and as the colonial administrative city. With independence these cities became the countries' capitals. The inclusion of DCAP in the model does not imply that these historical forces are unimportant. They are, however, captured by the country fixed effects. Thus, a positive coefficient would suggest that countries such as Brazil (in the sample period) and the United States (much earlier) that have created new capitals have probably decreased primacy.

It is not our purpose to fully discuss these theories and empirical studies in depth. Instead we take the variables suggested by the proponents of the various theories and include them in the variable vector, z_t in equation (1). The intent is to see if these variables are predicted to enter the model with high posterior probability, which would provide support for the underlying theories, and to decide if their addition affects or eliminates the relationship between primacy and economic factors reported in the previous section.

4 The Data

A problem arises in studies of urban primacy because unmeasurable variables, such as geography, history, institutions, and politics, which vary substantially over the cross section, vary slowly (if at all) over time. Nevertheless, the cross-sectional variation in these variables affects both agglomeration benefits and transportation costs, and thus affects the relationship between primacy and economic variables. The a priori importance of these unmeasurable variables implies that if they are omitted in cross-section analysis, the usual estimators may be seriously biased.

The confounding effects of unmeasurable variables in a cross section can be reduced or eliminated by using panel data and a fixed-effects estimator. The unmeasured variables are controlled with dummy variables for each country–fixed effects. Within country variation over time then permits the calculation of the effects of various independent variables on the dependent variable, primacy.

We use a panel consisting of data in five year increments, beginning in 1960 and extending to 1990 for 30 countries from Asia and the Americas. The countries from Asia and the Americas in the sample were chosen systematically, meeting three conditions; they (1) had a total population of two million or more in 1990, (2) were a nation-state not a city-state, eliminating Hong Kong and Singapore, and (3) were not a socialist or ex-socialist country. These criteria give 33 countries. Because of substantial data unavailability, we drop Haiti, Jamaica, and Nepal, leaving the thirty countries–11 from Asia, 17 from Latin America, and 2 from North America–listed in the Appendix.

5 Estimator

Choosing an appropriate subset of regressors to use in a linear statistical model is recurring problem in econometric analysis. In Feldstien's (1983) a useful model is not one that is 'true' or 'realistic' but one that is parsimonious, plausible, and informative. Toward that end our goal is to apply recent developments in Bayesian statistics to the problem of model specification in urban economics.

Geweke (Geweke 1994) considers the standard regression variable selection problem using a proper informative prior distribution for each parameter. The linear model is denoted

$$y = X\beta + e \qquad e \sim N(0, \sigma^2 I_T) \tag{2}$$

where y is $Tx1$ vector of observations on a random dependent variable and X is TxK matrix of T observations on K independent variables. In Geweke's setup k^* out of the K parameters each have a nonzero coefficient with prior probability 1, and there is positive probability that any combination of the remaining $K-k^*$ variables have coefficients equal to 0. In addition it is assumed that in the prior distribution all parameters are mutually independent.[1] The investigator's prior distributions for each of the coefficients and the parameter σ in equation (1) are mutually independent. With prior probability $\underline{p}_i$, $\beta_i = 0$; conditional on $\beta_i \neq 0$ the prior distribution of β_i is $N(\underline{\beta}_i, \tau_i^2)$, possibly truncated to the interval (λ_i, ν_i).

The prior distribution of σ is

$$\underline{v\sigma}^2/\sigma^2 \sim \chi^2(\underline{v}) \tag{3}$$

The prior distribution is proper and informative, but not conjugate. Geweke asserts that it is useful because it is relatively easy to elicit an informative prior of this type and its simplicity makes the computations much easier to do.

The computational procedure employed by Geweke is a Gibbs sampler with complete blocking. Computations proceed in the following way. A value for each coefficient is drawn in turn from its distribution conditional on $\beta_l (l \neq j)$ and σ is drawn conditional on β.

Algorithm

The conditional distributions are relatively simple and can be found in Geweke (1994). In this section we will merely detail the actual algorithm used to obtain our results.

1. Given an initial estimate of the parameters (we used least squares $\hat{\beta} = (X'X)^{-1}X'y$) obtain the residual:

$$z_t = y_t - \sum_{l \neq j} \hat{\beta}_l x_{il} \tag{4}$$

[1] This assumption can be weakened and is the subject of additional research.

2. Compute an estimate of an omitted coefficient

$$b = \sum_{t=1}^{T} x_{tj} z_t / \sum_{t=1}^{T} x_{tj}^2 \tag{5}$$

and its precision

$$\omega^2 = \sigma^2 / \sum_{t=1}^{T} x_{tj}^2 \tag{6}$$

3. Compute

$$\sigma_*^2 = (\omega^{-2} + \tau^{-2})^{-1} \tag{7}$$

and

$$\bar{\beta}_j = \sigma_*^2 (\omega^{-2} b + \tau^{-2} \underline{\beta}_j) \tag{8}$$

4. Compute the conditional Bayes factor in favor of $\beta_j \neq 0$, versus $\beta_j = 0$:

$$\begin{aligned} BF = & \exp[\bar{\beta}_j^2 / 2\sigma_*^2 - \underline{\beta}_j^2 / 2\tau_j^2](\sigma_* / \tau_j) \\ & \{\Phi[(\nu_j - \bar{\beta}_j)/\sigma_*] - \Phi[(\lambda_j - \bar{\beta}_j)/\sigma_*]\} \\ & \{\Phi[(\nu_j - \underline{\beta}_j)/\tau_j] - \Phi[(\lambda_j - \underline{\beta}_j)/\tau_j]\}^{-1} \end{aligned} \tag{9}$$

5. Compute the posterior probability that $\beta_j = 0$ using

$$\bar{p}_j = \frac{\underline{p}_j}{\underline{p}_j + (1 - \underline{p}_j) BF} \tag{10}$$

6. Take a random draw, u, from a uniform [0,1]. If $\bar{p}_j < u$, then draw β_j from $TN_{(\lambda_j, \nu_j)}(\bar{\beta}_j, \sigma_*^2)$. Otherwise, set $\beta_j = 0$.

7. Conditional on all β_j,

$$\left[\underline{\nu}\underline{\sigma}^2 + (y - X\beta)'(y - X\beta)\right] / \sigma^2 \sim \chi^2(\underline{\nu} + T) \tag{11}$$

The gibbs sampler works in the usual way. After an initial value for (β, σ) is drawn from the prior distribution (we started ours at the least squares estimates), the parameters $\beta_1, \beta_2, \ldots, \beta_K, \sigma$ are drawn in succession from their respective conditional posterior distributions. The posterior probability that a coefficient is zero is computed by taking the proportion of the gibbs samples in which $\beta_j = 0$. The posterior model probabilities can also be computed; however, when K is large, as it is in our example, computing posterior model probabilities for each of the 2^K possible choices and summarizing those in a meaningful way is necessarily difficult. Hence, we defer their computation until later iterations of the paper.

We also estimated the model using a more traditional approach, stepwise regression. We do so not to advocate its use, but to point out the its obvious deficiencies in model selection and serve as a contrast to the Bayesian approach. SAS's PROC STEPWISE is used (using the forward and backward options).

6 Estimation and Results

The Geweke procedure requires a vector that contains the prior probabilities with which each of the parameters enters the model, a prior mean, and a prior precision vector. In addition, inequality restrictions can be introduced.

We use two prior probability vectors. The first sets $\underline{p}_j = .5$ for each variable in the system. The other forces each of the country dummies (D_i) to enter the model (i.e., $\underline{p}_j = 0$ for $i = 1, \ldots, 30$ while the other parameters again enter the model with probability equal to .5.

The prior mean is allowed to take two values. The first is the OLS estimates from the full model. The other is formed using $b_D = (X_D' X_D)^{-1} X_D' y$ where $X_D = I_{30} \otimes j_T$, j_T is a $Tx1$ vector of ones. In effect, b_D are the country means for natural logarithm of urban primacy. The second prior mean uses b_D as the prior mean for the country dummies and 0 for the slopes, i.e., $\underline{\beta}_2' = \{b_D' | 0'\}$.

Prior precision also takes two values. The first is the square root of the OLS precision ($\sqrt{diag(X'X)/\hat{\sigma}^2}$); The second is similar in spirit to one suggested by ...... which assumes that a large change in x_j leads to a large change in y. A large change is here defined as a 1 standard deviation change in the variable. Thus, precision of each parameter is set to be σ_y / σ_{x_j}. Thus, our results consist of 8 possible scenarios. The posterior probabilities that $\beta_j = 0$ for each appears in Table 2. These results are based on a total of 40,000 Gibbs samples with the first 10,000 discarded.[2] In addition, $\underline{v} = 200$.

Inequality restrictions are also used as prior information. Economic theory suggests that the coefficients on LLAND and LGDP are non-negative while the coefficient on LPOP is non-positive; these were enforced in each scenario. The competing theories for the effects of the other variables on primacy are unable to yield unambiguous restrictions for the remaining parameters.

[2] For this preliminary version of the paper, the convergence of the Gibbs sampler was not explored.

The ordinary least squares results are shown in table 1. Many of the counrty dummies are significantly different from zero at the 5% level. Among the important economics variables, the coefficients of LGDP, LPOP are significant and have the right signs; that of LLAND has the correct sign, but is not significant. Among the other variables LLABOR, LEDUC, DCAP, DICT, are signigicantly different from zero at the 5% level whereas LURDISP is nearly so. The effects of the other variables on primacy are not statistically significant.

In Table 2 the results from a backward stepwise procedure executed in SAS 6.12 are shown. All of the insignificant continuous variables from the OLS results are dropped in addition to LEDUC. The previously insignificant LURDISP is included. Surprisingly, all 30 of the country dummies appear. None of the estimated coefficients change sign. LGDP, LPOP, LURDISP, and DICT coefficients get a bit smaller in magnitude whereas the other increase to a small degree.

The forward stepwise results appear in Table 3. In this instance a few of the dummy variables do no make it into the model. The results is a large change in the magnitudes of the country dummies. On the other hand, each of the significant continuous variables from OLS are included in the model along with LLAND and the marginally insignificant LURDISP. The magnitudes of several coefficients change by a fairly large amount (relative to the size of their OLS standard errors), including LPOP, LGDP, and LLABOR. The reduction in estimated standard errors relative to those for OLS could lead a naive researcher to conclude that greater efficiency has been achieved; more likely, the standard errors may seriously understate the actual precision with which the stepwise procedure has estimated the model's coefficients.

The first of the Bayesian model selection procedures appears in Table 4. OLS has been used as a prior mean, and the prior precision is based on the least squares covariance matrix estimator. The country dummies are forced to enter the model while the other variables enter with prior probability of .5. Notice that LGDP, LPOP, LLAND, DCAP, LLABOR, and LEDUC are excluded from the model with probability 0 (that means they are included with probability 1). The variables DICT and LURDISP enter the model with near certainty. It is highly unlikely that GDPCGR and GDPGR enter the model. LEXP and LPOPGR are excluded from the model with probabilities of .362 and .351, respectively. More interestingly, note how the posterior mean is affected by the probability of exclusion. If the probability of exclusion is near zero, then very little shrinkage toward zero occurs. As the probability increases, the degree of shrinkage increases as well. Also, note that the posterior mean of GDPCGR has changed sign relative to its OLS estimate.

The simulation above was repeated with the prior probability that the country dummy coefficients are zero is set to .5. In this case there is little change in the results and thus the output from this simulation is not reported.

Table 5 contains results based on a prior mean equal to the OLS estimates and a smaller prior precision than that used in Table 4 (i.e., $\tau_j = \sigma_y/\sigma_{x_j}$). Each coefficient is expected to enter the model with 50% prior probability. The results are remarkably similar to those in Table 4. The exception is that LEXP now has a high probability of being zero. Surprisingly, each of the country dummies is selected with certainty. The results using the same prior mean and precision that force the coefficients of the country dummies to appear in the model are very similar to these and are not reported.

In Table 6 the prior mean of the slopes is zero while the country dummies have prior means equal to their respective country sample means of urban primacy. The prior precision is the ratio of the standard deviation of y to that of each regressor. Each of the parameters, including the country dummies, enters the model with prior probability of 50%. In this case, DCAP, LGDP, and LPOP enter the model with certainty. LLAND, LURDISP, and LEDUC has very small probability of being zero. The biggest change involves the variable DICT, which now has a relatively large probability of being zero (about .8). Several of the dummy variables have positive probability of being zero, including Ecuador, El-Salvador, Honduras, Panama, Venezuela, Japan, Taiwan, and the US. Although the proximity of the prior mean to zero seems to have some affect on this result, Canada, India, and Pakistan all have prior means less than 1, yet each enters the model with certainty.

The priors that yield the results in Table 7 are similar to those for Table 6. The only difference is that the country dummies are forced to enter the model. When this happens, LGDP and LEDUC have much higher probabilities of being zero. The inclusion of these variables is thus very sensitive to the treatment of the country fixed effects.

Table 8 is similar to Table 6. In this case the change occurs with prior precision. Precision is much higher when based on OLS as it is in this set of results than when based on σ_y/σ_{x_j}. LLAND and LEDUC now have much higher probabilities of being zero while DICT has a much lower probability of being zero compared to Table 6. Higher precision about the nonzero country dummy variables results in fewer with high probabilites of being omitted. In the final analysis, only YEAR, DCAP and LLABOR have probabilities of being zero smaller than 10% in this setting.

Finally, in Table 9 we find results for the combination of priors found in Table 8, except that the country dummies are forced into the model. The effect relative to the results in Table 8 is minimal.

Variable	B Value	Std Error	t Ratio	Approx Prob
C1	5.964585	2.9828	2.000	0.0472
C2	7.053863	3.3470	2.107	0.0366
C3	4.552137	2.5866	1.760	0.0803
C4	6.010804	3.2750	1.835	0.0683
C5	3.809629	2.9452	1.293	0.1976
C6	5.892112	2.7357	2.154	0.0327
C7	5.338560	2.9539	1.807	0.0725
C8	4.561175	2.3621	1.931	0.0552
C9	5.073967	2.5842	1.963	0.0513
C10	4.502159	2.6095	1.725	0.0863
C11	4.371914	2.5329	1.726	0.0862
C12	6.235191	2.6752	2.331	0.0210
C13	4.210673	2.5073	1.679	0.0950
C14	7.761833	3.7792	2.054	0.0416
C15	7.243089	3.4173	2.120	0.0355
C16	5.835867	3.3766	1.728	0.0858
C17	5.746068	3.0692	1.872	0.0630
C18	5.106657	2.8387	1.799	0.0739
C19	6.572971	3.2069	2.050	0.0420
C20	6.626230	3.1868	2.079	0.0391
C21	3.841227	2.3236	1.653	0.1002
C22	5.003979	2.4499	2.043	0.0427
C23	6.437550	2.8548	2.255	0.0254
C24	7.370627	3.1310	2.354	0.0197
C25	5.827820	2.8562	2.040	0.0429
C26	4.831161	2.8578	1.691	0.0928
C27	8.695596	3.2241	2.697	0.0077
C28	5.668316	3.5018	1.619	0.1074
C29	5.390254	2.4622	2.189	0.0300
C30	4.672893	2.8096	1.663	0.0982
YEAR	-0.008081	0.0312	-0.259	0.7961
LGDP	0.097615	0.0483	2.021	0.0449
LPOP	-0.780064	0.2767	-2.819	0.0054
LLAND	0.172447	0.1166	1.479	0.1412
DCAP	0.490284	0.1777	2.760	0.0064
LLABOR	0.523794	0.2490	2.103	0.0369
LEDUC	0.274560	0.1317	2.085	0.0386
LEXP	-0.044246	0.0525	-0.843	0.4006
DICT	0.065640	0.0286	2.297	0.0229
LURDISP	-0.113895	0.0596	-1.912	0.0576
GDPCGR	0.006720	0.0273	0.247	0.8056
GDPGR	-0.009447	0.0276	-0.343	0.7323
LPOPGR	0.084542	0.0856	0.988	0.3245

Table 1: OLS estimates of the urban primacy model using all independent variables

```
R-square = 0.98867412   C(p) = 35.84191478
```

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	36	487.77963777	13.54943438	417.07	0.0001
Error	172	5.58782295	0.03248734		
Total	208	493.36746072			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
C1	6.28541602	1.46948640	0.59436210	18.30	0.0001
C2	6.54831835	1.79102498	0.43428064	13.37	0.0003
C3	4.65389599	1.36685840	0.37661749	11.59	0.0008
C4	5.94238862	1.62778132	0.43295641	13.33	0.0003
C5	4.45183443	1.42316782	0.31789227	9.79	0.0021
C6	6.05292105	1.37759962	0.62718863	19.31	0.0001
C7	5.19929748	1.50226797	0.38914254	11.98	0.0007
C8	4.57248577	1.25575908	0.43073113	13.26	0.0004
C9	4.97943753	1.37750449	0.42451078	13.07	0.0004
C10	4.57781085	1.34521143	0.37622598	11.58	0.0008
C11	4.15894604	1.34534700	0.31046463	9.56	0.0023
C12	6.04306418	1.43954873	0.57250049	17.62	0.0001
C13	4.20323599	1.37362752	0.30418919	9.36	0.0026
C14	7.43843593	1.93401094	0.48057325	14.79	0.0002
C15	6.79241748	1.75716736	0.48544125	14.94	0.0002
C16	5.56498114	1.65274974	0.36832177	11.34	0.0009
C17	5.50956333	1.56386781	0.40322595	12.41	0.0005
C18	4.83865306	1.47596875	0.34914742	10.75	0.0013
C19	6.57757575	1.63520852	0.52565358	16.18	0.0001
C20	6.31102959	1.62809626	0.48815132	15.03	0.0002
C21	3.96205420	1.23409484	0.33485589	10.31	0.0016
C22	5.16810100	1.31253095	0.50368274	15.50	0.0001
C23	6.39381699	1.47810939	0.60788502	18.71	0.0001
C24	7.20570327	1.62066138	0.64221785	19.77	0.0001
C25	5.56179364	1.48837507	0.45364816	13.96	0.0003
C26	4.58957660	1.44870301	0.32606239	10.04	0.0018
C27	8.64625776	1.74090432	0.80134724	24.67	0.0001
C28	6.12618247	1.66365380	0.44052254	13.56	0.0003
C29	5.59210405	1.24519576	0.65522342	20.17	0.0001
C30	4.68099285	1.43052855	0.34785407	10.71	0.0013
LGDP	0.06596543	0.03766569	0.09964490	3.07	0.0817
LPOP	-0.55454071	0.12048132	0.68824276	21.18	0.0001
DCAP	0.55824015	0.16785549	0.35932316	11.06	0.0011
LLABOR	0.56231818	0.21979874	0.21263171	6.55	0.0114
DICT	0.05737264	0.02766284	0.13974309	4.30	0.0396
LURDISP	-0.10629797	0.05713178	0.11246273	3.46	0.0645

Table 2: Backward Stepwise regression results

R-square = 0.98885798   C(p) = 31.05225956

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	35	487.87034963	13.93915285	438.68	0.0001
Error	173	5.49711110	0.03177521		
Total	208	493.36746072			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
C1	0.73678115	0.16769504	0.61337347	19.30	0.0001
C2	1.23239814	0.24961076	0.77457574	24.38	0.0001
C4	0.32423500	0.20420061	0.08011137	2.52	0.1142
C5	-1.32665279	0.21167021	1.24819801	39.28	0.0001
C6	1.11117936	0.11100532	3.18397127	100.20	0.0001
C7	0.19293322	0.09444056	0.13261277	4.17	0.0426
C8	0.51784759	0.13847213	0.44439339	13.99	0.0003
C9	0.59538562	0.08364983	1.60973713	50.66	0.0001
C12	1.59581852	0.09958796	8.15909091	256.78	0.0001
C13	-0.18495974	0.09815377	0.11283109	3.55	0.0612
C14	1.13909209	0.34427913	0.34784414	10.95	0.0011
C15	1.23722594	0.21849939	1.01879310	32.06	0.0001
C17	0.41192956	0.09726247	0.56995918	17.94	0.0001
C18	0.17373030	0.08929328	0.12028239	3.79	0.0533
C19	0.97022055	0.14683916	1.38722082	43.66	0.0001
C20	1.06398731	0.26142890	0.52632514	16.56	0.0001
C21	-0.18785058	0.13518778	0.06135339	1.93	0.1665
C22	0.73337655	0.10461816	1.56145064	49.14	0.0001
C23	1.47291647	0.08518837	9.49913323	298.95	0.0001
C24	1.89741770	0.12102680	7.81000274	245.79	0.0001
C25	0.86902465	0.10898757	2.02021995	63.58	0.0001
C26	-0.11094879	0.09385992	0.04439900	1.40	0.2388
C27	3.00141899	0.16416050	10.62195254	334.28	0.0001
C28	-0.41936927	0.20732381	0.13001194	4.09	0.0446
C29	1.09969601	0.15656447	1.56764353	49.34	0.0001
C30	-0.18501664	0.10627618	0.09630263	3.03	0.0835
YEAR	-0.05129099	0.01269344	0.51881405	16.33	0.0001
LGDP	0.12818373	0.02787472	0.67194420	21.15	0.0001
LPOP	-0.37332604	0.05818644	1.30804119	41.17	0.0001
LLAND	0.21876172	0.05385141	0.52436885	16.50	0.0001
DCAP	0.63414839	0.07183092	2.47655101	77.94	0.0001
LLABOR	0.19910114	0.13288944	0.07132717	2.24	0.1359
LEDUC	0.28719112	0.11551829	0.19639433	6.18	0.0139
DICT	0.06098264	0.02725852	0.15903631	5.01	0.0266
LURDISP	-0.13408277	0.05745060	0.17307934	5.45	0.0208

Table 3: Forward Stepwise regression results

Variable Name	Prior Mean	Prior Precision	Posterior Mean	Probability Coeff=0 Prior	Posterior
C1	5.9646	14.6721	5.9853	0.00	0
C2	7.0539	13.5837	7.0253	0.00	0
C3	4.5521	14.6721	4.5400	0.00	0
C4	6.0108	14.6721	6.0244	0.00	0
C5	3.8096	14.6721	3.8279	0.00	0
C6	5.8921	14.6721	5.9058	0.00	0
C7	5.3386	14.6721	5.3450	0.00	0
C8	4.5612	14.6721	4.5759	0.00	0
C9	5.0740	14.6721	5.0872	0.00	0
C10	4.5022	14.6721	4.5127	0.00	0
C11	4.3719	14.6721	4.3652	0.00	0
C12	6.2352	14.6721	6.2227	0.00	0
C13	4.2107	14.6721	4.1796	0.00	0
C14	7.7618	14.6721	7.7327	0.00	0
C15	7.2431	13.5837	7.2123	0.00	0
C16	5.8359	14.6721	5.8616	0.00	0
C17	5.7461	14.6721	5.7531	0.00	0
C18	5.1067	14.6721	5.0953	0.00	0
C19	6.5730	14.6721	6.5834	0.00	0
C20	6.6262	14.6721	6.6030	0.00	0
C21	3.8412	14.6721	3.8402	0.00	0
C22	5.0040	14.6721	5.0017	0.00	0
C23	6.4376	14.6721	6.4492	0.00	0
C24	7.3706	14.6721	7.3687	0.00	0
C25	5.8278	14.6721	5.8225	0.00	0
C26	4.8312	14.6721	4.8417	0.00	0
C27	8.6956	14.6721	8.6551	0.00	0
C28	5.6683	14.6721	5.7100	0.00	0
C29	5.3903	14.6721	5.3901	0.00	0
C30	4.6729	14.6721	4.6925	0.00	0
YEAR	-0.0081	359.3060	-0.0032	0.50	0.6638
LGDP	0.0976	1259.8721	0.0916	0.50	0
LPOP	-0.7801	793.0043	-0.7859	0.50	0
LLAND	0.1724	677.6999	0.1825	0.50	0
DCAP	0.4903	71.0175	0.4993	0.50	0
LLABOR	0.5238	58.3933	0.4935	0.50	0
LEDUC	0.2746	118.3621	0.2571	0.50	0
LEXP	-0.0442	235.3336	-0.0261	0.50	0.3627
DICT	0.0656	138.3052	0.0635	0.50	0.003367
LURDISP	-0.1139	97.6824	-0.1161	0.50	0.0093
GDPCGR	0.0067	283.5257	-0.0003	0.50	0.8977
GDPGR	-0.0094	419.5596	-0.0003	0.50	0.8827
LPOPGR	0.0845	71.5421	0.0527	0.50	0.3515

Table 4: Bayesian Estimates using $\underline{\beta}$ =OLS and prior precision as the square root of diagonal elements of $X'X/\hat{\sigma}^2$. Slopes enter the model with 50% probability. Inequality restrictions are imposed on the coefficients of LPOP, LLAND, and LGDP.

Variable Name	Prior Mean	Prior Precision	Posterior Mean	Probability Coeff=0 Prior	Probability Coeff=0 Posterior
C1	5.9646	4.7840	5.9664	0.50	0
C2	7.0539	5.1545	6.9962	0.50	0
C3	4.5521	4.7840	4.5189	0.50	0
C4	6.0108	4.7840	6.0305	0.50	0
C5	3.8096	4.7840	3.8505	0.50	0
C6	5.8921	4.7840	5.9071	0.50	0
C7	5.3386	4.7840	5.3564	0.50	0
C8	4.5612	4.7840	4.6544	0.50	0
C9	5.0740	4.7840	5.1067	0.50	0
C10	4.5022	4.7840	4.5844	0.50	0
C11	4.3719	4.7840	4.3863	0.50	0
C12	6.2352	4.7840	6.2263	0.50	0
C13	4.2107	4.7840	4.1508	0.50	0
C14	7.7618	4.7840	7.6594	0.50	0
C15	7.2431	5.1545	7.1507	0.50	0
C16	5.8359	4.7840	5.8992	0.50	0
C17	5.7461	4.7840	5.7860	0.50	0
C18	5.1067	4.7840	5.1254	0.50	0
C19	6.5730	4.7840	6.5535	0.50	0
C20	6.6262	4.7840	6.5728	0.50	0
C21	3.8412	4.7840	3.8729	0.50	0
C22	5.0040	4.7840	5.0211	0.50	0
C23	6.4376	4.7840	6.4565	0.50	0
C24	7.3706	4.7840	7.3686	0.50	0
C25	5.8278	4.7840	5.8513	0.50	0
C26	4.8312	4.7840	4.8938	0.50	0
C27	8.6956	4.7840	8.5855	0.50	0
C28	5.6683	4.7840	5.7367	0.50	0
C29	5.3903	4.7840	5.3884	0.50	0
C30	4.6729	4.7840	4.7225	0.50	0
YEAR	-0.0081	0.4340	-0.0004	0.50	0.9672
LGDP	0.0976	0.5806	0.0765	0.50	0
LPOP	-0.7801	0.5769	-0.7995	0.50	0
LLAND	0.1724	0.5156	0.2199	0.50	0
DCAP	0.4903	2.1125	0.5536	0.50	0
LLABOR	0.5238	2.2224	0.4973	0.50	0
LEDUC	0.2746	1.6196	0.2250	0.50	0
LEXP	-0.0442	1.3582	-0.0057	0.50	0.7362
DICT	0.0656	1.3978	0.0618	0.50	0.0873
LURDISP	-0.1139	1.7811	-0.1005	0.50	0.1124
GDPCGR	0.0067	0.3071	-0.0000	0.50	0.9912
GDPGR	-0.0094	0.3043	-0.0000	0.50	0.9968
LPOPGR	0.0845	1.9592	0.0156	0.50	0.7476

Table 5: Bayesian Estimates using $\underline{\beta}$ =OLS as the prior mean, σ_y/σ_x as the prior precision, and $\underline{p}_j = .5$. Inequality restrictions are imposed on LPOP, LLAND, and LGDP

Variable Name	Prior Mean	Prior Precision	Posterior Mean	Probability Coeff=0 Prior	Posterior
C1	2.3160	4.7840	1.3232	0.50	0
C2	0.7381	5.1545	1.0770	0.50	0
C3	1.0553	4.7840	0.3158	0.50	0.0003
C4	0.2413	4.7840	0.4256	0.50	0
C5	0.0959	4.7840	-0.7727	0.50	0
C6	2.4238	4.7840	1.4910	0.50	0
C7	0.9271	4.7840	0.3469	0.50	3.333e-005
C8	1.6845	4.7840	0.6198	0.50	0
C9	1.3437	4.7840	0.6872	0.50	0
C10	0.3022	4.7840	0.0125	0.50	0.8349
C11	0.7096	4.7840	0.0054	0.50	0.9248
C12	2.2297	4.7840	1.6626	0.50	0
C13	0.6047	4.7840	-0.0006	0.50	0.9227
C14	0.1571	4.7840	1.0729	0.50	0
C15	1.1933	5.1545	1.2903	0.50	0
C16	0.6949	4.7840	0.0228	0.50	0.8555
C17	0.9683	4.7840	0.4687	0.50	0
C18	0.8313	4.7840	0.1830	0.50	0.09587
C19	1.7709	4.7840	1.2689	0.50	0
C20	0.5107	4.7840	1.0428	0.50	0
C21	1.1294	4.7840	0.0064	0.50	0.9546
C22	1.9496	4.7840	1.0322	0.50	0
C23	2.2470	4.7840	1.6393	0.50	0
C24	2.3168	4.7840	1.9880	0.50	0
C25	1.3519	4.7840	0.8382	0.50	0
C26	0.6715	4.7840	-0.0027	0.50	0.9385
C27	3.4100	4.7840	3.1746	0.50	0
C28	0.5676	4.7840	0.0369	0.50	0.8052
C29	2.7885	4.7840	1.5292	0.50	0
C30	0.9548	4.7840	0.0033	0.50	0.9637
YEAR	0.0000	0.4340	-0.0401	0.50	0.0089
LGDP	0.0000	0.5806	0.1253	0.50	0
LPOP	0.0000	0.5769	-0.2280	0.50	0
LLAND	0.0000	0.5156	0.0545	0.50	0.08297
DCAP	0.0000	2.1125	0.4752	0.50	0
LLABOR	0.0000	2.2224	0.0639	0.50	0.5323
LEDUC	0.0000	1.6196	0.2040	0.50	0.0457
LEXP	0.0000	1.3582	-0.0064	0.50	0.8247
DICT	0.0000	1.3978	0.0075	0.50	0.8176
LURDISP	0.0000	1.7811	-0.1005	0.50	0.1204
GDPCGR	0.0000	0.3071	-0.0000	0.50	0.9917
GDPGR	0.0000	0.3043	-0.0000	0.50	0.9955
LPOPGR	0.0000	1.9592	0.0024	0.50	0.9352

Table 6: Bayesian Estimates using $\underline{\beta_2}$ as the prior mean and σ_y/σ_x as the prior precision. Inequality restrictions are imposed on LGDP. LPOP, and LLAND

Variable Name	Prior Mean	Prior Precision	Posterior Mean	Probability Coeff=0 Prior	Posterior
C1	2.3160	4.7840	1.8836	0.00	0
C2	0.7381	5.1545	1.1526	0.00	0
C3	1.0553	4.7840	0.8593	0.00	0
C4	0.2413	4.7840	0.5563	0.00	0
C5	0.0959	4.7840	0.0056	0.00	0
C6	2.4238	4.7840	2.1311	0.00	0
C7	0.9271	4.7840	0.8527	0.00	0
C8	1.6845	4.7840	1.5954	0.00	0
C9	1.3437	4.7840	1.3094	0.00	0
C10	0.3022	4.7840	0.6560	0.00	0
C11	0.7096	4.7840	0.6950	0.00	0
C12	2.2297	4.7840	2.2172	0.00	0
C13	0.6047	4.7840	0.5072	0.00	0
C14	0.1571	4.7840	0.7402	0.00	0
C15	1.1933	5.1545	1.3127	0.00	0
C16	0.6949	4.7840	0.7252	0.00	0
C17	0.9683	4.7840	1.0406	0.00	0
C18	0.8313	4.7840	0.8875	0.00	0
C19	1.7709	4.7840	1.5845	0.00	0
C20	0.5107	4.7840	0.9547	0.00	0
C21	1.1294	4.7840	0.9426	0.00	0
C22	1.9496	4.7840	1.7600	0.00	0
C23	2.2470	4.7840	2.1653	0.00	0
C24	2.3168	4.7840	2.3882	0.00	0
C25	1.3519	4.7840	1.4824	0.00	0
C26	0.6715	4.7840	0.6708	0.00	0
C27	3.4100	4.7840	3.4694	0.00	0
C28	0.5676	4.7840	0.6026	0.00	0
C29	2.7885	4.7840	2.4272	0.00	0
C30	0.9548	4.7840	0.7504	0.00	0
YEAR	0.0000	0.4340	-0.0113	0.50	0.6256
LGDP	0.0000	0.5806	0.0072	0.50	0.9301
LPOP	0.0000	0.5769	-0.1494	0.50	0
LLAND	0.0000	0.5156	0.1387	0.50	0
DCAP	0.0000	2.1125	0.4729	0.50	0
LLABOR	0.0000	2.2224	0.2693	0.50	0.0372
LEDUC	0.0000	1.6196	0.0074	0.50	0.8795
LEXP	0.0000	1.3582	-0.0007	0.50	0.947
DICT	0.0000	1.3978	0.0237	0.50	0.5558
LURDISP	0.0000	1.7811	-0.0095	0.50	0.9024
GDPCGR	0.0000	0.3071	0.0001	0.50	0.992
GDPGR	0.0000	0.3043	0.0000	0.50	0.9941
LPOPGR	0.0000	1.9592	0.0001	0.50	0.9419

Table 7: Bayesian Estimates using $\underline{\beta_2}$ as the prior mean and σ_y/σ_x as the prior precision. Country dummies enter with certainty and are not constrained to be positive.

Variable Name	Prior Mean	Prior Precision	Posterior Mean	Probability Coeff=0 Prior	Posterior
C1	2.3160	14.6721	2.1058	0.50	0
C2	0.7381	13.5837	0.9695	0.50	0
C3	1.0553	14.6721	1.0105	0.50	0
C4	0.2413	14.6721	0.3528	0.50	0.001167
C5	0.0959	14.6721	0.0270	0.50	0.622
C6	2.4238	14.6721	2.2684	0.50	0
C7	0.9271	14.6721	0.8527	0.50	0
C8	1.6845	14.6721	1.5914	0.50	0
C9	1.3437	14.6721	1.3242	0.50	0
C10	0.3022	14.6721	0.4193	0.503	0.
C11	0.7096	14.6721	0.6815	0.50	0
C12	2.2297	14.6721	2.2474	0.50	0
C13	0.6047	14.6721	0.6459	0.50	0
C14	0.1571	14.6721	0.4754	0.50	0.0004
C15	1.1933	13.5837	1.2458	0.50	0
C16	0.6949	14.6721	0.5625	0.50	0
C17	0.9683	14.6721	0.9108	0.50	0
C18	0.8313	14.6721	0.7959	0.50	0
C19	1.7709	14.6721	1.7116	0.50	0
C20	0.5107	14.6721	0.7228	0.50	0
C21	1.1294	14.6721	1.0268	0.50	0
C22	1.9496	14.6721	1.8826	0.50	0
C23	2.2470	14.6721	2.2032	0.50	0
C24	2.3168	14.6721	2.3310	0.50	0
C25	1.3519	14.6721	1.3548	0.50	0
C26	0.6715	14.6721	0.5527	0.50	0
C27	3.4100	14.6721	3.5434	0.50	0
C28	0.5676	14.6721	0.5141	0.503	0.0003
C29	2.7885	14.6721	2.5880	0.50	0
C30	0.9548	14.6721	0.8303	0.50	0
YEAR	0.0000	359.3060	-0.0208	0.50	0.0846
LGDP	0.0000	1259.8721	0.0109	0.50	0.3712
LPOP	0.0000	793.0043	-0.0192	0.50	0.3502
LLAND	0.0000	677.6999	0.0105	0.50	0.599
DCAP	0.0000	71.0175	0.1780	0.50	0.0534
LLABOR	0.0000	58.3933	0.2056	0.50	0.03423
LEDUC	0.0000	118.3621	0.0180	0.50	0.5574
LEXP	0.0000	235.3336	-0.0047	0.50	0.6782
DICT	0.0000	138.3052	0.0233	0.50	0.4188
LURDISP	0.0000	97.6824	-0.0310	0.50	0.4252
GDPCGR	0.0000	283.5257	0.0010	0.50	0.8598
GDPGR	0.0000	419.5596	0.0002	0.50	0.8368
LPOPGR	0.0000	71.5421	-0.0014	0.50	0.7357

Table 8: Bayesian Estimates using $\underline{\beta_2}$ as the prior mean and the diagonal elements of the inverse of the estimated OLS covariance matrix as the prior precision. All coefficients enter the model with 50% prior probability. Only LGDP, LLAND, and LPOP have inequality restrictions.

Variable Name	Prior Mean	Prior Precision	Posterior Mean	Probability Coeff=0 Prior	Posterior
C1	2.3160	14.6721	2.1264	0.00	
C2	0.7381	13.5837	0.9535	0.00	0
C3	1.0553	14.6721	1.0166	0.00	0
C4	0.2413	14.6721	0.3675	0.00	0
C5	0.0959	14.6721	0.0805	0.00	0
C6	2.4238	14.6721	2.2819	0.00	0
C7	0.9271	14.6721	0.8542	0.00	0
C8	1.6845	14.6721	1.5962	0.00	0
C9	1.3437	14.6721	1.3238	0.00	0
C10	0.3022	14.6721	0.4381	0.00	0
C11	0.7096	14.6721	0.6780	0.00	0
C12	2.2297	14.6721	2.2418	0.00	0
C13	0.6047	14.6721	0.6453	0.00	0
C14	0.1571	14.6721	0.4772	0.00	0
C15	1.1933	13.5837	1.2366	0.00	0
C16	0.6949	14.6721	0.5666	0.00	0
C17	0.9683	14.6721	0.9107	0.00	0
C18	0.8313	14.6721	0.7958	0.00	0
C19	1.7709	14.6721	1.7151	0.00	0
C20	0.5107	14.6721	0.7266	0.00	0
C21	1.1294	14.6721	1.0379	0.00	0
C22	1.9496	14.6721	1.8904	0.00	0
C23	2.2470	14.6721	2.2059	0.00	0
C24	2.3168	14.6721	2.3300	0.00	0
C25	1.3519	14.6721	1.3523	0.00	0
C26	0.6715	14.6721	0.5556	0.00	0
C27	3.4100	14.6721	3.5399	0.00	0
C28	0.5676	14.6721	0.5530	0.00	0
C29	2.7885	14.6721	2.6043	0.00	0
C30	0.9548	14.6721	0.8389	0.00	0
YEAR	0.0000	359.3060	-0.0193	0.50	0.1208
LGDP	0.0000	1259.8721	0.0095	0.50	0.4403
LPOP	0.0000	793.0043	-0.0136	0.50	0.3602
LLAND	0.0000	677.6999	0.0054	0.50	0.6668
DCAP	0.0000	71.0175	0.1932	0.50	0.03
LLABOR	0.0000	58.3933	0.1993	0.50	0.0454
LEDUC	0.0000	118.3621	0.0085	0.50	0.578
LEXP	0.0000	235.3336	-0.0068	0.50	0.6376
DICT	0.0000	138.3052	0.0221	0.50	0.425
LURDISP	0.0000	97.6824	-0.0286	0.50	0.4682
GDPCGR	0.0000	283.5257	0.0008	0.50	0.865
GDPGR	0.0000	419.5596	0.0005	0.50	0.8394
LPOPGR	0.0000	71.5421	-0.0025	0.50	0.7341

Table 9: Bayesian Estimates using $\underline{\beta_2}$ as the prior mean and the diagonal elements of the inverse of the estimated OLS covariance matrix as the prior precision. The slope coefficients enter the model with 50% prior probability. Only LGDP, LLAND, and LPOP have inequality restrictions.

7 Analysis

To analyze and summarize the effects of the various specifications of prior information on the selection of regressors in this model is a difficult task. To help accomplish this, Table 10 summarizes the probabilities that each coefficient is zero for each of the priors chosen.

Few clear cut results emerge; however, some general observations are:

1. Combining prior means of zero with high precisions produced using the OLS covariance matrix tends to increase the probability that coefficients are zero. Hardly a surprise.
2. In most instances the results were not very sensitive to the prior probabilities placed on the country dummy variables. Recall, they were either forced to be nonzero or were given prior probabilities of .5. The exception is seen by comparing column (3) and (4). When the country dummies are forced into the model, LLABOR displaces LGDP and LURDISP. This is due to the probable exclusion of many of the country dummies in (3).
3. As a coefficient became more likely to be zero, the degree of shrinkage toward zero of the OLS coefficient estimate increased. This could mean that the estimator has a relatively good mean square error risk performance. Certainly, this is worth looking into.
4. DCAP is important in every instance.
5. GDPCGR and GDPGR are unimportant in every instance.
6. In most of the models the economic size variables LGDP, LPOP, and LLAND had low probabilites of being zero. One exception is the model in column three where LGDP has a 93% probability of being zero. The other exception occurs when a zero prior is used with OLS based precision (columns (7) and (8)) where LLAND has greater than 50% probability of being zero.
7. Most of the other slopes have both high and low probabilities of being zero depending on the prior selected.
8. If pressed into stating which of the priors we prefer, we would select that in column 2. Our predisposition is that the dummy variables should either all be in or all be out of the model. Since we think it important to account for unobservable country differences that they should all be in. OLS is a sensible and conservative prior mean, though our degree of preference for it is small relative to the other.

9. The results for the preferred set of priors which appear in Table 4 are quite similar to the ones from the backward stepwise procedure in Table 2. All of the dummy variables enter the model; LGDP, LPOP, DCAP, LLABOR, and DICT enter the model and the coefficient estimates are reasonably similar to the posterior means from the Bayesian procedure. The biggest difference is the implied zero coefficients from Stepwise for LLAND and LEDUC which each have posterior means of approximately .22.

10. The main problem with using OLS precision with this estimator is that no use is made of the nonzero covariances associated with it. If the estimator is modified to use this information, our preference for it might change.

11. The combination of priors from column (3) of Table 10 yield model results somewhat similar to forward stepwise. A few of the estimated dummies (C3, C10, C11 and C16) that are omitted by stepwise also have nonzero probabilities of being zero in this scenario. Also, the coefficient on LGDP is very similar in both cases. The biggest differences in point estimates are for LLABOR and DICT.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Prior Mean=	OLS	OLS	Slopes=0	Slopes=0	OLS	OLS	Slopes=0	Slopes=0
Prior Precision=	Std D	Std D	Std D	Std D	OLS	OLS	OLS	OLS
Prior Prob of 0=	.5	Slopes=.5	.5	Slopes=.5	.5	Slopes=.5	.5	Slopes=.5
Variable				Probability that coefficient is zero				
C1	0	0	0	0	0	0	0	0
C2	0	0	0	0	0	0	0	0
C3	0	0	0.0003	0	0	0	0	0
C4	0	0	0	0	0	0	0.001167	0
C5	0	0	0	0	0	0	0.622	0
C6	0	0	0	0	0	0	0	0
C7	0	0	3.3e-005	0	0	0	0	0
C8	0	0	0	0	0	0	0	0
C9	0	0	0	0	0	0	0	0
C10	0	0	0.8349	0	0	0	0	0
C11	0	0	0.9248	0	0	0	0	0
C12	0	0	0	0	0	0	0	0
C13	0	0	0.9227	0	0	0	0	0
C14	0	0	0	0	0	0	0.0004	0
C15	0	0	0	0	0	0	0	0
C16	0	0	0.8555	0	0	0	0	0
C17	0	0	0	0	0	0	0	0
C18	0	0	0.09587	0	0	0	0	0
C19	0	0	0	0	0	0	0	0
C20	0	0	0	0	0	0	0	0
C21	0	0	0.9546	0	0	0	0	0
C22	0	0	0	0	0	0	0	0
C23	0	0	0	0	0	0	0	0
C24	0	0	0	0	0	0	0	0
C25	0	0	0	0	0	0	0	0
C26	0	0	0.9385	0	0	0	0	0
C27	0	0	0	0	0	0	0	0
C28	0	0	0.8052	0	0	0	0.0003	0
C29	0	0	0	0	0	0	0	0
C30	0	0	0.9637	0	0	0	0	0
YEAR	0.9672	0.9514	0.0089	0.6256	0.6346	0.6638	0.0846	0.1208
LGDP	0	0	0	0.9301	0	0	0.3712	0.4403
LPOP	0	0	0	0	0	0	0.3502	0.3602
LLAND	0	0	0.08297	0	0	0	0.599	0.6668
DCAP	0	0	0	0	0	0	0.0534	0.03
LLABOR	0	0	0.5323	0.0372	0	0	0.03423	0.0454
LEDUC	0	0.02043	0.0457	0.8795	0	0	0.5574	0.578
LEXP	0.7362	0.8168	0.8247	0.947	0.3422	0.3627	0.6782	0.6376
DICT	0.0873	0.1251	0.8176	0.5558	0.021	0.003367	0.4188	0.425
LURDISP	0.1124	0.1169	0.1204	0.9024	0.0158	0.0093	0.4252	0.4682
GDPCGR	0.9912	0.9914	0.9917	0.992	0.9122	0.8977	0.8598	0.865
GDPGR	0.9968	0.9963	0.9955	0.9941	0.8945	0.8827	0.8368	0.8394
LPOPGR	0.7476	0.8235	0.9352	0.9419	0.3283	0.3515	0.7357	0.7341

Table 10: Summary of the probabilites that each coefficient is zero

8 Critique

The procedure used to select the likely regressors suffers from several deficiencies. First, the results are somewhat sensitive to our selection of prior mean and precision. This may largely be a function of sample size; only in models with large numbers of observations would we normally expect the posterior distribution to be robust to changes in the prior (Geweke 1994). Still, many economists are leery of using techniques that yield results so dependent on users input. George and McCulloch (1996) have proposed an alternative to Geweke's estimator that can be used in a more or less automatic way. It remains to be seen how the George and McCulloch procedure compares to the one used here.

Another weakness of this study is the small number of Gibbs samples taken. It is well known that convergence of the Markov chain is crucial to obtaining useful results and that the number of samples required to reach convergence increases with the degree of collinearity. In this set of data the largest condition number is 166, suggesting severe collinearity. The small number of samples chosen was mainly due to the limited time available to the authors to complete the paper for this conference. In subsequent research we expect to do a better job of assuring that convergence has occurred. In the end, a more efficient algorithm is badly needed since several of the prior combinations eventually ended up in an infeasible portion of the parameter space that the Gibbs sampler could not get out of.

Another shortcoming involves the naive assumption that the parameters are mutually independent. At a minimum it seems that the sampling from the posterior distribution should come from the conditional truncated normal as Geweke subsequently has done. In this preliminary study we wanted to gain some experience with Geweke's Bayesian model selection procedure in its simplest form and to assess its usefulness before plunging into its more complicated variants.

Despite these shortcomings, the results are fairly promising. The procedure yielded probabilities and estimates that were consistent with both our prior expectations about what the proper model should be and reasonably robust with respect to the various priors employed. One plan is to study the empirical risk properties of a much smaller model (4 or 5 variables). Given the rather interesting pattern of shrinkage, it is our conjecture that the estimator will exhibit good squared error risk properties under reasonable conditions involving specification of the prior. Also, it may prove that this estimator has good component-wise risk properties, a feat that has proven elusive.

In addition, an efficient means of assessing posterior model probability is in the works which should aid in the final analysis of likely regressor combinations.

Appendix

Countries

The 30 countries, in alphabetical order, are: Argentina, Bangladesh, Bolivia , Brazil, Canada, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El-Salvador, Guatemala, Honduras, India, Indonesia, Japan, Korea (south), Malaysia, Mexico, Pakistan, Panama, Paraguay, Peru, Philippines, Sri Lanka, Taiwan, Thailand, United States, Uruguay, and Venezuela.

Variables and Sources

An 'L' before the variable indicates that the natural logarithm has been taken. For instance, $LPOP = \ln(POP)$.

PRIMACY: the ratio of the largest city's population to that of the second largest city. Urban population data are from The UN World Urbanization Prospects: The 1992 Revision, The Europa World Yearbook, The Statesman's Yearbook, and the World Development Report.

GDP: gross domestic product. GDP is calculated from the Penn World Tables (Mark 5.6).

GDPC: gross domestic product per capita. Source: The Penn World Tables (Mark 5.6).

POPDENS: population density which is the ratio of the total population to the arable land. Sources: total population data are from The UN World Urbanization Prospects: The 1992 Revision; arable land data are from FAO Production Yearbook (see www.fao.org.).

DCAP: dummy variable that equals 1 if the capital city is also the largest city and equals 0 otherwise. Sources: UN Demographic Yearbook, and World Urbanization Prospects: The 1992 Revision.

EXP: exports of goods and nonfactor services as a percentage of GDP. Sources: World Tables, 1994 and different issues of World Development Report.

LLABOR: share of labor outside agriculture. It is calculated as 1 minus the percentage of economically active population in agriculture. Source: FAO Production Yearbook.

DICT: dictatorship variable (1 = free, 2 = partially free, 3 = not free). Sources: different issues of Gastil's Freedom in the World, and Bollen (1990), for cross-sectional data, it is averaged over 1975, 1980, 1985, and 1990.

EDUC: is average years of schooling of people 25 years and older from Barro and Lee (1993), and from different issues of the UN Human Development Report

URBDISP: The urban-rural disparity is the ratio of ratio of output per worker in nonagriculture (i.e., the percentage of nonagriculture GDP divided by percentage of labor force in nonagriculture) to output per worker in agriculture (i.e., the percentage contribution of agriculture to total GDP divided by the percentage of the labor force in agriculture). The data for percentage contribution of agriculture to total GDP are from various issues of World Tables and World Development Report, and data for percentage of the labor force in agriculture are from FAO Production Yearbook

FDI: Foreign direct investment as a proportion of GDP. FDI data are from various issues of IMF Balance Of Payments Yearbook and IMF International Financial Statistics.

GPDCGR: per capita GDP growth

GDPGR: GDP growth

POPGR: population growth

* Data for the Republic of China (Taiwan) are from different issues of The Republic of China Statistical Yearbook, The Europa Yearbook, and The Statesman's Yearbook.

References

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

Bairoch, P., *Cities and Economic Development: From the Dawn of History to The Present*, Chicago: The University of Chicago Press, 1988.

Benson, B. and M.D. Faminow, "Location Choice and Urban Growth in a Rent-Seeking Society," *Public Finance Quarterly*, 1988, *16*, 158–178.

Carroll, G. R., "National City-size Distributions: What Do We Know after 67 Years of Research?," *Progress in Human Geography*, 1982, *6* (1).

Castells, M, *The Urban Question: A Marxist Approach*, Cambridge: MIT Press, 1977.

Feldstein, Martin S., "Inflation, Tax Rules and Investment: Some

Econometric Evidence," *Econometrica*, 1983, *50*, 826–862.

George, Edward I. and Robert E. McCulloch, "Approaches for Bayesian Model Selection," *University of Texas, working paper*, 1996, *November.*

Geweke, John, "Variable Selection and Model Comparison in Regression," *Working Paper*, May 1994.

Gilbert, A. G. and J. Gugler, *Cities, Poverty, and Development: Urbanization in the Third World*, New York: Oxford University Press, 1992.

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

Jefferson, M., "The Law of the Primate City," *Geographical Review*, April 1939, *29*, 226–232.

Kennedy, Peter, *A Guide to Econometrics*, fourth ed., Cambridge, MA: MIT Press, 1998.

Krugman, Paul, "Urban Concentration: The Role of Increasing Returns and Transportation Costs," *International Regional Science Review*, 1996, *19*, 5–30.

Lipton, M., *Why Poor People Stay Poor: Urban Bias in World Development*, Cambridge, MA: Harvard University Press, 1977.

London, B., "Structural Determinants of Third World Urban Change: An Ecological and Political Economic Analysis," *American Sociological Review*, 1987, *52*, 28–43.

Mills, E. S. and B. W. Hamilton, *Urban Economics*, 5th ed., New York: HarperCollins College Publishers, 1994.

Moomaw, Ronald L. and Ali M. Shatter, "Urbanization as a Factor in Economic Growth: An Empirical Study," *Journal of Economics*, 1993, *19* (2), 1–6.

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

Sheppard, E., "City Size Distributions and Spatial Economic Change," *International Regional Science Review*, 1982, *7* (2), 127–151.