NBER TECHNICAL WORKING PAPER SERIES

EASTERN DATA AND WESTERN ATTITUDES

Edward E. Leamer

Technical Working Paper No. 114

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 1991

Prepared for the Conference on Econometrics of Short and Unreliable Time Series at the Institute for Advanced Studies, Vienna, June 14-15, 1991. This paper is part of NBER's research program in International Studies. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #114
October 1991

EASTERN DATA AND WESTERN ATTITUDES

ABSTRACT

Most studies of the economies of Eastern Europe by Western analysts depend substantially on Western data and Western attitudes. Usually this dependence is implicit and concealed. An explicit and transparent treatment may yield better results, both for the individual analyst and for the profession overall. This article proposes and illustrates an econometric method for pooling Western and Eastern data. The pooled estimates depend on doubt about the Western attitudes, on the degree of experimental contamination in Western and Eastern data and on the similarity of Western and Eastern structures. The method is illustrated by a study of the determinants of the growth rates of developed and developing countries.

Edward E. Leamer
Anderson Graduate School
of Management
UCLA
Los Angeles, CA 90024
and NBER

Eastern Data and Western Attitudes

by

Edward E. Leamer[1]

Empirical studies of the Eastern European economies for the foreseeable future will have to make due with data sets that are limited in terms of quantity, quality and relevance. Facing this scarcity of useful data, analysts will have to import data from analogous Western countries. Many of these imports will be buried in the baggage of ideas that Western economists have formed from observation of Western economies. These hidden stowaway ideas will make it difficult to evaluate the inferences that Western economists draw from Eastern data. We will be forced to guess if an analyst is acting as if Hungary is more analogous to Austria or more analogous to Spain. It may be better to pool Western with Eastern data in a formal econometric way, and thereby to develop a language that facilitates the conversation about the strength and the importance of the analogies on which the pooling depends.

This paper proposes an econometric method of pooling Western and Eastern data for the estimation of a linear regression model. The method is Bayesian and uses prior information about the regression coefficients. The pooled estimates depend on three parameters: (1) δ, the lack of confidence in Western attitudes, (2) ρ, the degree of similarity of Western and Eastern economic structures, and (3) λ_1, the amount of contamination in the Eastern and Western data caused by measurement errors, left-out variables, simultaneity and the like.

Assumptions about these three items will be necessary whenever data sets are pooled whether formally or informally. The formal treatment suggested in this paper is simple and direct. It makes obvious how the pooling depends on these three assumptions. The simplicity and transparency of this framework should facilitate the conversation that ought to occur about the role that Western and Eastern analogies play in Eastern policymaking.

But a formal approach is not without serious shortcomings. A formal analysis requires more work. More importantly, mathematical, numerical and cognitive limitations dictate great simplifications and a high degree of inflexibility in a formal approach. Though analyzing Eastern data will necessarily depend on the confidence in Western attitudes, the degree of similarity of the structures and the amount of experimental contamination, these vague concepts are translated into a precise mathematical model for inference in a way that may leave you so uncomfortable that the mathematics impedes the conversation. Ultimately it is up to the consumer to decide if the cost in terms of inflexibility is worth the increase in clarity.

The principal methodological contribution of this paper is the addition of the experimental bias parameters into the pooling problem. The traditional econometric methods of pooling data sets are intended to deal with data that are limited in terms of quantity, but not in terms of quality and relevance. The experimental bias parameters allow the data sets to be of doubtful relevance and quality.

The traditional methods for pooling data sets use the random coefficients model in which parameters applying to different experiments are treated as if they were a random sample out of a population.

Methods such as "the random coefficients model", "Type II Analysis of Variance", "Bayes" or "empirical Bayes" differ primarily in how they treat uncertainty about the hypothetical population of parameters. For example, Aigner and Leamer(1984), use an empirical Bayes approach to pool time-of-use pricing experiments that were performed by different electric utilities. DeMouchel and Harris(1981) use Bayes and empirical Bayes approaches to "combine cancer experiments in man and other species".

These approaches generally presume that the data are perfect in quality and relevance. Another tradition deals with dubious data by referring to errors in measurement in the variables. Measurement error causes bias in the estimates of regression coefficients, which can be corrected only given some information about the probable amount of the measurement errors. Absent that additional information, the model is underidentified and a data set admits a set of equally good estimates - the so-called errors-in-variables bound. See for example, Leamer(1987), Klepper and Leamer(1984), Aigner et. al. (1984) and Patefield(1981).

In a companion paper, Leamer(1991) directly combines these two statistical traditions to deal with data sets that are both brief and noisy, but with identical structures and no prior information. The errors-in-variable model is complex to begin with and when it is combined with the random coefficients model which would allow differences in structures, the analysis becomes quite difficult. In this paper, the errors-in-variable model is reparameterized to allow a somewhat easier treatment. The explanatory variables are entered twice in the regression model, once to represent the "true" or "structural" coefficients and a second time to represent the experimental bias associated with measurement errors, and other statistical problems that cause bias. This "contaminated" regression model has been used by Leamer(1974) and Leamer(1978) to discuss data-instigated models.

Traditionally, errors-in-variables issues would be analyzed with a measurement model and prior information about the bias would be entered indirectly through prior information about the signal to noise ratio in the measurement model. Here the path is the opposite direction. Prior information about the signal to noise ratio is being introduced implicitly through prior information about the bias parameters. The choice of approach depends on economy and accuracy. It is clearly more economical to use the bias parameters rather than the signal-to-noise ratios since the econometric procedures that refer to bias parameters are an order of magnitude easier than procedures that refer to a measurement model. But accuracy is another matter. If your prior information refers directly to signal-to-noise ratios, you may find it very difficult to think about the bias parameters which are complicated functions of the signal and noise covariance matrices.

The symbols that represent the three critical inputs are (1) δ, the doubt about the prior (2) ρ , the structural similarity,and (3) λ, the experimental bias. The traditional pooling model uses the data to estimate ρ, the degree of similarity in the structural parameters, but takes special values of δ and λ_i: $\delta = \infty$ (No prior information), and $\lambda_i = 0$ (No experimental bias).

Doubt about the values of these "hyperparameters" can be treated either with a sensitivity analysis or with an estimation approach. The discussion that follows presumes that there are only two data sets to be pooled.[2] With a sample size of only two it is impossible with accuracy to infer much about the distribution from which the regression parameters are drawn and a sensitivity analysis may be the preferred approach. Both Bayesian estimation of these hyperparameters and also sensitivity analysis are proposed.

The method is illustrated by a study of the determinants of the growth rates of developed and developing countries. These two data subsets yield estimates that are very similar, which makes the results not very sensitive to the form of pooling and which points toward low experimental bias and a high degree of similarity.

1.0 Bayesian pooling of contaminated data sets

A Bayesian approach for the pooling of contaminated data sets makes use of the contaminated regression model:

$$y_i = X_i \beta_i + X_i \theta_i + \epsilon_i \qquad i = 1,2$$

$$\epsilon_i \sim N(0, \sigma_i^2 I)$$

where y_i is an $n_i \times 1$ vector of observations of the "dependent" variable, X_i is an $n_i \times k$ matrix of observations of the k "explanatory" variables, β is the k×1 vector of "true" parameters and θ is the k×1 bias vector representing the experimental contamination due to measurement errors or any other statistical pathologies. This model suffers from an extreme multicollinearity problem: all the variables enter twice, once to capture the structural effects β and again to capture the experimental contamination θ. The informational deficiencies of this underidentified model can be overcome by supplementing the data information with prior information. The prior that seems natural should embody the ideas that the experimental contamination is probably small, and that the structural parameters are probably similar in the two data sets.

Smallness of the experimental contamination can be captured with a prior located at the origin:

$$\theta_i \sim N(0,\ V_i),$$

where V_i is the prior covariance matrix, the smaller its value the smaller is the probable experimental contamination. A prior for the structural parameters that captures information about their probable sizes as well as the idea that they are similar is the normal distribution

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \sim N\left(\begin{bmatrix} p \\ p \end{bmatrix} , \begin{bmatrix} U & \rho U \\ \rho U & U \end{bmatrix} \right)$$

Here the parameter ρ is the correlation of the structural parameters across the two data sets, the vectors p represent the most likely values of these structural parameters and the covariance matrix U measures the likely departures from p. Incidentally, this parameterization allows a relative lack of information about the values of β but confidence that $\beta_1 - \beta_2$ is small. This can be accomplished by selecting a large value of U and a value of ρ close to one such that $\mathrm{Var}(\beta_1 - \beta_2) = 2U(1-\rho)$ is small.

With these as the elements, the full prior covariance matrix takes the form

$$V = \mathrm{Var}\left(\begin{bmatrix} \beta_1 \\ \beta_2 \\ \theta_1 \\ \theta_2 \end{bmatrix} \right) = \begin{bmatrix} U & \rho U & 0 & 0 \\ \rho U & U & 0 & 0 \\ 0 & 0 & V_1 & 0 \\ 0 & 0 & 0 & V_2 \end{bmatrix}$$

The corresponding sample precision matrix is

$$N = \begin{bmatrix} X_1'X_1/\sigma_1^2 & 0 & X_1'X_1/\sigma_1^2 & 0 \\ 0 & X_2'X_2/\sigma_2^2 & 0 & X_2'X_2/\sigma_2^2 \\ X_1'X_1/\sigma_1^2 & 0 & X_1'X_1/\sigma_1^2 & 0 \\ 0 & X_2'X_2/\sigma_2^2 & 0 & X_2'X_2/\sigma_2^2 \end{bmatrix},$$

and the sample cross-product matrix is

$$r = \begin{bmatrix} X_1'y_1/\sigma_1^2 \\ X_2'y_2/\sigma_2^2 \\ X_1'y_1/\sigma_1^2 \\ X_2'y_2/\sigma_2^2 \end{bmatrix}$$

If we denote the stacked vector of parameters by γ:

$$\gamma = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \theta_1 \\ \theta_2 \end{bmatrix} \qquad E(\gamma) = \begin{bmatrix} p \\ p \\ 0 \\ 0 \end{bmatrix}$$

the Bayesian posterior moments are(e.g. Leamer(1978, p.78)):

$$E(\gamma \mid y_1, X_1, y_2, X_2, \sigma_1^2, \sigma_2^2) = (N + V^{-1})^{-1}(r + V^{-1}E(\gamma)) \qquad (1)$$

$$\mathrm{Var}(\gamma \mid y_1, X_1, y_2, X_2, \sigma_1^2, \sigma_2^2) = (N + V^{-1})^{-1} \qquad (2)$$

These two posterior moments condition on the residual variances σ_1^2 and σ_2^2, which rarely would be known. A prior distribution for these parameters is required to deal with the uncertain case, but, unfortunately, there is no prior distribution that leads to a tractable nonapproximate posterior distribution. A traditional prior distribution would be the product of gamma distributions on σ_1^{-2} and σ_2^{-2}. Then after integrating these parameters from the likelihood function, the posterior is the product of a Student distribution for (β_1,θ_1) times a Student distribution for (β_2,θ_2) times the normal prior for $(\beta_1,\theta_1,\beta_2,\theta_2)$. This complicated density can have many modes but will look like a unimodal normal distribution if the two Student pieces are approximately normal. This occurs when the data and prior combine to firmly establish the values of σ_1^2 and σ_2^2 and we then revert approximately to the former case of known σ_1^2 and σ_2^2. In this paper, I will act as if the values of σ_1^2 and σ_2^2 can be accurately estimated from the data and treated as known. The estimates that I will use are:

$$\sigma_i^2 = y_i'(I-X_i(X_i'X_i)^{-1}X_i')y_i/(n_i-k) \tag{3}$$

where n_i is the dimension of y_i and k is the dimension of β. An improved approximation would make use of the prior information about the slope parameters and use the (iterative) estimate

$$\sigma_i^2 = (y_i - X_i\ \beta_i)'(y_i - X_i\ \beta_i)/n_i \tag{4}$$

where β_i is evaluated at, or at least near, the posterior mode. The approximation (3) will be accurate if (3) and (4) are not very different and if the degrees of freedom n_i-k are both adequately large.

Sensitivity Analysis.

It is highly unlikely that the prior parameters (U, ρ, V_1, V_2, p) could be chosen with complete confidence. If the data are not sufficiently informative about these parameters a sensitivity analysis is required to identify those inferences that are too sensitive to the choice of prior to be taken seriously. The sensitivity analysis is facilitated if the number of prior parameters is reduced. A natural constraint is

$$V_i = \lambda_i U,$$

where λ_i measures the relative importance of experimental contamination. Furthermore, since it will be difficult to carry on a sensitivity analysis for all elements of the prior covariance matrix U, it seems better only to perturb U by a scalar δ:

$$U = \delta^2 U_0$$

Thus for this sensitivity analysis the initial prior covariance matrix U_0 and the vector of prior means p are taken as given and the scalars λ_1, λ_2, δ and ρ are perturbed.

To decide if an inference is excessively sensitive to the choice of these parameters we need to select an sensible amount of perturbation which requires clear thinking about the meaning of each parameter. The parameter δ is a discount rate applicable to the prior information. Formally, it is the factor by which the prior standard errors are multiplied. A value of $\delta = 1$ selects the initial prior covariance matrix. A value of $\delta = 2$ selects a prior covariance matrix with standard errors multiplied by two. A range of $1/4 \leq \delta \leq 4$ seems like a large range.

The parameter ρ measures the similarity of the structural parameters across data sets. Though this correlation could be negative, it is unlikely that there would be many settings in which that would be a sensible choice. For that matter, the reason for pooling must be that the structures are adequately similar and values of ρ in excess of .5 see, seem sensible.

To think about the choice of λ_i it is instructive to consider the problem of a single contaminated data set. Then the posterior mean is a weighted average of the prior mean p and the OLS estimate with weights that depend on λ_i. In the next paragraph it is shown that, as the sample size increases, the weights on prior and sample converge respectively to $\lambda_i/(1+\lambda_i)$ and $1/(1+\lambda_i)$. Thus λ_i measures the resistance to the new information: the larger is the value of λ_i, the less weight is put on the sample result asymptotically. The traditional value is λ_i

= 0, meaning that in a sufficiently large sample, the prior is altogether discarded. A value of λ_i equal to one selects a degree of experimental error that asymptotically assigns half the weight to the prior and half to the data.

To show the influence of λ_i as the sample size grows, we may explore the contaminated model with a single data set. Leamer(1978, pp.295-299) reports the following posterior mean:

$$E(\beta_1 | y_1, X_1) = D^{-1} (U^{-1} p + [N_1 - N_1(N_1 + \lambda_1^{-1}U^{-1})^{-1} N_1] b_1)$$

where b_1 is the OLS estimate $(X_1'X_1)^{-1}X_1'y_1$, $N_1 = X_1'X_1/\sigma_1$, and $D = U^{-1} + N_1 - N_1(N_1 + \lambda_1^{-1}U^{-1})^{-1} N_1$ where $N_1 = X_1'X_1/\sigma_1^2$. In words, the posterior mean is a weighted average of the prior mean and the sample estimate. If there were no contamination, that is if $\lambda_1=0$, then the weight on the prior would be the prior precision U^{-1} and the weight on the data would be the data precision N_1. The effect of the sample contamination is to reduce the sample weight to $[N_1 - N_1(N_1 + \lambda_1^{-1}U^{-1})^{-1} N_1]$ $= N_1(N_1 + \lambda_1^{-1}U^{-1})^{-1}(N_1 + \lambda_1^{-1}U^{-1} - N_1) = (I+\lambda_1^{-1}U^{-1}N_1^{-1})^{-1}\lambda_1^{-1}U^{-1}$. Thus as the sample size grows in the sense that $\lambda_1^{-1}U^{-1}N_1^{-1}$ converges to the zero matrix, the weight on the sample converges to $\lambda_1^{-1}U^{-1}$, compared with the sample-independent weight on the prior of U^{-1}.

Estimation

The prior parameters λ_1, λ_2, δ and ρ can be perturbed to see how much they matter. They can also be estimated since there is at least some sample information about their values. For example, if the two ordinary least squares estimates are both very precise and very different, the data are incompatible with the assumption that the parameters are very similar ($\rho = 0$) and without experimental error ($\lambda_1=0$).

The posterior distribution for a parameter is proportional to the product of the prior times the marginal likelihood

$$f(y_1, y_2 | X_1, X_2, \lambda_1, \lambda_2, \delta, \rho)$$

which is the joint marginal density for y_1, y_2 evaluated at the sample values. The label "marginal likelihood" refers to the fact that the parameters β and θ have been integrated from the likelihood function. This integration is rather easily accomplished. We may stack the model to form:

$$\begin{bmatrix} y_1/\sigma_1 \\ y_2/\sigma_2 \end{bmatrix} = \begin{bmatrix} X_1/\sigma_1 & 0 \\ 0 & X_2/\sigma_2 \end{bmatrix} \begin{bmatrix} \beta_1+\theta_1 \\ \beta_2+\theta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1/\sigma_1 \\ \epsilon_2/\sigma_2 \end{bmatrix}$$

which in obvious notation can be written as:

$$Y = X\gamma^* + E$$

Then given the normal prior distribution for the vector γ, the marginal distribution for Y is normal with mean and covariance

$$E(Y) = XE(\gamma^*), \qquad Var(Y) = X\,Var(\gamma^*)X' + I.$$

The corresponding marginal likelihood is therefore:

$$f(Y|X,\lambda_1,\lambda_2,\delta,\rho) \propto |Var(Y)|^{-1/2}\exp(-(Y-XE(\gamma^*))'[Var(Y)]^{-1}(Y-XE(\gamma^*))/2) \quad (4)$$

where the dependence of this statistic on the parameters λ_1, λ_2, δ, and ρ is implicit (Var(Y) depends on them.).

This marginal likelihood involves the inverse and determinant of a high-dimensional matrix, a calculation which stretches the computing capacity of most machines. The dimensionality of the problem can be reduced using the following matrix results:

$$|Var(Y)| = |X\,Var(\gamma^*)X' + I| = |Var(\gamma^*)|\;|\,[Var(\gamma^*)]^{-1} + X'X\,|$$

$$[Var(Y)]^{-1} = I - X\,([Var(\gamma^*)]^{-1} + X'X)^{-1}X'$$

Using these formulae one needs to compute inverses and determinants of matrices of order 2k where k is the number of variables. This contrasts with Var(Y) which is of order $n_1 + n_2$ where n_i is the number of observations of type i. Typically $n_1 + n_2$ is a much larger number than 2k.

The covariance matrix of the parameters can be expressed in Kronecker notation as $\mathrm{Var}(\gamma^*) = (\mathbf{A} \otimes U_0)\, \delta^2$, where

$$\mathbf{A} = \begin{bmatrix} 1+\lambda_1 & \rho \\ \rho & 1+\lambda_2 \end{bmatrix}$$

With **A** a 2×2 matrix and U_0 a 7×7 matrix, results on Kronecker products imply

$$|\mathrm{Var}(\gamma)| = |\mathbf{A} \otimes U_0\, \delta^2| = |\mathbf{A}|^7 |U_0|^2\, \delta^{28} = [(1-\rho^2)\lambda_1\lambda_2]^7 |U_0|^2\, \delta^{28}$$

$$[\mathrm{Var}(\gamma)]^{-1} = \mathbf{A}^{-1} \otimes U_0^{-1}\, \delta^{-2}$$

2.0 An Example: The Convergence Hypothesis

In a companion paper, Leamer(1991) has studied the determinants of the growth rate of per capita GNP using a cross-section of countries as has been done by Barro(1991). That paper deals with the incompatibility of the developed country and developing country subsamples using an errors-in-variables approach but with the assumption that the regression coefficients are the same in both subsamples. Here we combine errors-in-variables concerns with doubt about the similarity of the coefficients. Doubt about the similarity of coefficients is expressed with a random coefficients model that allows variability of regression parameters across observational units. Errors-in-variables concerns enter through a bias parameter vector.

Table 1 contains a list of variables together with various summary statistics. The dependent variable is the average annual growth rate of real per capita gross domestic product from 1960 to 1985. The 98 countries that form the data set grew 2.2 per cent per annum on the average. The lowest per capita growth rate was minus 1.7 per cent; the highest was 7.4 per cent. Seven variables are hypothesized to determine the growth rate. A variable of special interest is the initial per capita GDP which is the focus of the convergence literature. A negative coefficient on initial GDP would indicate that countries which are initially ahead tend to grow more slowly; thus there is convergence of per capita GDP. Barro chooses to control for a number of other effects including school enrollment rates, government expenditures, frequency of revolutions and coups, frequency of assassinations and last a measure of macro-economic disequilibrium - the deviation of investment prices from their mean.

Table 2 contains the prior parameters and the OLS estimates of this model. The prior means and standard errors in the first column were chosen to approximate the author's state of mind prior to analyzing this data set. The coefficient of -0.01 on GDP60 means that it would take a change of initial per capita GDP by 1($1000) to change the growth rate by .01(1 per cent). Similarly, it would take a 10 per cent change in the secondary school enrollment rate, or a 20 per cent change in the primary school enrollment rate, or a 100 per cent in the share of government, or .1 more revolutions per year or 1 more assassinations per year to alter the growth rate by .01 (1 per cent). As I think of this, I am revealing an attitude that the measured share of government probably doesn't matter very much but that political upheavals do matter a lot. Prior standard errors are all set equal to the prior means to suggest a fair degree of confidence in the sign of the coefficients.

The pooled OLS estimates are smaller than the prior means, but all the same sign. The exception is the coefficient on government, which is more than ten times as large as the prior. The other big differences are the coefficients on revolutions and assassinations which are a very small fraction of the prior means. Thus the data seem to want to make me change my mind: government matters more than I thought; revolutions and assassinations don't matter as much.

The regressions on subsets of the data are remarkably similar. The developed country estimates are almost all slightly smaller than the developing country estimates. One exception is the coefficient on government expenditures which is much smaller for the developed than the developing countries. Perusal of Table 2 thus suggests one puzzle: what is the size of the government effect? Are there assumptions that could justify an estimate as big as the pooled OLS of -0.13 or the even greater developing-country OLS of -0.16? Why is the estimate for the developed countries so much smaller (-0.04)?

Various Bayes estimates are reported in Tables 3, 4, and 5. These three tables indicate respectively how the Bayes estimates depend on the three critical inputs: (1) Inaccuracy of the prior information (δ); (2) experimental contamination (λ_1 and λ_2); and (3) structural similarity(ρ). The central case around which the sensitivity analysis is performed uses the prior inaccuracy defined by the standard errors in Table 2 (δ=1), a significant amount of experimental bias (λ_1= λ_2=1) and a moderate degree of structural similarity (ρ = .5).

Table 3 is designed to demonstrate the sensitivity of the estimates to the choice of δ, the factor that multiplies the prior covariance matrix. A large value of δ indicates relatively little confidence in the prior estimates. The first column of numbers in Table 3 has the prior estimates which apply if the prior is taken to be exact, $\delta = 0$. Note that the estimates are then the same for both developed and developing countries. The last column contains the separate OLS regressions, applicable if the prior is fully diluted ($\delta = \infty$) but also if there is no structural similarity (ρ=0) and no experimental contamination $\lambda = 0$.

The central case is reported in the column headed by $\delta = 1$. These numbers might be compared with the pooled OLS estimates reported in Table 3. Although the Bayes estimates for the two subsets are very similar, they are very different from the pooled OLS estimates in the sense that eight of the fourteen Bayes estimates are outside the two-standard error confidence sets around the pooled OLS estimates. The Bayes estimate of the effect of government expenditure, in particular, is much smaller than the pooled OLS estimate.

Nothing very dramatic happens as the inaccuracy of the prior information is increased or diluted. In particular, diluting the prior does not uniformly increase the size of the coefficient on government expenditure. This table indicates generally that the greatest departures from the pooled OLS estimates occur when the prior estimates are held with great confidence (small δ). This need not have been the case, and is not the case for some of the coefficients.

Table 4 has the Bayes estimates for different values of the experimental bias parameters: λ_1(developed) and λ_2 (developing). Note that the middle column of this table is the same as the middle column in Table 3. For the reasons discussed above, a value of λ equal to one indicates that the prior and data will be equally weighted if the sample size is infinite. A value equal to 10 indicates that 10/11 = 0.91 of the weight is placed on the prior. A value equal to .1 indicates that a weight of 0.1/1.1 = 0.09 is placed on the prior.

The convergence effect is largest when λ_1 is large. Then the developed data are treated as noisy and consequently more weight is put on the relatively large estimate using the developing data. The estimates of the effect of government expenditures vary little with λ. One other pattern that emerges in Table 4 is that the Bayes estimates generally depart more from the pooled OLS estimates as the values of λ increase. Again this seems to repeat the lesson from the previous table. The prior is substantially different from the pooled OLS estimate and changes that tend to put more weight on the prior have more dramatic effects on the estimates.

Table 5 indicates the effects of changing the degree of similarity of the developed and developing regressions. A value of $\rho = 0$ indicates complete dissimilarity and no pooling. A value of $\rho = .99$ indicates that the regression coefficients are virtually identical. The similarity of the estimates does increase with ρ. In the case of the convergence coefficient, it is the coefficient of the developing data set that does the most adjustment. For the government effect, both coefficients get larger with ρ.

This small sea of numbers challenges our cognitive ability. It would be very helpful if the data could focus attention on a subset. Marginal likelihoods offer one tool for focussing. These statistics point to parameters that are most favored by the data, in the same way that likelihood ratios would. The difference is that instead of maximizing the likelihood function, a marginal likelihood is found by integrating the likelihood function using the prior distribution as weights of integration.

Ratios of marginal likelihoods evaluated at different sets of parameter values are called Bayes factors. The Bayes factors comparing different values of the prior parameters with the central case are reported in Tables 6, 7, and 8. These Bayes factors are all defined relative to the central model, which accordingly has the Bayes factor of one in each of the tables. The Bayes factors in Table 6 point to a sharper prior distribution. With the grid of δ-values in the table, the maximum Bayes factor occurs at $\delta = .5$, suggesting that the prior standard errors could be reduced by fifty per cent.

It is more difficult to improve upon the central case by varying either the experimental bias parameters. The Bayes factor in Table 7 is maximized at 56.1 when there is no experimental contamination, $\lambda = 0$, in both data sets. In this display the Bayes factors are slightly higher above than below the diagonal, indicating a very slight preference for relatively accurate developing data set. This result presumably comes from the (relative) similarity of the prior and developing OLS estimates.

The data information about the degree of similarity is the weakest. The Bayes factors in Table 8 grow slightly with the degree of similarity, attaining the level of 2.14 when $\rho = .99$.

Generally a data analysis must combine estimation and sensitivity analysis. When the data are adequately informative, estimation is the preferred approach. These Bayes factors select δ, the prior distrust parameter, with an adequate degree of precision. But neither the experimental bias λ nor the structural similarity can be estimated with adequate accuracy and the sensitivity analysis that was previously discussed comes into play. By limiting the sensitivity analysis to parameter values with adequately high Bayes factors, we can in this case narrow down the estimates very precisely.

3.0 Conclusion

This article proposes and illustrates an econometric method for pooling two noisy data sets. The pooled estimates depend on parameters measuring doubt about the prior, experimental contamination and structural similarity. Implicit or explicit reference must be made to these three concepts regardless of the method that is used to pool data. The explicit mathematical reference to these concepts in this paper is intended to facilitate the discussion about the assumptions that underlie a pooling exercise. The method is illustrated by a study of the determinants of the growth rates of developed and developing countries. These two data subsets yield estimates that are very similar, which makes the results not very sensitive to the form of pooling and which points toward low experimental bias and a high degree of similarity.

ENDNOTES

[1] Support from NSF grant SES 8910950 is gratefully acknowledged. Able computational assistance was provided by Rodrigo Fuentes. The data set has been provided by Robert Barro of Harvard University and the NBER and is available on request.

[2] If there are more than two data sets, additional contamination parameters, λ, would be required. Furthermore, a single correlation parameter ρ indicating the similarity of structures is unlikely to be judged adequate. A variety of higher dimensional correlation matrices might be used but one that has some appeal begins with data sets ordered by their similarity and assigns a correlation depending on the differences in their orders. (Like a stationary time series.)

Reference

Aigner, Dennis J., Hsiao, Cheng, Kapteyn, Arie and Wansbeek, Tom (1984), "Latent Variables in Econometrics," in Handbook of Econometrics, Volume II, eds. Zvi Griliches and Michael D. Intrilligator, Amsterdam: North-Holland Publishing Company, 1321-1393.

Aigner, Dennis and Leamer, Edward E. (1984), "Estimation of Time-of-Use Pricing in the Absence of Experimental Data: An Application of the Methodology of Data Transfer," Journal of Econometrics, 26, 205-227.

Barro, Robert J. (1991), "Economic Growth in a Cross Section of Countries," Quarterly Journal of Economics, 407-443.

DeMouchel, W.H. and Harris, J.E. (1981), "Bayes and Empirical Bayes Methods for Combining Cancer Experiments in Man and Other Species," Technical Report No. 24, Department of Mathematics, M.I.T., February.

Klepper, Steven and Leamer, Edward E. (1984), "Consistent Sets of Estimates for Regressions with Errors in All Variables," Econometrica, 52, 163-183.

Leamer, Edward E. (1974), "False Models and Post-Data Model Construction," Journal of the American Statistical Association, 69, 122-131.

Leamer, Edward E. (1978), Specification Searches, New York: John Wiley and Sons, Inc..

Leamer, Edward E. (1991), "Measurement Errors and the Convergence Hypothesis," Anderson Graduate School of Management.

Patefield, W.M. (1981), "Multivariate Linear Relationships: Maximum Likelihood Estimation and Regression Bounds," Journal of the Royal Statistical Society, Series B, 43, 342-352.

TABLE 1

VARIABLE NAMES and DESCRIPTIVE STATISTICS

NAME	DEFINITION	MEAN	STD. DEV.	MIN	MAX
GR6085	Growth rate of real per capita GDP 1960-1985	0.022	0.019	-0.017	0.074
GDP60	1960 value of real per capita GDP (1980 base year , $1,000)	1.917	1.813	0.208	7.380
SEC60	1960 secondary-school enrollment rate	0.226	0.214	0.010	0.860
PRIM60	1960 primary-school enrollment rate	0.785	0.312	0.050	1.440
GOVEXP	Average from 1970 to 1985 of the ratio of real government cons. (exclusive of defense and education) to real GDP.	0.107	0.053	0.000	0.245
REVCOU	Number of revolutions and coup per year (1960-1985 or sub-sample)	0.180	0.231	0.000	1.150
ASSASS	Number of assassinations per million population per year (1960-1985 or sub-sample)	0.226	0.462	0.000	2.850
PPI60DEV	Magnitude of the deviation 1960 PPP value for the investment deflator from the sample mean	0.001	0.340	-0.494	1.827

TABLE 2

OLS and Prior Estimates

VARIABLES			OLS ESTIMATES	
	PRIOR	POOLED	DEVELOPED	DEVELOPING
GDP60	-0.01 (0.01)	-0.0074 (0.001)	-0.005 (0.001)	-0.011 (0.002)
SEC60	0.1 (0.1)	0.0327 (0.011)	0.025 (0.010)	0.023 (0.020)
PRIM60	0.05 (0.05)	0.0224 (0.006)	0.022 (0.014)	0.027 (0.008)
GOVEXP	-0.01 (0.01)	-0.1332 (0.029)	-0.042 (0.040)	-0.161 (0.034)
REVCOUP	-0.1 (0.1)	-0.0236 (0.007)	-0.019 (0.041)	-0.025 (0.008)
ASSASS	-0.01 (0.01)	-0.0028 (0.003)	0.000 (0.006)	-0.002 (0.004)
PPPIDEV	0 (1)	-0.0067 (0.004)	0.004 (0.013)	-0.008 (0.005)

TABLE 3

Sensitivity Respect to Prior Accuracy
(lambdas = 1, rho = 0.5)

Variables	Prior	d=0.1	d=0.5	d=1	d=2	d=10	OLS
DEVELOPED							
GDP60	-0.01 (0.01)	-0.0092	-0.0077	-0.0076	-0.0075	-0.0075	-0.005 (0.001)
SEC60	0.1 (0.1)	0.0693	0.0561	0.0553	0.0548	0.0550	0.025 (0.010)
PRIM60	0.05 (0.05)	0.0458	0.0432	0.0427	0.0432	0.0428	0.022 (0.014)
GOVEXP	-0.01 (0.01)	-0.0100	-0.0108	-0.0127	-0.0207	-0.0162	-0.042 (0.040)
REVCOUP	-0.1 (0.1)	-0.0768	-0.0555	-0.0530	-0.0505	-0.0517	-0.019 (0.041)
ASSASS	-0.01 (0.01)	-0.0093	-0.0057	-0.0046	-0.0042	-0.0042	0.000 (0.006)
PPPIDEV	0 (1)	-0.0014	-0.0008	-0.0008	-0.0007	-0.0008	0.004 (0.013)
DEVELOPING							
GDP60	-0.01 (0.01)	-0.0096	-0.0086	-0.0085	-0.0086	-0.0085	-0.011 (0.002)
SEC60	0.1 (0.1)	0.0741	0.0527	0.0501	0.0477	0.0487	0.023 (0.020)
PRIM60	0.05 (0.05)	0.0457	0.0456	0.0462	0.0481	0.0470	0.027 (0.008)
GOVEXP	-0.01 (0.01)	-0.0100	-0.0113	-0.0142	-0.0266	-0.0196	-0.161 (0.034)
REVCOUP	-0.1 (0.1)	-0.0608	-0.0497	-0.0492	-0.0478	-0.0486	-0.025 (0.008)
ASSASS	-0.01 (0.01)	-0.0091	-0.0057	-0.0048	-0.0047	-0.0046	-0.002 (0.004)
PPPIDEV	0 (1)	-0.0020	-0.0033	-0.0033	-0.0033	-0.0033	-0.008 (0.005)

TABLE 4

Sensitivity with Respect to Experimental Bias
(rho = 0.5, delta = 1)

	Lambda_1 = 0.1			Lambda_1 = 1			Lambda_1 = 10		
Variables	ambda_2 = 0.1	Lambda_2 = 1	ambda_2 = 10	Lambda_2 = 0.1	Lambda_2 = 1	ambda_2 = 10	Lambda_2 = 0.1	Lambda_2 = 1	ambda_2 = 10
DEVELOPED									
GDP60	-0.0058	-0.0058	-0.0058	-0.0075	-0.0076	-0.0076	-0.0089	-0.0092	-0.0095
SEC60	0.0326	0.0341	0.0353	0.0473	0.0553	0.0628	0.0593	0.0744	0.0900
PRIM60	0.0374	0.0373	0.0373	0.0424	0.0427	0.0428	0.0469	0.0479	0.0486
GOVEXP	-0.0129	-0.0127	-0.0118	-0.0129	-0.0127	-0.0118	-0.0126	-0.0124	-0.0115
REVCOUP	-0.0286	-0.0312	-0.0335	-0.0442	-0.0530	-0.0609	-0.0575	-0.0736	-0.0894
ASSASS	-0.0019	-0.0022	-0.0024	-0.0039	-0.0046	-0.0053	-0.0060	-0.0073	-0.0088
PPPIDEV	0.0001	0.0001	0.0001	-0.0017	0.0008	0.0001	-0.0030	-0.0016	-0.0003
DEVELOPING									
GDP60	-0.0081	-0.0080	-0.0079	-0.0082	-0.0085	-0.0088	-0.0083	-0.0090	-0.0096
SEC60	0.0209	0.0439	0.0636	0.0221	0.0501	0.0764	0.0232	0.0557	0.0891
PRIM60	0.0450	0.0446	0.0440	0.0453	0.0462	0.0465	0.0456	0.0476	0.0492
GOVEXP	-0.0146	-0.0142	-0.0123	-0.0146	-0.0142	-0.0123	-0.0145	-0.0140	-0.0121
REVCOUP	-0.0188	-0.0430	-0.0629	-0.0197	-0.0492	-0.0757	-0.0204	-0.0550	-0.0889
ASSASS	-0.0023	-0.0041	-0.0059	-0.0026	-0.0048	-0.0072	-0.0028	-0.0057	-0.0088
PPPIDEV	-0.0063	-0.0030	0.0004	-0.0064	-0.0033	-0.0005	-0.0065	-0.0035	-0.0006

TABLE 5

Sensitivity Respect to Regression Similarities
(lambdas = 1, delta = 1)

Variables	rho=0	rho=0.5	rho=0.9	rho=0.99
DEVELOPED				
GDP60	-0.0076	-0.0076	-0.0077	-0.0078
SEC60	0.0645	0.0553	0.0490	0.0475
PRIM60	0.0427	0.0427	0.0435	0.0438
GOVEXP	-0.0109	-0.0127	-0.0141	-0.0144
REVCOUP	-0.0627	-0.0530	-0.0467	-0.0454
ASSASS	-0.0055	-0.0046	-0.0042	-0.0041
PPPIDEV	-0.0001	-0.0008	-0.0019	-0.0022
DEVELOPING				
GDP60	-0.0091	-0.0083	-0.0080	-0.0078
SEC60	0.0568	0.0501	0.0475	0.0474
PRIM60	0.0479	0.0462	0.0444	0.0439
GOVEXP	-0.0138	-0.0142	-0.0144	-0.0144
REVCOUP	-0.0563	-0.0492	-0.0458	-0.0453
ASSASS	-0.0059	-0.0048	-0.0042	-0.0041
PPPIDEV	-0.0036	-0.0033	-0.0026	-0.0023

TABLE 6

Bayes Factors for Different
Values of Prior Accuracy

Delta	Bayes Factor
0.10	3.99E-10
0.20	2.35E+00
0.30	2.00E+02
0.40	3.59E+02
0.50	2.00E+02
1.00	1.00E+00
2.00	3.95E-04
10.00	1.94E-13

TABLE 7

Bayes Factors for Different
Values of Experimental Bias

Lambda 1 Lambda 2	0	0.1	1	10
0	56.1084	41.5369	6.3868	0.0295
0.1	41.2674	30.9494	5.0295	0.0244
1	6.0710	4.8100	1.0000	0.0062
10	0.0241	0.0201	0.0053	4.32E-05

TABLE 8

Bayes Factors for Different
Values of Regression Similarity

Rho	Bayes Factor
0.00	0.66
0.50	1.00
0.90	1.80
0.99	2.14