BAYESIAN ELICITATION DIAGNOSTICS

Edward E. Leamer

University of California, Los Angeles

UCLA Dept. of Economics
Working Paper # 555
April 1989

/DIAGNOSTICS/ABSTRACT.DOC

ABSTRACT

Bayesian Elicitation Diagnostics

by
Edward E. Leamer
Department of Economics
UCLA
Los Angeles, CA 90024

(213) 825-3557
Messages: 825-2443

The subject of this paper is elicitation diagnostics that indicate if a prior distribution has to be measured accurately. An elicitation diagnostic forms a question that compares the information in the prior distribution with the information in the given sample. One elicitation diagnostic identifies a family of prior distributions that are so diffuse that they are practically equivalent to the "completely" diffuse prior. Another elicitation diagnostic identifies a family of prior distributions that concentrate enough mass in the neighborhood of zero that they are practically equivalent to the dogmatic prior which sets a parameter exactly equal to zero.

Elicitation diagnostics for the normal linear regression model are reported. The prior distribution is assumed to be normal with mean zero, diffuse on one subset of parameters and with an unknown prior covariance matrix V for the other parameters. The question is whether one can act as if $V = 0$ or, alternatively, as if $V = \infty$. One diagnostic is a matrix V^* such that if the prior covariance matrix satisfies $V < V^*$ then one might as well act as if $V = 0$. Another diagnostic is a matrix V_* such that if the prior covariance matrix satisfies $V_* < V$, then one might as well act as if $V = \infty$.

/DIAGNOSTICS/BAYDIAG.DOC

Bayesian Elicitation Diagnostics

by
Edward E. Leamer
Department of Economics
UCLA
Los Angeles, CA 90024

(213) 825-3557
Messages: 825-2443

1.0 Introduction

The subject of this paper is diagnostic statistics that indicate if a prior distribution has to be measured accurately. If the sample size is small, one might as well set certain parameters to their prior means since the data will contain little information that would suggest other estimates. If the sample size is large, one might as well use a diffuse prior and "let the data speak". But which samples are "small" and which are "large"? It depends on *both* the sample *and* the prior distribution. Regardless of what the sample may be, there are prior distributions that are practically equivalent to the diffuse prior and there are also prior distributions that are practically equivalent to a dogmatic prior. Thus to determine if a sample is small or large, we must ask a question that compares the sample information with the prior information.

An elicitation diagnostic forms a question that compares the information in the prior distribution with the information in the given sample. One elicitation diagnostic identifies a family of prior distributions that are so diffuse that they are practically equivalent to the "completely" diffuse prior. Another elicitation diagnostic identifies a family of prior distributions that concentrate enough mass in the neighborhood of zero that they are practically equivalent to the dogmatic prior which sets a parameter exactly equal to zero. The question that is asked is whether the subject's prior distribution falls in either of these two classes. If an affirmative answer can be given to either of these two elicitation questions then there is no need to go to the expense of a more accurate elicitation of the prior distribution.

Elicitation diagnostics for the normal linear regression model are reported in this paper. The prior distribution is assumed to be normal with mean zero, diffuse on one subset of parameters and with an unknown prior covariance matrix V for the other parameters. The question is whether one can act as if $V = 0$ or, alternatively, as if $V = \infty$. One diagnostic is a matrix V^* such that if the prior covariance matrix satisfies $V < V^*$ then one might as well act as if $V = 0$. Another diagnostic is a matrix V_* such that if the prior covariance matrix satisfies $V_* < V$, then one might as well act as if $V = \infty$.

A plethora of "diagnostics" are the latest addition to the econometric tool kit. The Durbin-Watson statistic and the adjusted R^2 have been used as "diagnostic statistics" for several decades. But recently the list of diagnostics has expanded dramatically to include among others: Godfrey's test of residual serial correlation, Ramsey's RESET test of functional form, Jarque-Bera's test of the normality of regression residuals, tests for heteroscedasticity, the Chow test of the stability of the regression coefficients, Sargan's misspecification test, Sargan's test of serial correlation of instrumental variable residuals, ARCH tests and leverage plots.

The use of diagnostic statistics presents a challenge to statistical theory, classical or Bayesian. Traditional statistical theory deals with the evaluation of <u>planned</u> responses to <u>hypothetical</u>

data sets. Indeed it is impossible to compute sampling properties without a set of plans indicating the response to the data for every conceivable data set. The use of a diagnostic statistic to criticize a model seems to be an advance announcement that the planned response is not fully committed and may be revised when the actual data are observed. This seems to call for a major overhaul of our theories of statistical inference.

But I would argue that there are three different kinds of diagnostics and a different theory is appropriate for each. The three kinds of diagnostics are:

1) Pre-test diagnostics which select between a pair of alternative estimates.
2) Elicitation diagnostics which indicate if the inferences are sensitive to the choice of prior distribution and which call for a more accurate measurement of the state of mind.
3) Criticisms which suggest a "fundamental" change in the model and/or prior distribution.

It is only criticisms that call for a major overhaul of our theories of inference.

Many of the diagnostics that are traditionally employed in the econometrics literature are not used as criticisms that might precipitate an unplanned, unpredictable response to the data. These _"pre-test" diagnostics_ play a part in a complex multi-stage method of estimation of a very general model. A statistic is a pretest diagnostic if both the general model and the response to the data can be fully defined (more-or-less) before the data are observed. The proper evaluation of pretest diagnostics involves either the study of the sampling properties of these complex procedures, or the search for a prior distribution that could justify them. For example, Kennedy and Simons(1989) have studied the estimation of a model with first-order serial correlation with the first step using the Durbin-Watson statistic to determine if adjustment for serial correlation is warranted. They find this method of estimation to be substantially inferior in terms of mean squared error to a Bayesian (one-step) method that integrates the first-order autocorrelation parameter from the likelihood function. Not all responses to data can or should be fully planned because a complete set of plans applicable to every conceivable data set is prohibitively costly to formulate. Plans accordingly will be formulated only for data sets that are regarded to be probable, and responses to improbable data sets will be formulated only if and when they are observed. Contrast for example the scatter of observations in Figures 1 and 2. The R^2's and t-values are the same, but the plan of regressing y on x seems not very good for the scatter in Figure 2 - the message seems to be something else entirely. One would not sensibly have planned for this possibility since it seems so remote, but once these data are observed the original plan of regressing y on x seems highly inappropriate, and cries out for revision.

The actual response to real data can differ from the planned response to hypothetical data either because the initial parameter space is inadequate (as in Figure 2) or because the initial (implicit or explicit) prior distribution is judged inadequate. Diagnostic statistics and data displays can help to stimulate revisions in either the parameter space or the prior distribution. I will call a diagnostic that is intended to suggest a "genuinely" new parameter space a "criticism." A diagnostic that suggests a "revised" prior distribution is also called a "criticism." A diagnostic that suggests that the prior should be more accurately elicited, but which does not change the state of mind of the observer, will be called an "elicitation" diagnostic. The difference between a criticism and an elicitation diagnostic is that a criticism presents a double-counting problem since the data are used both to "formulate" the model and to "estimate" it. An elicitation diagnostic in principle is not subject to this double counting problem though in practice it may be difficult to reveal an aspect of the data set to an observer through the elicitation process without also altering his/her state of mind.

Figure 1

A Probable Scatter

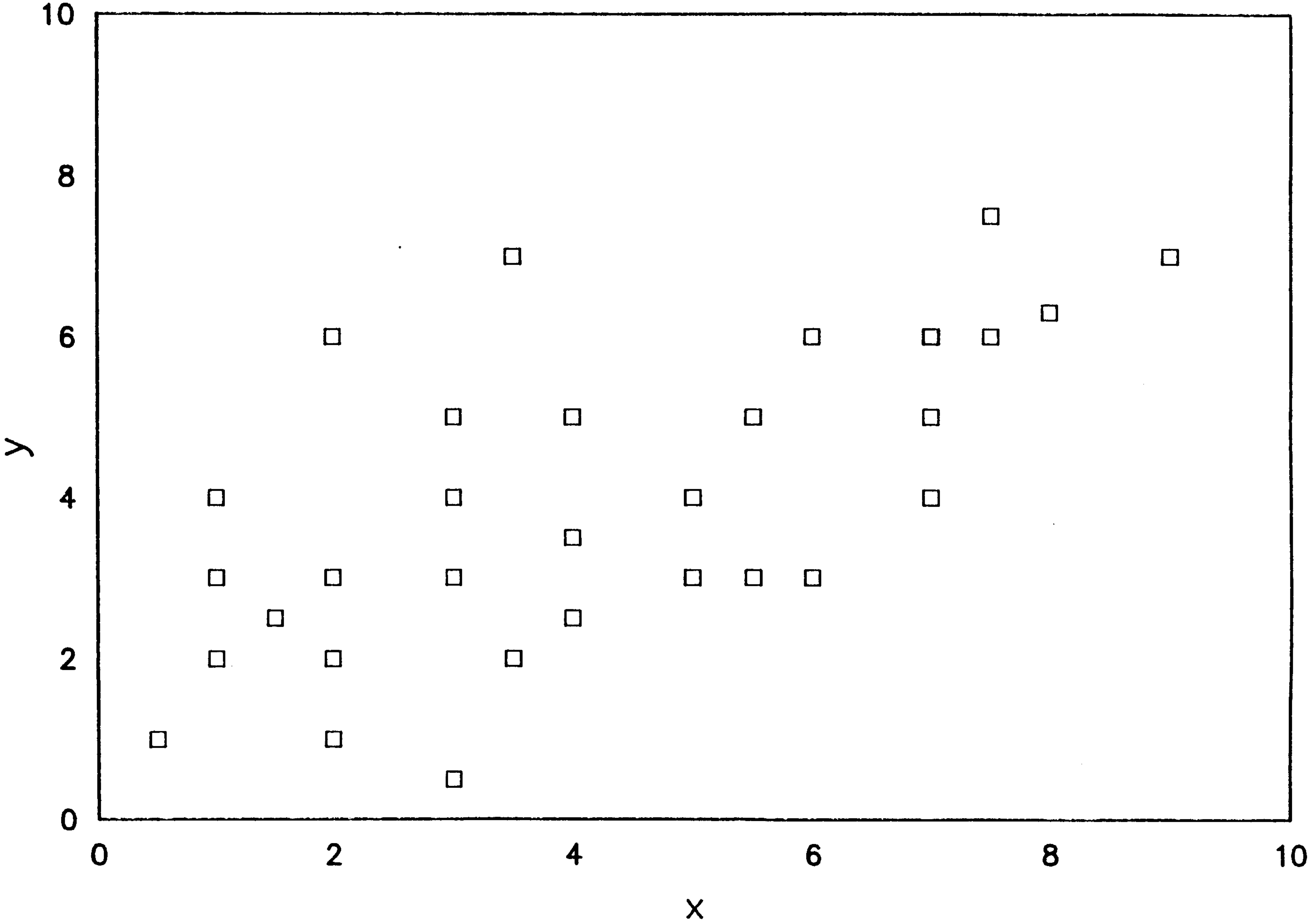
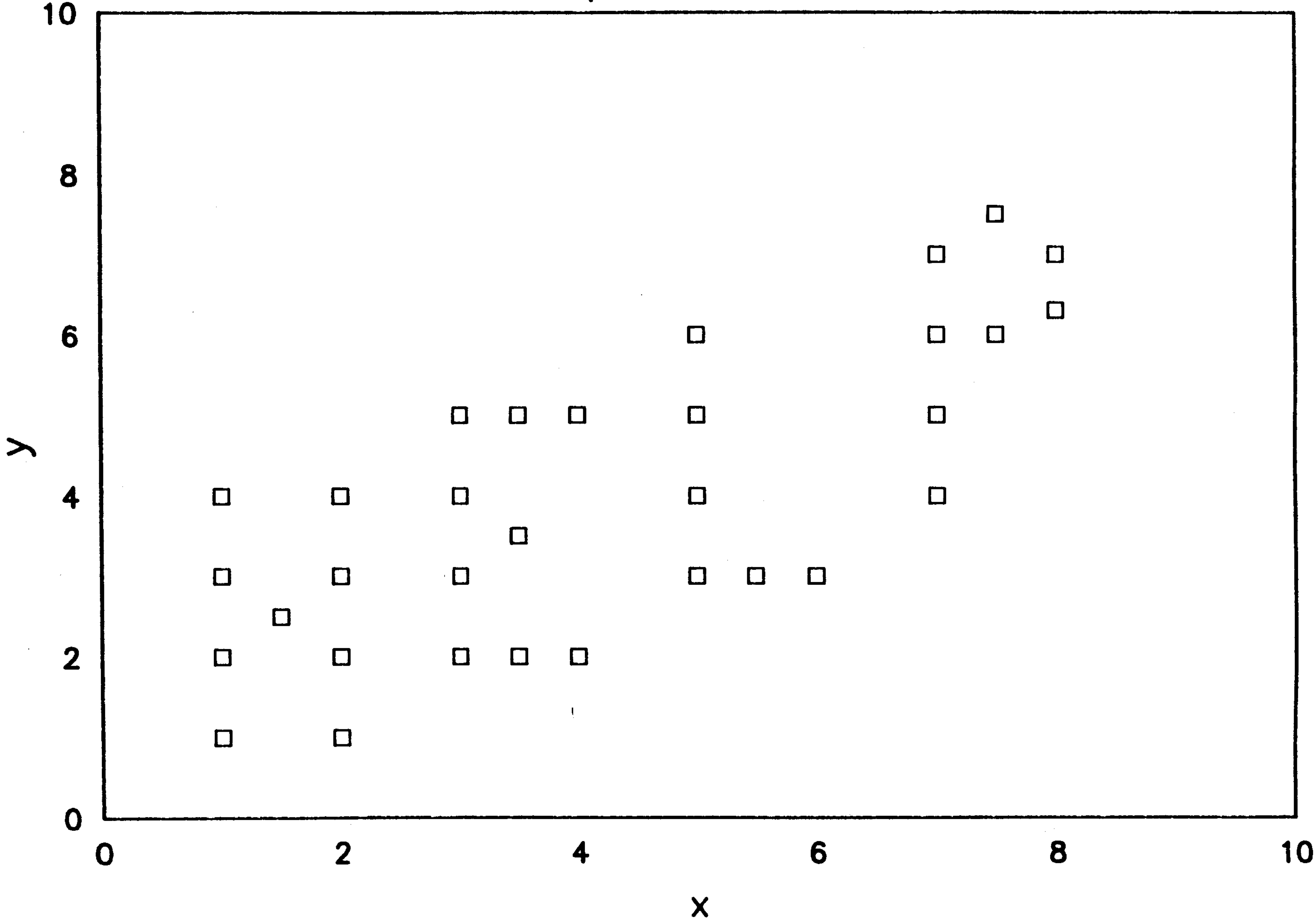


Figure 2

An Improbable Scatter



2.0 Elicitation Diagnostics: General Approach

The elicitation of a prior distribution can be done most efficiently after the data are observed since there are many prior distributions that are practically equivalent to the diffuse prior, and there are many others that are practically equivalent to partially dogmatic priors. An elicitation diagnostic is a statistic measuring the worst-case estimation inaccuracy caused by the use of either the diffuse prior or a partially dogmatic prior as an approximation to the true prior distribution.

A general problem of elicitation can be described in the following way. Suppose that the parameter vector can be divided into a subset β over which the prior distribution can be accurately and costlessly elicited and a subset of "nuisance" parameters γ over which the prior can be elicited only at a cost. Suppose further that the decision problem is to estimate $\theta = G(\beta, \gamma)$ with quadratic loss. Then the optimal estimator of θ is the posterior mean for $G(\beta, \gamma)$. Given the prior distribution $f(\beta, \gamma) = f(\gamma|\beta)\, p(\beta)$ and the data distribution $h(y|\, \beta, \gamma)$ the optimal estimator is

$$\hat{\theta}(f) = E(\, G(\beta, \gamma) \mid y, f) =$$

$$\int G(\beta,\gamma)\, h(y|\beta,\gamma)\, f(\gamma|\beta)p(\beta)\, d\beta d\gamma \,/ \int h(y|\beta,\gamma)\, f(\gamma|\beta)p(\beta)\, d\beta d\gamma ,$$

and the corresponding posterior loss is

$$E(\, (\theta - \hat{\theta})^2 \mid y , f) = \mathrm{Var}(\theta \mid y , f)$$

Note that the posterior loss and the choice of estimator are written explicitly as functionals of f, the conditional prior distribution of the nuisance parameter γ given β. If it is costly to measure f perfectly, one may be willing to use an approximation f* which has the effect of increasing the expected loss because it alters the apparent location of the posterior distribution. The loss if f* is used instead of the true prior f is

$$E(\theta - \hat{\theta}*)^2 \mid y , f) = Var(\theta \mid y , f) + (\hat{\theta} - \hat{\theta}*)^2$$

The percentage increase in expected loss due to approximating f by f* is therefore

$$D(f,f*) = \% \Delta \text{ Loss} = (\hat{\theta} - \hat{\theta}*)^2 / Var(\theta \mid y , f).$$

Here D is a functional that measures the difference between the functions f and f*.

An elicitation problem can then be described in the following way. Suppose that although f is measured inaccurately, it is known that $f \in F$. From this class select an approximation f*. This approximation may minimize the difference D(f,f*) for $f \in F$, or it may minimize the difference D subject to some "simplicity constraint" C(f) = 0. The worst-case cost of mismeasurement of the prior distribution is the maximum percentage reduction in the loss if f is accurately measured:

$$d(F,f*) = \max_{f \in F} D(f,f*)$$

Let the "width" of the family F be

$$\text{Width} = W(F),$$

where W may be a vector function. Suppose further that one were prepared to tolerate values of D that are below some tolerance level τ. Then an elicitation diagnostic is the maximum width of the prior distribution that is consistent with tolerance level τ:

$$\text{Diagnostic}(\tau) = \max_{\{F \mid d(F,f*) \leq \tau\}} W(F)$$

If this diagnostic is large, there is little incentive to measure the prior distribution accurately since the family of distributions that are "essentially" the same as f* is very wide. If this diagnostic is small, the family of priors that are essentially the same as f* is very narrow, and it may be worthwhile to measure the prior more accurately.

The elicitation question, expressed directly, is: Given your tolerance level τ, are you confident that

$$f \in \{F \mid W(F) \leq \text{Diagnostic}(\tau)\} \text{ ?}$$

This query can admit a legitimate answer only if the question is genuinely intelligible, which depends upon the function W that is used to measure the width of the family. What seem like the most natural measures of the width of a family of distributions can cause formidable mathematical problems. The width of a family of distributions could be measured by the couple $W(F) = [\epsilon\ ,\ c]$ where c is the maximum chance that γ is more than an ϵ away from zero

$$c = \max_{f \in F} \text{Prob}_f(\ |\gamma| > \epsilon\ )$$

The corresponding diagnostic is a combination of values for c and ϵ indicating the amount of mass that must be close to zero for one to act as if all the mass of the prior were are at zero. An affirmative answer to the question posed by any pair of these diagnostics then justifies the use of the dogmatic prior that sets γ exactly equal to zero.

It is not generally an easy task to solve for values of $(c,\ \epsilon)$ that imply families of priors that are practically equivalent to a dogmatic prior. Furthermore, if γ is a vector, some arbitrary choice of the measurement of length of a vector $|\gamma|$ would have to be made. For these reasons, it seems clear that some further restrictions would have to be imposed, even though the restrictions need not be fully believed.

In this paper, the families of prior distributions that are used are normal with given means and with intervals of covariance matrices.

These intervals of covariance matrices have a one dimensional representation of the form $\lambda_* S \le V \le \lambda^* S$ where S is the sample covariance matrix and λ is a scalar. The family that is practically equivalent to the dogmatic prior (V=0) has $\lambda_* = 0$. The family that is practically equivalent to the diffuse prior ($V = \infty$) has $\lambda^* = \infty$. Comments are made below about the propriety of this class of distributions.

3.0 Diagnostics for the Simple Linear Regression Model: One Focus Variable and One Doubtful Variable

The framework is first formulated in terms of a simple regression and then extended to the general multivariate case. It is assumed that data are drawn from the following linear regression process:

$$y_t = \alpha + \beta x_t + \gamma z_t + \epsilon_t, \quad t=1,2,\ldots,n$$

where y_t, x_t and z_t are observable variables, α, β and γ are unobservable constants, and ϵ_t is a sequence of unobservables that are drawn independently from a normal distribution with mean zero and known variance which is without loss of generality set to one. In practice, of course, the variance of ϵ is not likely to be known, but the results that follow are approximately correct if the data are divided by the square root of a suitable estimate of the variance.

The parameter of interest is β on which the prior distribution is assumed (for now) to be diffuse. The prior for the coefficient (γ) of the doubtful variable is normal with mean zero variance v. The problem is how to select the prior variance v. If v is set to zero, the z-variable is omitted and a simple regression of y on x is estimated. If v is set to infinity, the z-variable is included and a simple regression of y on x and z is estimated. Other values for v select estimates between these two extremes.

It is conceivable that the value for v could be elicited with an infinite sequence of questions of the form: "Is v in the interval $c_1 \le v \le c_2$?" If this is the only form of elicitation question, then the value of v would be precisely determined only if an affirmative answer were given to this question with $c_1 = c_2$. But, even if an unambiguous affirmative could be given, the probability of selecting $c_1 = c_2 = v$ is zero, and thus no perfectly precise elicitation of v could occur in a finite period of time. Any real data analysis must therefore make use of approximate values for v, either because v is in fact an interval, not a number, or because it is prohibitively expensive to elicit v perfectly. An important question that has not been addressed in the literature on the elicitation of priors is how best to elicit an approximate value for v. This involves both the selection of the sequence of questions (values for c_1 and c_2) and the point at which the elicitation should terminate. Clearly the solution to this elicitation problem should depend on the nature of the data evidence. For example, in settings in which the data are weak, it may be desirable to select the value for v more accurately.

Here I consider only the elicitation questions: "Is $v \le v^*$?" and "Is $v \ge v_*$?" The value of v^* is selected such that an affirmative answer implies that the approximation $v = 0$ is adequate. The value of v_* is selected such that an affirmative answer implies that the approximation $v = \infty$ is adequate. Thus an affirmative answer to the first question implies that the z-variable can be neglected; an affirmative answer to the second question implies that the z-variable should be included. If neither question is answered affirmatively, then the prior distribution would have to elicited more accurately with a series of questions of the form "Is v in the interval $c_1 \le v \le c_2$?" An interesting question that is not further considered here is what values of c_1 and c_2 offer the most expected information from each elicitation question, given the data and the negative answers to the initial queries "Is $v \le v^*$?" and "Is $v \ge v_*$?"

I assume that interest centers on β. The posterior mean and variance of β are calculated in the usual way:

$$\hat{\beta}(v) \equiv E(\beta|\mathbf{y},\mathbf{x},\mathbf{z},v)$$
$$= [\mathbf{x}'\mathbf{y} + (\mathbf{x}'\mathbf{y}\mathbf{z}'\mathbf{z}-\mathbf{x}'\mathbf{z}\mathbf{z}'\mathbf{y})v] \;/\; [\mathbf{x}'\mathbf{x} + (\mathbf{x}'\mathbf{x}\mathbf{z}'\mathbf{z}-\mathbf{x}'\mathbf{z}\mathbf{z}'\mathbf{x})v]$$
$$\hat{V}(v) \equiv Var(\beta|\mathbf{y},\mathbf{x},\mathbf{z},v) = 1 \;/\; [\mathbf{x}'\mathbf{x} - (\mathbf{x}'\mathbf{z})^2/\; (\mathbf{z}'\mathbf{z} + v^{-1}]$$

where **y** , **x**, **z** are data vectors with the means removed and where the dependence of these moments on the prior variance v is made explicit. The question considered here is whether v has to be specified, or can we proceed instead as if v = 0, in which case the variable z is omitted and the following approximate values are used:

$$\hat{\beta}(0) = \mathbf{x}'\mathbf{y}/\; \mathbf{x}'\mathbf{x}$$

$$\hat{V}(0) = 1 \;/\; \mathbf{x}'\mathbf{x}$$

For reference, note that

$$\hat{\beta}(\infty) = (\mathbf{x}'\mathbf{y}\mathbf{z}'\mathbf{z}-\mathbf{x}'\mathbf{z}\mathbf{z}'\mathbf{y}) \;/\; (\mathbf{x}'\mathbf{x}\mathbf{z}'\mathbf{z}-\mathbf{x}'\mathbf{z}\mathbf{z}'\mathbf{x}) = \mathbf{x}'\mathbf{M}_z\mathbf{y} \;/\; \mathbf{x}'\mathbf{M}_z\mathbf{x}$$

$$\hat{V}(\infty) = 1 \;/\; [\mathbf{x}'\mathbf{x} - (\mathbf{x}'\mathbf{z})^2/\; \mathbf{z}'\mathbf{z}\;] = 1/\mathbf{x}'\mathbf{M}_z\mathbf{x}$$

where $\mathbf{M}_z = \mathbf{I} - \mathbf{z}(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}$. Furthermore, we may write the posterior mean as a weighted average of the two extremes:

$$\hat{\beta}(v) = w\;\hat{\beta}(0) + (1-w)\;\hat{\beta}(\infty)$$

where $w = \mathbf{x}'\mathbf{x} \;/\; [\;\mathbf{x}'\mathbf{x} + \mathbf{z}'\mathbf{z}\;\mathbf{x}'\mathbf{M}_z\mathbf{x}\;v] = 1\;/[1 + \mathbf{z}'\mathbf{M}_x\mathbf{z}v]$.

3.1 <u>Tolerance defined in terms of percentage change of coefficient</u>

One might as well set v = 0 if selecting the correct value doesn't make much of a difference. Suppose that the values of v that are practically equivalent to v = 0 are those for which

$$|\hat{\beta}(v) - \hat{\beta}(0)|/|\hat{\beta}(0)| \le \tau$$

where τ is the tolerance level. This interval of values of the prior variance can be written as $v \le v^*$ and v^* can be found by first finding the range of weights

$$|(1-w)(\hat{\beta}(\infty) - \hat{\beta}(0))/\hat{\beta}(0)| \le \tau$$

$$|(1-w)| \le \tau \ |\hat{\beta}(0)/ (\hat{\beta}(\infty) - \hat{\beta}(0))| \equiv \tau / d_0$$

where $1-w = z'M_x zv/ [\ 1+ z'M_x zv]$ and

$$d_0 = |(\hat{\beta}(\infty) - \hat{\beta}(0)| / |\hat{\beta}(0)|$$

Solving for the value of v*: $z'M_x z\ v* = (\tau/d_0)\ [\ 1+ z'M_x zv*]$

$$v* = [(\tau/d_0)/ (1 - (\tau/d_0))\ ]\ \mathrm{Var}(\gamma|v{=}\infty) \qquad \text{if} \quad \tau/d_0 \le 1$$

$$v* = \infty \qquad \text{if} \quad \tau/d_0 > 1$$

where $\mathrm{Var}(\gamma|v{=}\infty) = 1/z'M_x z$ is the least squares variance for γ in the unconstrained model. This interval of prior distributions that are approximately the same as the sharp prior that $v = 0$ is wider if the tolerance τ is great and if the data are relatively uninformative about γ. Thus you might as well omit the variable if the data do not strongly suggest that it belongs.

The forgoing comment suggests that this diagnostic can be related to the size of the t-statistic on the coefficient of z in the full regression. This t-statistic is

$$t_\gamma = (z'M_x y)/(z'M_x z)^{1/2}$$

where $M_x = I - x(x'x)^{-1}x$ Then the difference in the estimates can be written as:

$$\delta = \hat{\beta}(0) - \hat{\beta}(\infty) = \frac{x'y}{x'x} - \frac{x'yz'z - x'zz'y}{x'xz'z - x'zz'x}$$

$$= \frac{x'z\ [\ x'xz'y - x'zx'y\ ]}{x'x\ [\ x'xz'z - x'zz'x\ ]} = \frac{x'z\ z'M_x y}{x'x\ z'M_x z}$$

$$= t_\gamma\ x'z\ /\ x'x\ (z'M_x z)^{1/2} = t_\gamma\ r_{xz}\ /\ (x'M_z x)^{1/2}$$

where $r_{xz} = \mathbf{x'z}/\sqrt{\mathbf{x'xz'z}}$ is the correlation between x and z. The percentage difference in the estimates with the unconstrained estimate as the denominator is then

$$d_\infty = |(\hat\beta(0) - \hat\beta(\infty)) \,/\, \hat\beta(\infty)| = |t_\gamma \; r_{xz} \;\; (\mathbf{x'M_zx})^{1/2} / \; \mathbf{x'M_zy}| = |r_{xz}(t_\gamma/t_\beta)|$$

where t_β is the t-statistic for testing β=0 in the unconstrained regression. Incidentally, this result is the foundation for Leamer's(1975) observation that the sign of a coefficient cannot change when a variable with a less significant t-value is omitted, $|t_\gamma/t_\beta| \le 1$. Using this value for d_∞, it is a simple matter to solve for d_0 [1]

$$d_0 = |(\hat\beta(\infty) - \hat\beta(0)| \,/\, |\hat\beta(0)| = 1 \,/\, |1 + (t_\beta/t_\gamma r_{xz})|$$

A lower bound for the prior variance can be obtained by finding the set of values v that imply essentially the same estimate as v = ∞:

$$|\hat\beta(v) - \hat\beta(\infty)|/|\hat\beta(\infty)| \le \tau$$

This inequality can be written as

$$|w(\hat\beta(\infty) - \hat\beta(0))/\hat\beta(\infty)| \le \tau$$

which can be manipulated as:

$$|r_{xz}(t_\gamma/t_\beta)| \le \tau[\; 1 + \mathbf{z'M_xz}v]$$

Thus

$$v_* = [(d_\infty/\tau). - 1] \; \mathrm{Var}(\gamma|v{=}\infty) \qquad \text{if} \;\; d_\infty/\tau \ge 1$$

$$v_* = 0 \qquad \text{if} \;\; d_\infty/\tau < 1$$

The width of this set of priors that are practically to the diffuse prior increases with the tolerance level τ and increases as the data get more informative about γ and Var(γ|v=∞) gets smaller.

[1] Using $|(x-y)/y| = |1/(1+[x/(y-x)])|$

In summary: (a) When the sample size is small, one might as well just omit the confounder variable z. The parameter γ is thought probably to be small, and in the absence of data evidence to the contrary, one might as well set the estimate to zero. (b) When the sample size is large, ordinary regression on the full model is likely to be a good approximation. Though there exists prior information that γ is small, when the data information is rich enough, this prior can be ignored. (c) Thus, the precise value for v has to be selected only for intermediate sample sizes.

Incidentally, it is possible to have $v_* \leq v^*$, in which case, whatever the value of v, it can be rounded off either to 0 or to ∞, or possibly to either one. For that matter, if the tolerance is great and the difference between the constrained and unconstrained estimates is small, then it is possible to have $v_* = 0$ and $v^* = \infty$, in which case it doesn't matter which estimate in the interval of possible estimates is selected.

3.2 Tolerance relative to the standard error

The foregoing defines the tolerable level of sensitivity in terms of the percentage change in the estimate. Another approach compares the change in the estimate with its standard error. This can be justified formally be referring to a problem of estimating β with quadratic loss. The expected posterior loss is then:

$$\text{Exp. Loss} = E[(\beta - \hat{\beta})^2 | \mathbf{x},\mathbf{z},\mathbf{y},v] = \hat{V}(v) + (\hat{\beta}(v) - \hat{\beta})^2$$

If one were prepared to tolerate a 100 τ^2 per cent increase in the expected loss, one could act as if v = 0 if

$$(\hat{\beta}(v) - \hat{\beta}(0))^2 / \hat{V}(v) \leq \tau^2$$

This set of values for v is:

loosely speaking, compares the size of the t-value of the included variable with t-value of the doubtful variable.

We can also find the set of priors that are equivalent to the diffuse prior $v = \infty$. If one were prepared to tolerate a 100 τ^2 per cent increase in the expected loss, one could act as if $v = \infty$ if

$$(\hat{\beta}(v) - \hat{\beta}(\infty))^2 / \hat{V}(v) \leq \tau^2$$

This set of values for v is:

$$w^2 (\hat{\beta}(\infty) - \hat{\beta}(0))^2 / \hat{V}(v) \leq \tau^2$$

which again is a quadratic inequality in v:

$$\delta^2 \mathbf{x}'\mathbf{x} \leq \tau^2 (\mathbf{z}'\mathbf{z}v + 1)(\mathbf{z}'M_x\mathbf{z}v + 1)$$

As above, it is conservative to use an underestimate of the variance with v=0, in which case the inequality becomes

$$\delta^2 \mathbf{x}'\mathbf{x} \leq \tau^2 (\mathbf{z}'M_x\mathbf{z}v + 1)^2$$

which implies:

$$v_* = \mathrm{Var}(\gamma | v{=}\infty)[s/\tau - 1] \quad \text{if } \tau/s \leq 1$$

$$v_* = 0 \quad \text{if } \tau/s > 1$$

where $s = |\delta| (\mathbf{x}'\mathbf{x})^{1/2} = |t_\gamma \, r_{xz}| / (1 - r_{xz}^{\;2})^{1/2}$.

Note that if $\tau/s > 1$ then $v_* = 0$ and $v^* = \infty$, which means that the tolerance is sufficiently high relative to the difference in the estimates that it makes only an unimportant difference whether the constrained or unconstrained estimate is selected, or anything in between.

4.0 Multivariate Cases

The multivariate case is more difficult because the minimal set of prior distributions that are practically equivalent to either the diffuse or the dogmatic prior distributions cannot be described in terms of a finite set of interpretable diagnostics. The solution that is proposed here is to assume that the prior covariance matrix is proportional to the sample covariance matrix. The set of prior distributions that are practically equivalent to either the diffuse prior or the dogmatic prior is then described in terms of intervals of values of the scale factor that multiplies this prior covariance matrix. A researcher is expected to be able to answer questions of the form: "Can you comfortably assert that the prior variance of any linear combination of parameters is smaller than λ^* times its sample variance, or can you comfortably assert that the variance any linear combination of parameters is larger than λ_* times its sample variance?" Because these questions refer to any linear combination of parameters, these apparently simple questions are really an infinite number of questions. I am not certain that this infinite set of questions can be answered in a finite time, but I will proceed as if it could.

First we consider the case in which the prior distribution is diffuse on a subset of coefficients. Assume that the $n \times 1$ vector $\mathbf{y}$ is distributed normally with mean $\mathbf{X}\beta + \mathbf{Z}\gamma$ and covariance matrix $\mathbf{I}$, where $\mathbf{X}$ is an $n \times q$ observable matrix , $\mathbf{Z}$ is an $n \times p$ observable matrix, β is a $q \times 1$ vector of unobservables and γ is a $p \times 1$ vector of unobservables. Suppose, for now, that the prior distribution for β is diffuse and the prior distribution for γ is normal with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{V}$.

To summarize, the model is

$y \sim N(X\,\beta + Z\,\gamma,\ I)$

$\beta \sim N(0,\ \infty)$

$\gamma \sim N(0, V)$

Then let the unconstrained estimate of β and γ, and the corresponding precision matrices be:

$$b = (X'M_zX)^{-1}\ X'M_zy$$

$$g = (Z'M_xZ)^{-1}\ Z'M_xy$$

$$H_\beta = X'M_zX,$$

$$H_\gamma = Z'M_xZ,$$

where

$$M_z = I - Z(Z'Z)^{-1}Z'$$

$$M_x = I - X(X'X)^{-1}X'.$$

And let the regression of the z-variables on the x-variables be:

$$P = (X'X)^{-1}X'Z.$$

Then the posterior mean of β can be written as:

$$\hat{\beta}(V) = (X'X)^{-1}X'(y - Z\hat{\gamma}(V)) = \hat{\beta}(0) - P\,\hat{\gamma}(V).$$

where the posterior mean for γ is

$$\hat{\gamma}(V) = (Z'M_xZ + V^{-1})^{-1}\ Z'M_xy.$$

The corresponding posterior covariance matrix is

$$\mathrm{Cov}(\beta,\gamma|V) = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + V^{-1} \end{bmatrix}^{-1}$$

The issue of interest is assumed to be the linear combination $\psi'\beta + \eta'\gamma$ where ψ and η are vectors of constants. The posterior mean of this linear combination can be written as

$$\psi'\hat{\beta}(V) + \eta'\hat{\gamma}(V) = \psi'\hat{\beta}(0) + (\eta' - \psi'P)\ \hat{\gamma}(V) = \psi'\hat{\beta}(0) + \phi'\hat{\gamma}(V)$$

where

$$\phi' = (\eta' - \psi' P).$$

Observe further that the dogmatic prior sets the estimate of γ to zero, and the estimate of the issue is:

$$\psi'\hat{\beta}(0) + \eta'\hat{\gamma}(0) = \psi'\hat{\beta}(0)$$

For later reference, we will need the following results:

1) The unconstrained estimate of β is

$$\mathbf{b} \equiv \hat{\beta}(\infty) = \hat{\beta}(0) - \mathbf{P}\,\mathbf{g} \qquad (1)$$

2) The χ^2 statistic for testing $\gamma = 0$ is

$$\chi^2_\gamma = \mathbf{g}'\mathrm{H}_\gamma \mathbf{g}$$

3) The t-statistic for testing $\psi'\beta = 0$ is

$$t_\psi = \psi'\mathrm{b}/[\psi'\mathrm{H}_\beta^{-1}\psi]^{1/2}.$$

4) The t-statistic for testing $\phi'\gamma = (\eta' - \psi'\mathbf{P})\gamma = 0$ is

$$t_\phi = \phi'\mathbf{g}\;/[\phi'\mathrm{H}_\gamma^{-1}\phi]^{1/2}.$$

5) The t-statistic for testing the "Hausman hypothesis" that $\psi'\beta(0)$ is unbiased, $-\psi'\mathrm{P}\;\gamma = 0$, is

$$t_H = \psi'\mathrm{P}\mathbf{g}\;/[\psi'\mathrm{PH}_\gamma^{-1}\mathrm{P}\psi]^{1/2}.$$

Note that if the issue does not involve the omitted variable, that is if $\eta = 0$, then $t_\phi = t_H$. Further, using the standard rule for the inverse of this form we have:

$$(\mathbf{X}'\mathbf{M}_z\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1}$$
$$= (\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{M}_x\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Thus:

$$\mathrm{PH}_\gamma^{-1}\mathrm{P}' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}[\mathbf{Z}'\mathbf{M}_x\mathbf{Z}]^{-1}\mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{M}_z\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1}\;.$$

4.1 Partially Informative Priors Equivalent to the Dogmatic Prior V = 0.

The first problem that is considered is to find a set of prior distributions that are practically equivalent to the dogmatic prior

$V = 0$ in the sense of implying an estimate that is indistinguishable from the estimate obtained if the z-variables are omitted. The following theorem dealing with the set of estimates corresponding to a family of prior distributions with covariance matrix bounded from above $V \leq V^*$ is taken from Leamer(1982):

Theorem 1: The extreme values of $\psi'\hat{\beta}(V) + \eta'\hat{\gamma}(V)$ for $V \leq V^*$ are:

$$\psi'\hat{\beta}(0) + \phi'[H_\gamma + V^{*-1}]^{-1}H_\gamma g/2$$
$$\pm [g'H_\gamma[H_\gamma + V^{*-1}]^{-1}H_\gamma g]^{1/2}[\phi'[H_\gamma + V^{*-1}]^{-1}\phi]^{1/2}/2 \qquad (2)$$

where $\phi' = (\eta' - \psi'P)$

The problem now is to find a "large" value for V^* such that this interval of estimates is practically equivalent to a point. A "large" value can be found by selecting V^* up to a proportionality constant $V^* = \lambda^* V_0$, and then finding the maximal value for λ^* such that the interval of estimates(2) is adequately short. There are of course an infinity of possible values for V_0. This is what makes the multivariate case more difficult than the univariate case. One possibility is that the researcher sensibly and economically selects a personal value for V_0. When it takes a lot of effort to think about V_0, it may be useful to use the sample precision matrix $V_0 = H_\gamma^{-1}$. In that event the interval (2) takes a very convenient form.

Theorem 1': The extreme values of $\psi\hat{\beta}(V) + \eta'\hat{\gamma}(V)$ for $V \leq \lambda^* H_\gamma^{-1}$ are:

$$\psi'\hat{\beta}(0) + \phi' g\lambda^*/2(1+\lambda^*)$$
$$\pm [g'H_\gamma g]^{1/2}[\phi' H_\gamma^{-1}\phi]^{1/2}\lambda^*/2(1+\lambda^*) \qquad (2')$$

The first proposed diagnostic is the maximum value of λ^* such that the posterior mean is essentially the same as the constrained estimate for all V satisfying $V \le \lambda^* H_\gamma^{-1}$ where "essentially the same" is defined as making a percentage difference that is less than some tolerance τ:

$$|\psi'\hat{\beta}(V) + \eta'\hat{\gamma}(V) - \psi'\hat{\beta}(0)|/|\psi'\hat{\beta}(0)| \le \tau .$$

Using (2′) we may solve for the extreme values of this tolerable difference as $[\lambda^*/(1+\lambda^*)]s$ where

$$s = (|\phi' g| + [\chi^2_\gamma]^{1/2}[\phi' H_\gamma^{-1}\phi]^{1/2}) /2|\psi'\hat{\beta}(0)|$$

$$= (|t_\phi| + [\chi^2_\gamma]^{1/2}) [\phi' H_\gamma^{-1}\phi]^{1/2}/2|\psi'\hat{\beta}(0)|$$

Then the diagnostic becomes:

$$\lambda^*(\psi,\eta) = (\tau/s) / (1 - (\tau/s)),$$

where I have indicated the dependence of this diagnostic on (ψ,η) to make clear that it depends on the issue of interest. The value of this diagnostic will be large, and the corresponding family of prior distributions will be wide, when s is small. This will occur when the γ parameters are insignificant, collectively and as the linear combination $\phi'\gamma$. Also s will be small when the constrained estimate of the issue is large compared with the unconstrained sampling standard error of $\phi'\gamma$ ($[\phi' H_\gamma^{-1}\phi]^{1/2}$).

Note, incidentally, that if $\psi = 0$, then $s = \infty$ and $\lambda^* = 0$, which in words means that there are no priors other than the dogmatic prior that are equivalent to the dogmatic prior. This is correct because the dogmatic prior in this case, with $\psi = 0$, implies a zero estimate of the issue, and any other prior will imply an estimate that is proportionately different from zero by an infinite amount.

4.2 Partially Informative Priors Equivalent to the Diffuse Prior $V = \infty$.

The next problem is to find a set of prior distributions that are equivalent to the diffuse prior in which case the unconstrained estimate can be used. We will make use of the following theorem taken from Leamer(1982):

Theorem 2: The extreme values of $\psi'\hat{\beta}(V) + \eta'\hat{\gamma}(V)$ for $V_* \leq V$ are:

$$\psi'\hat{\beta}(0) + \phi'[H_\gamma + V_*^{-1}]^{-1}[H_\gamma + V_*^{-1}/2]g$$
$$\pm [g'[H_\gamma^{-1} + V_*]^{-1}g]^{1/2}[\phi'[H_\gamma V_* H_\gamma + H_\gamma]^{-1}\phi]^{1/2}/2 \qquad (3)$$

For the case of a prior precision matrix proportional to the sample precision this specializes to:

Theorem 2': The extreme values of $\psi\hat{\beta}(V) + \eta'\hat{\gamma}(V)$ for $\lambda_* H_\gamma^{-1} \leq V$ are:

$$\psi'\hat{\beta}(0) + \phi' g(\lambda_*+1/2)/(1+\lambda_*) \pm [g'H_\gamma g]^{1/2}[\phi' H_\gamma^{-1}\phi]^{1/2}/2(1+\lambda_*) \qquad (3')$$

We now select the smallest value for λ_* such that for all V satisfying $\lambda_* H_\gamma^{-1} \leq V$ the estimate is essentially the same as the estimate with the diffuse prior $V = \infty$:

$$|\psi'\hat{\beta}(V) - \psi'\hat{\beta}(\infty) + \eta'\hat{\gamma}(V) - \eta'\hat{\gamma}(\infty)\ |/|\psi'\hat{\beta}(\infty) + \eta'\hat{\gamma}(\infty)\ | \leq \tau$$

where τ is the tolerance level. Using (3') and (1) we may solve for the extreme values of this tolerable difference $[1/(1+\lambda_*)]$ s where

$$s = (\ |\phi' g| \pm [\chi^2_\gamma]^{1/2}[\phi' H_\gamma^{-1}\phi]^{1/2}|\ )/2|\psi' b + \eta' g\ |$$
$$= (|t_\phi| + [\chi^2_\gamma]^{1/2}|\ )\ [\phi' H_\gamma^{-1}\phi]^{1/2}/2|\psi' b + \eta' g\ |$$

The extreme value of λ_* is the solution to $[1/(1+\lambda_*)]\ s = \tau$. Thus

$$\lambda_* = s/\tau - 1.$$

The value for s further simplifies if $\eta = 0$, that is if the issue of interest involves only the coefficients on which the prior is diffuse.

Let

$$m = \psi' PH_\gamma^{-1}P'\psi / \psi'(X'M_zX)^{-1}\psi$$

$$= 1 - \psi'(X'X)^{-1}\psi/\psi'(X'M_zX)^{-1}\psi$$

Then

$$s = \; = m^{1/2} \left(|t_H| + [\chi^2_\gamma]^{1/2}| \right) / 2 \; |t_\psi|$$

The value of λ_* will be small and the corresponding set of prior distributions will be large when s is small and the tolerance τ is large.

4.3 <u>Fully Informative Priors Equivalent to the Diffuse Prior $V = \infty$.</u>

The next result allows the prior to be informative on <u>all</u> the parameters, not just a subset. The question that is raised is whether the diffuse prior, $V = \infty$, is a good approximation in the sense that the posterior mean is practically the same as the unconstrained ordinary regression. It is assumed that the $n \times 1$ vector **y** is distributed normally with mean $X\beta$ and covariance matrix **I**, where **X** is an $n \times q$ observable matrix, and β is a $q \times 1$ vector of unobservables. It is also assumed that the prior distribution for β is normal with mean vector 0 and covariance matrix **V**. Then the posterior mean for β is

$$\hat{\beta}(V) = (X'X + V^{-1})^{-1}X'y.$$

The following theorem from Leamer(1982) uses the notation:

$$H = X'X$$

$$b = (X'X)^{-1}X'y.$$

Theorem 3: The extreme values of $\psi'\hat{\beta}(V)$ for $V_* \le V$ are:

$$\psi' f \pm [c]^{1/2}[\psi' A^{-1}\psi]^{1/2} \qquad (4)$$

where $f = (HV_*H + H)^{-1}(HV_*Hb + Hb/2)$

$$c = b'V_*^{-1}(H + V_*^{-1})^{-1}Hb/4$$

$$A = H\,V_*\,H + H$$

This specializes when the prior precision is proportional to the sample precision to:

Theorem 3′: The extreme values of $\psi'\hat{\beta}(V)$ for $\lambda_* H^{-1} \le V$ are:

$$\psi' b(\lambda_* + 1/2)/(1+\lambda_*) \pm [b'Hb]^{1/2}[\psi' H^{-1}\psi]^{1/2}/2(1+\lambda_*) \qquad (4')$$

A minimal value for λ_* is selected such that for all V satisfying $\lambda_* H^{-1} \le V$

$$|\psi'\hat{\beta}(V) - \psi' b|/|\psi' b| \le \tau$$

where τ is the tolerance level. Using (4′) we may solve for the extreme values of this tolerable difference

$$\lambda_* = s/\tau - 1 \quad , \text{ where}$$

$$s = [\ 1 + [\chi^2_\beta]^{1/2}/|t_\psi|\ ]\ /\ 2$$

Thus, using this notion of sensitivity, the prior can be treated as if it were completely diffuse if interest focuses on parameters with large t-values. (I leave as an exercise the problem of finding the linear combination with the maximum t-value.)

4.4 <u>Tolerance relative to the standard error</u>

The other kind of diagnostic compares the difference in the estimates with their standard error. These are straightforward variations of the diagnostics already considered and can be justified with reference to the problem of estimation with quadratic loss that has been discussed earlier.

The first diagnostic for the partially informative case is the maximum value of λ^* such that the posterior mean is essentially the same as the constrained estimate for all V satisfying $V \le \lambda^* H_\gamma^{-1}$:

$$(\psi'\hat{\beta}(V) + \eta'\hat{\gamma}(V) - \psi'\hat{\beta}(0))^2 / (\psi,\eta)'\mathrm{Cov}(\beta,\gamma|V)(\psi,\eta) \le \tau^2 \qquad (5)$$

where τ is the tolerance level. Solving this problem and the ones analogous to it for the other cases seems very difficult, but a narrower set of priors that also satisfy this inequality can be found by replacing the denominator with the variance of the linear combination corresponding to the sharp prior **V** = 0. Paralleling the results above we may solve for the extreme values of this tolerable difference as $[\lambda^*/(1+\lambda^*)]s$ where

$$s = (|t_\phi| + [\chi^2_\gamma]^{1/2})\ [\phi' H_\gamma^{-1}\phi]^{1/2}/2\sqrt{(\psi,\eta)'\mathrm{Cov}(\beta,\gamma|\mathbf{V{=}0})(\psi,\eta)}$$
$$= (|t_\phi| + [\chi^2_\gamma]^{1/2})\ [\phi' H_\gamma^{-1}\phi]^{1/2}/2\ (\psi'(X'M_zX)^{-1}\psi)^{1/2}$$

Then the diagnostic becomes:

$$\lambda^*(\psi,\eta) = (\tau/s)\ /\ (1 - (\tau/s)).$$

Note that if the linear combination of interest does not involve the doubtful variables, that is if $\eta = 0$, then

$$s = (|t_H| + [\chi^2_\gamma]^{1/2})\ [\psi'\mathbf{P}H_\gamma^{-1}\mathbf{P}\psi]^{1/2}/2\sqrt{\psi'(X'M_zX)^{-1}\psi}$$
$$= (|t_H| + [\chi^2_\gamma]^{1/2})\ m^{1/2}/2$$

where $m = 1 - \psi'(\mathbf{X'X})^{-1}\psi/\psi'(\mathbf{X'M_zX})^{-1}\psi$ and t_H is the t-statistic for testing the Hausman hypothesis $\psi'\mathbf{P}\gamma = 0$.

Note also that if the linear combination involves only the doubtful variables, $\psi = 0$, then this value of s is ∞ and $\lambda^* = 0$, meaning that there are no prior distributions that are equivalent to the dogmatic prior. This is formally appropriate because if one sets **V** = 0, then one acts as if the linear combination of interest were known perfectly, and relative to this zero standard error any small change in the estimate is unlimited in importance. This result, however, raises questions concerning the tolerance definition in terms of the percentage increase in the loss associated with this one decision since there may be many other decisions that cause so much risk that this additional risk is inconsequential even though a small change in the estimate would imply an infinite per centage increase in this risk. Further, if the problem were forecasting or control rather than estimation, the size of the residual variance affects the expected loss and setting a parameter to zero can never increase the expected loss by an infinite percentage. More on this subsequently.

The second diagnostic for the partially informative case is the smallest value for λ_* such that for all V satisfying $\lambda_* H_\gamma^{-1} \leq V$ the estimate is essentially the same as the estimate with the diffuse prior $V = \infty$ in the sense that inequality (5) is satisfied. If the denominator of (5) is not zero when $V = 0$, then a family of priors that is equivalent to the diffuse prior can be found by setting the variance in the denominator to the smallest possible which occurs when $V = 0$. Using the result above, this implies the value

$$\lambda_* = s/\tau - 1.$$

where

$$s = (|t_\phi| + [\chi^2_\gamma]^{1/2}|)\ [\phi' H_\gamma^{-1}\phi]^{1/2}/2((\psi,\eta)'\mathrm{Cov}(\beta,\gamma|V=0)(\psi,\eta))^{1/2}$$

The value for s further simplifies if $\eta = 0$, that is if the issue of interest involves only the coefficients on which the prior is diffuse. Then following the algebra above we have

$$s = m^{1/2}\ (|t_H| + [\chi^2_\gamma]^{1/2}|)\ /2\ .$$

Note that if the issue in question involves only the doubtful variables, that is if $\psi=0$, then this value of s is infinite and $\lambda_* = \infty$, meaning that there is no prior that is practically equivalent to the diffuse prior. But this solution is inappropriate because it allows the variance in the denominator to go to zero when in fact it is bounded away from zero. A better solution would be to define the class of priors by separately optimizing the numerator and denominator over the class of priors $\lambda_* H_\gamma^{-1} \leq V$

$$\max\ (\psi'\hat{\beta}(V) + \eta'\hat{\gamma}(V) - \psi'\hat{\beta}(0))^2 / \min\ (\psi,\eta)'\mathrm{Cov}(\beta,\gamma|V)(\psi,\eta) \leq \tau^2$$

The denominator is minimized at $V = \lambda_* H_\gamma^{-1}$ since for $V_1 < V_2$, $\mathrm{Cov}(\beta,\gamma|V_1) < \mathrm{Cov}(\beta,\gamma|V_2)$. From the discussion above, the numerator of this expression is

$$\max\ (\psi'\hat{\beta}(V) + \eta'\hat{\gamma}(V) - \psi'\hat{\beta}(0))^2 = (|t_\phi| + [\chi^2_\gamma]^{1/2}|)\ [\phi' H_\gamma^{-1}\phi]^{1/2}\ /\ 2(1+\lambda_*)$$

The denominator is a relatively complicated function of λ_*, except when the issue does not involve the diffuse parameters, $\psi = 0$. Then the denominator becomes

$$\eta'(H_\gamma + H_\gamma/\lambda_*\ )^{-1}\ \eta' = \eta'(H_\gamma\ )^{-1}\eta\ \lambda_*/(1+\lambda_*)$$

Dividing the previous expression by this, using $\phi = \eta$, implies a value

$$\lambda_* = s\ /\ \tau$$

where

$$s = (|t_\eta| + [\chi^2_\gamma]^{1/2}|)/2$$

Another diagnostic can be found if the prior covariance is restricted to be proportional to the sample covariance matrix, $V = \lambda H^{-1}$. (This is different from the solutions above which only impose an inequality that makes reference to the sample covariance matrix, and do not assume that the prior covariance matrix is known up to a factor of proportionality.) With $V = \lambda H^{-1}$, we have

$$\hat{\beta}(\lambda) = b\ \lambda\ /\ (\ 1 + \lambda)$$

$$\mathrm{Cov}(\beta) = H^{-1}\ \ \lambda\ /\ (\ 1 + \lambda)$$

and

$$|\psi'\hat{\beta}(\lambda) - \psi' b|/(\psi'\mathrm{Cov}(\beta)\psi)^{1/2} = [\psi' b/(\psi' H^{-1}\psi)^{1/2}\ ]\ /\ [\ \lambda\ (\ 1 + \lambda\ )\ ]^{1/2}$$

Thus λ_* is the solution to

$$t_\psi^2 / \tau^2 = \lambda_*\ (\ 1 + \lambda_*)$$

This quadratic has the relevant solution

$$\lambda_* = -1/2 + (1 + 4t_\psi^2/ \tau^2)^{1/2}/2$$

4.5 Summary

All of these results are summarized in three tables. Table 1 contains the assumptions and various summary statistics. Table 2 contains the diagnostics that indicate families of priors that are equivalent to the dogmatic prior which implies that variables are omitted from the equation. Table 3 contains the diagnostics that indicate the range of prior distributions that are equivalent to the diffuse prior which implies estimates equal to the unconstrained ordinary least squares estimates.

What you see in these tables is that the traditional "misspecification" test statistics, namely the χ-squared statistic for the omitted variables and the Hausman statistic for zero bias, do play an important role in determining the values of these elicitation diagnostics. But none of these elicitation diagnostics suggests a simple study of the size of either of these misspecification test statistics alone. My inference is that the diagnostic statistics that we are traditionally using are probably the right thing to be examining, but we are not using them in the most appropriate way, certainly not if they are used implicitly as elicitation diagnostics.

5.0 An Example

The preceeding algebra is probably best understood by way of an example. For want of a better data set, I will reanalyze the murder data for which I previously reported bounds in Leamer(1982) and Leamer(1983). These bounds answer the question: If the prior distribution has certain properties, what properties must the posterior distribution have? In this paper I turn this question around to form: If the posterior distribution has certain properties, what properties must the prior distribution have?

The variables are listed in Table 4. The dependent variable is the murder rate per hundred thousand population, observed for 44 states in 1950. The explanatory variables are divided into three sets. There are four deterrent variables that characterize the criminal justice system, or in economic parlance, the expected out-of-pocket cost of crime. There are four economic variables that measure the opportunity cost of crime. And there are four social/environmental variables that possibly condition the taste for crime. This leaves unmeasured only the expected rewards for criminal behavior, though these are possibly related to the economic and social variables and are otherwise assumed not to vary from state to state.

Individuals with different experiences and different training will find different subsets of the variables to be candidates for omission from the equation. Five different lists of doubtful variables are reported in Table 5. A right winger expects the punishment variables to have an effect, but treats all other variables as doubtful. He wants to know whether the date still favor the large deterrent effect, if he omits some of these doubtful variables. The rational maximizer takes the variables that measure the expected economic return of crime as important, but treats the taste variables as doubtful. The eye-for-an-eye prior treats all variables as doubtful except the probability of execution. An individual with the bleeding heart prior sees murder as the result of economic impoverishment. Finally, if murder is thought to be a crime of passion then the punishment variables are doubtful.

The prior elicitation questions that will be posed subsequently refer to the least-squares standard errors reported in the first column of Table 6. These elicitation questions compare the standard errors of whatever prior information you may have with these sample standard errors. In order to facilitate the introspection that is necessary to answer these elicitation questions, Table 6 contains an interpretation of each standard error based on the assumption that the estimate is equal to a multiple of its standard error. The number that is reported is the amount by which the explanatory variable would have to change in order for the murder rate to change by 1/100,000. For example, the estimated standard error for PC, the conditional probability of conviction, is 4.0. An estimate of 4.0 implies that an increase in the conditional probability of execution by 1/4 = .25 changes the murder rate by 1/100,000. This number can be found in the column headed by a multiplier equal to one. If the estimate were ten times as large as the standard error, it would take only a 1/40 = .025 increase in the conditional probability of conviction to decrease the murder rate by 1/100,000. This number can be found in the column headed by a multiplier equal to ten. The last column in Table 6 contains the units of the variables, most of which are fractions(f).

The elicitation questions implicit in the diagnostics soon to be discussed select a posterior distribution and identify a family of prior distributions that are mapped by the data into it. Before discussing these diagnostics it is useful for purposes of contrast to do the mapping in the other way, that is to find an "agreeable" family of prior distributions and the corresponding family of posterior distributions. Consider, for example, the problem of selecting a family of prior distributions to characterize your uncertainty about the effect of the conviction probability on the murder rate. First find a lower bound for the prior variance: If the prior location is zero, what is the smallest prior standard error that you could comfortably allow? A standard error about the same size as the sample standard error of 4 seems right to me. What this means is that there is at least a fifty percent chance that the coefficient exceeds 4, and consequently that it takes less than a .25 increase in the conditional conviction probability to reduce the murder rate by 1/100,000. Next find an upper bound for the prior variance: If the prior location is zero, what is the largest prior standard error that you could comfortably allow? A number about five times the sample standard error seems about right to me. What this means is that there is at least a fifty percent chance that the coefficient is less than 20, and that it takes more than a 1/20 = .05 increase in the conditional conviction probability to reduce the murder rate by 1/100,000. Both of these numbers can be found in Table 6.

The choice of these upper and lower standard errors can be done only with some discomfort. That is the primary reason for taking the other approach of identifying a family of prior distributions that are practically equivalent to a dogmatic distribution or to a diffuse distribution. For example, a couple of diagnostics taken from Table 7 are $\lambda^* = 0$ and $\lambda_* = 1.48$ which apply to the first prior distribution in which the probability of conviction is a doubtful variable. These diagnostics form the questions: "Is your prior standard error less than zero times the sample standard error?" "Is your prior standard error more than 1.05 times the sample standard error?" The first question cannot have an affirmative answer and there can be no prior distribution that is equivalent to the dogmatic prior that merely sets this parameter to zero. The second question could have an affirmative answer, in which case the prior can be taken to be diffuse.

For purposes of discussion, the issue is assumed to be the effect of the conditional execution probability on the murder rate. The Bayesian elicitation diagnostics applicable to this issue for each of the five prior distributions are reported in Table 7. The first column indicates if the issue is an "included" or a "doubtful" variable. After that are two sets of four columns each, the first refering to diagnostics with the tolerance defined in terms of the percentage change in the estimate and the second set of four columns referring to diagnostics with the tolerance relative to the standard error. The first two columns in each group of four indicate the values of the statistics s^* and s_* which can be used to form diagnostics at any tolerance levels. The next two columns contain the elicitation diagnostics $\sqrt{\lambda^*}$ and $\sqrt{\lambda_*}$ corresponding to a ten per cent tolerance level. (The squareroot of λ is reported here in order to facilitate comparisons with standard errors rather than variances.) The elicitation questions posed by these two sets of diagnostics are: "Are your prior standard errors less than λ^* times the sample standard errors?" "Are your prior standard errors greater than λ_* times the sample standard errors?"

Consider first the case in which the tolerance is defined in terms of the percentage change in the estimate. The five values of λ^* are all very small; two values in fact are zero. This means that the prior distribution would have to be very concentrated in order to proceed as if it were dogmatic. The two zeroes occur because the constrained estimate is zero, and any coefficient different from zero is infinitely different in a per centage sense. Thus there can be no prior distribution that is practically equivalent to the dogmatic prior. The non-zero values for λ^* occur when the probability of execution is not a doubtful variable. Consider the right-wing prior which takes everything but these punishment variables as doubtful. The value of λ^* is .71. The elicitation question posed by this diagnostic is "Are your prior standard errors for the coefficients of the doubtful variables all less than .71 times the sample standard errors. The numbers in Table 6 can help to answer this question. Take a look at the column headed by the multiplier equal to .5 (which is close to .71). Referring to all but the punishment variables, is there a greater than fifty per cent chance that it would take a larger change than the value indicated to affect the murder rate by 1/100,000. Is it probable (50 per cent or more) that to change the murder rate by at least 1/100,000 would require more than: a .09 change in the per cent poor, a .061 change in the per cent unemployed, a .087 change in the fraction employed, a .37 change in the per cent nonwhite, a .026 change in the per cent youth, a .33 change in the per cent urban, a .028 change in the per cent male, a .032 change in the per cent of complete families and a 1.23 change in the Southern dummy(which is impossible). If you can answer affirmatively to all these questions and others like them, then you can act as if the prior distribution were dogmatic and all but the punishment variables could be omitted. I cannot, and I would not think that this dogmatic prior is a good approximation.

The other diagnostics, λ_* , identify a family of prior distributions that are equivalent to the diffuse prior. In the case of the crime of passion prior, which treats only the punishment variables as doubtful, the value of λ_* is only .57. Referring again to the numbers in Table 6 corresponding to a multiplier of .5, this diagnostic is asking the questions: "Are you pretty confident (at least fifty per cent) that in order to change the murder rate by 1/100,000 it would take less than a .5 increase in the conviction probability, less than a .22 increase in the execution probability and less than a 222 increase in the number of months of incarceration." My answer is yes, and if I were to use the bleeding heart prior, the approximaton that my prior is diffuse seems adequate.

The diagnostics applicable when the tolerance is defined relative to the standard errors do not seem to suggest that either the diffuse or the dogmatic prior is adequate.

Thus except in one case the message of these diagnostics is that this data set is neither so small that we can rely entirely on the prior information no so large that we can rely entirely on the data information. In retrospect, that seems like a predictable conclusion. It is hard to imagine that observation of 44 states could yield an estimates of the effect of executions that are impervious to the state of prior information, but it is equally hard to imagine that one could discard the data altogether.

This Bayesian approach to diagnostics can be contrasted to the traditional pretest diagnostics that help to select a method of estimation. The traditional diagnostic statistics for each of the five prior distributions are reported in Table 8. The first column indicates if the issue is an "included" or a "doubtful" variable. The second column contains the chi-square statistic for testing if the omitted variables have no effect. The third column reports the unconstrained estimate of the issue, and the fourth column the corresponding t-value. The last two columns contain the estimate of the issue with the doubtful variables omitted and the corresponding t-value.

Using the traditional levels of significance, the chi-squared statistics all would suggest that the restricted estimators are undesirable, the one possible exception being the crime-of-passion restrictions. The t-statistics for testing the Hausman hypothesis that the restricted estimator is consistent are small in two cases, which might suggest opting for the restricted estimator, nevermind the large chi-square statistic.

It is difficult for me to guess exactly how one might use the pretest diagnostics reported in Table 8 to select among the alternative estimates. It is even more difficult for me to suggest how they sensible could be used, since that would depend either on sampling properties or implicit prior distributions, neither of which do I have any inkling of. Thus my message: Bayesian elicitation diagnostics have a firm philosophical foundation, though they are often difficult to use in real settings. Pretest diagnostics, which seem easy to use, have a shaky philosophical foundation, and consequently yield a product whose value is very much in doubt.

/DIAGNOSTICS/REFERENC.DOC

REFERENCES

Berger, James (1984), "The Robust Bayesian Viewpoint," in Robustness of Bayesian Analysis, ed. J. Kadane, Amsterdam: North-Holland Publishing Company, 63-124.

Berger, James O. and O'Hagan, Anthony (1987), "Ranges of Posterior Probabilities for Unimodal Priors with Specified Quantiles," Invited Paper, Third Valencia International Meeting on Bayesian Statistics, Altea(Spain), 1-5 June.

Berkson, J. (1938), "Some Difficulties of Interpretation Encountered in the Application of the Chi-Squared Test," Journal of the American Statistical Association, 33, 526-542.

Hill, Bruce M. (1980), Review of Specification Searches by Edward E. Leamer, Journal of the American Statistical Association, 75, 252-253.

Hill, Bruce M. (1985), "Some Subjective Bayesian Considerations in the Selection of Models (with discussion)," Econometric Reviews, 4, 191-288.

Hill, Bruce M. (1988), "A Theory of Bayesian Data Analysis," University of Michigan, February.

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

Leamer, Edward E. (1975), "A Result on the Sign of Restricted Least Squares Estimates," Journal of Econometrics, 3 , 387-390.

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

Leamer, Edward E. (1982), "Sets of Posterior Means with Bounded Variance Priors," Econometrica, 50, 725-736.

Leamer, Edward E. (1983), "Let's Take the Con Out of Econometrics," American Economic Review, 73, 31-43.

Leamer, Edward E. (1985), "Sensitivity Analyses Would Help," American Economic Review, 75, 308-313.

Leamer, Edward E. (1987), "Econometric Metaphors," in Advances in Econometrics, ed. Truman F. Bewley, Cambridge: Cambridge University Press, 1-29.

Leamer, Edward E. (1989), "Planning, Criticism and Revision," Journal of Applied Econometrics, forthcoming.

Leamer, Edward E. and Chamberlain, G. (1976), "Matrix Weighted Averages and Posterior Bounds," Journal of the Royal Statistical Society, Series B, 38, 73-84.

McAleer, Michael, Pagan, Adrian R. and Volker, Paul A. (1985), "What Will Take the Con Out of Econometrics?," American Economic Review, 75, 293-307.

Pagan, Adrian (1987), "Twenty Years After: Econometrics, 1966 - 1986," University of Rochester, Working Paper No. 94.

Smith, A. F. M. (1986), "Some Bayesian Thoughts on Modelling and Model Choice"," The Statistician, 35, 97-102.

/DIAGNOSTICS/TABLES.DOC

Table 1

The Statistical Model and Summary Statistics

Model:

$$y \sim N(X\beta + Z\gamma, I)$$

$$\gamma \sim N(0, V)$$

Parameter Combination of Interest: $\psi'\beta + \eta'\gamma$, ψ and η given

Notation:

$$b = (X'M_zX)^{-1} X'M_zy$$

$$H_\gamma = \ = Z'M_xZ,$$

$$M_z = I - Z(Z'Z)^{-1}Z'$$

$$M_x = I - X(X'X)^{-1}X'.$$

$$P = (X'X)^{-1}X'Z.$$

$$\hat{\beta}(V) = (X'X)^{-1}X'(y - Z\hat{\gamma}(V)) = \hat{\beta}(0) - P\,\hat{\gamma}(V).$$

$$\hat{\gamma}(V) = (Z'M_xZ + V^{-1})^{-1} Z'M_xy.$$

$$\phi = (\eta - P'\psi)$$

$$m = 1 - \psi'(X'X)^{-1}\psi/\psi'(X'M_zX)^{-1}\psi$$

Test Statistics:

χ^2_γ tests $\gamma = 0$

t_H tests $-\psi'P\gamma = 0$

t_ϕ tests $(\eta' - \psi'P)\gamma = 0$

t_ψ tests $\psi'\beta = 0$

t_η tests $\eta'\gamma = 0$

Error Tolerance level: τ

Table 2

Diagnostics Indicating Priors Equivalent to the Dogmatic Prior

I. Constrained least squares with doubtful variables omitted ($V = 0$) is an adequate approximation if the prior variance satisfies

$$V \leq \lambda^{*} H_{\gamma}^{-1}$$

$$\text{where} \quad \lambda^{*} = (\tau/s) \; / \; (1 - (\tau/s)),$$

A. Tolerance τ Defined in Terms of Per Centage Change in Coefficients

1. Linear Combination of Interest Depends on Doubtful Coefficients ($\eta \neq 0$)

$$s_1 = (|t_{\phi}| + [\chi^2_{\gamma}]^{1/2}) \; [\phi' H_{\gamma}^{-1} \phi]^{1/2} \; / \; 2|\psi' \hat{\beta}(0)|$$

2. Linear Combination of Interest Depends on Diffuse Coefficients Only ($\eta = 0$)

$$s_2 = m^{1/2}(|t_H| + [\chi^2_{\gamma}]^{1/2}) \; |\psi' b / \psi' \hat{\beta}(0)| \; / \; 2|t_{\psi}|$$

B. Tolerance τ Defined in Terms of the Difference in Estimates Divided by the Standard Error

1. Linear Combination of Interest Depends on Doubtful Coefficients ($\eta \neq 0$)

$$s_3 = (|t_{\phi}| + [\chi^2_{\gamma}]^{1/2}|) \; [\phi' H_{\gamma}^{-1} \phi]^{1/2} \; / \; 2(\psi' (X' M_z X)^{-1} \psi)^{1/2}$$

2. Linear Combination of Interest Depends on Diffuse Coefficients Only ($\eta = 0$)

$$s_4 = m^{1/2} \; (|t_H| + [\chi^2_{\gamma}]^{1/2}) \; / \; 2$$

Table 3

Diagnostics Indicating Priors Equivalent to the Diffuse Prior

II. Unconstrained least squares with doubtful variables included ($V = \infty$) is an adequate approximation if the prior variance satisfies

$$\lambda_* H_\gamma^{-1} \le V$$

$$\text{where} \quad \lambda_* = s/\tau - 1.$$

A. Tolerance τ Defined in Terms of Per Centage Change in Coefficients

1. Linear Combination of Interest Depends on Doubtful Coeff.($\eta \neq 0$)

a. Partially Informative Prior

$$s_5 = (|t_\phi| + [\chi^2_\gamma]^{1/2}|)\ [\phi' H_\gamma^{-1}\phi]^{1/2} / 2|\psi' b + \eta' g|$$

b. Fully Informative Prior $X = 0$, $\psi = 0$

$$s_6 = [\ 1 + [\chi^2_\gamma]^{1/2}/|t_\eta|\ ] / 2$$

2. Linear Combination of Interest Depends on Diffuse Coefficients Only ($\eta = 0$)

a. Partially Informative Prior

$$s_7 = m^{1/2} (|t_H| + [\chi^2_\gamma]^{1/2}) / 2|t_\psi|$$

B. Tolerance Defined in Terms of the Difference in Estimates Divided by the Standard Error

1. Linear Combination of Interest Depends on Doubtful Coefficients ($\eta \neq 0$)

a. Issue depends on diffuse parameters, $\psi \neq 0$

$$s_8 = (|t_\phi| + [\chi^2_\gamma]^{1/2}|)\ [\phi' H_\gamma^{-1}\phi]^{1/2} / 2(\psi'(X'M_z X)^{-1}\psi)^{1/2}$$

b. Issue does not depend on diffuse parameters, $\psi = 0$

$$s_9 = (|t_\eta| + [\chi^2_\gamma]^{1/2}|)/2 \quad ; \quad \lambda_* = s/\tau$$

b. Fully Informative Prior $X = 0$, $\psi = 0$, $V = \lambda(Z'Z)^{-1}$

$$\lambda_* = -1/2 + (1 + 4t_\eta^2/\tau^2)^{1/2} / 2$$

2. Linear Combination of Interest Depends on Diffuse Coefficients Only ($\eta = 0$, $\psi \neq 0$)

a. Partially Informative Prior

$$s_{10} = m^{1/2} (|t_H| + [\chi^2_\gamma]^{1/2}) / 2$$

/DIAGNOSTICS/TABLES2'.DOC

Table 4

Variables Used In the Analysis

a. Dependent Variable
   M = Murder rate per 100,000, FBI estimate.
b. Independent Deterrent Variables
   PC = (Conditional) Probability of conviction for murder given commission. Defined by PC = C/Q, where C = convictions for murder, Q = M × NS, NS = state population.
   PX = (Conditional) Probability of execution given conviction (average number of executions 1946-50 divided by C).
   T = Median time served in months for murder by prisoners released in 1951.
   XPOS = A dummy equal to 1 if PX > 0.
c. Independent Economic Variables
   W = Median income of families in 1949.
   X = Percent of families in 1949 with less than one-half W.
   U = Unemployment rate.
   LF = Labor force participation rate.
d. Independent Social and Environmental Variables
   NW = Percent nonwhite.
   AGE = Percent 15-24 years old.
   URB = Percent urban.
   MALE = Percent male.
   FAMHO = Percent of families that are husband and wife both present families.
   SOUTH = A dummy equal to 1 for southern states (Alabama, Arkansas, Delaware, Florida, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia).
e. Weighting Variable
   SQRTNF = Square root of the population of the FBI-reporting region. Note that weighting is done by multiplying variables by SQRTNF.
f. Level of Observation

Observations are for 44 states, 35 executing and 9 nonexecuting. The executing states are: Alabama, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Illinois, Indiana, Kansas, Kentucky, Louisiana, Maryland, Massachusetts, Mississippi, Missouri, Nebraska, Nevada, New Jersey, New Mexico, New York, North Carolina, Ohio, Oklahoma, Oregon, Pennsylvania, South Carolina, South Dakota, Tennessee, Texas, Virginia, Washington, West Virginia.

The nonexecuting states are: Idaho, Maine, Minnesota, Montana, New Hampshire, Rhode Island, Utah, Wisconsin, Wyoming.

Table 5
Doubtful Variables for Each of Five Priors

	PC	PX	T	W	X	U	LF	NW	AGE	URB	MALE	FAMHO	SOUTH
Right wing				D	D	D	D	D	D	D	D	D	D
Rational Maximizer								D	D	D	D	D	D
Eye-for-Eye			D	D	D	D	D	D	D	D	D	D	D
Bleeding Heart	D	D	D					D	D	D	D	D	D
Crime of Passion	D	D	D										

Table 6
Change in the Explanatory Variable
Needed to Alter Murder Rate by 1/100,000

Coefficient Assumed to Equal Standard Error Times Multiplier

		Standard Error Multiplier $\sqrt{\lambda}$							
	Stnd Err-	.25	.50	1.00	2.50	5.00	10.00	15.00-	units
PC	4	1.000	.500	.250	.100	.050	.025	.017	f
PX	8.99	.445	.222	.111	.044	.022	.011	.007	f
T	.009	444	222	111	44	22	11	7	mnths
W	.002	2000	1000	500	200	100	50	33	$
X	22.3	.179	.090	.045	.018	.009	.004	.003	f
U	32.7	.122	.061	.031	.012	.006	.003	.002	f
LF	22.9	.175	.087	.044	.017	.009	.004	.003	f
NW	5.4	.741	.370	.185	.074	.037	.019	.012	f
AGE	77.1	.052	.026	.013	.005	.003	.001	.001	f
URB	6	.667	.333	.167	.067	.033	.017	.011	f
MALE	71.4	.056	.028	.014	.006	.003	.001	.001	f
FAMHO	61.8	.065	.032	.016	.006	.003	.002	.001	f
SOUTH	1.62	2.469	1.235	.617	.247	.123	.062	.041	f

Note: f = fraction

Table 7
Diagnostics

λ values correspond to ten percent tolerance

$\lambda^* = (.1/s) / \min[(1 - (.1/s),0]$

$\lambda_* = \min[\ (s/.1) - 1\ ,\ 0]$

Issue: Effect of Probability of Execution

Tol. defined in terms of Difference in the coefficients

Prior	i/d	Rel. to estimate				Rel. to Standard Error			
		s^*	s_*	$\sqrt{\lambda^*}$	$\sqrt{\lambda_*}$	s^*	s_*	$\sqrt{\lambda^*}$	$\sqrt{\lambda_*}$
Rt. Wing	i	2.180	1.674	.71	3.97	2.300	2.30	.21	4.69
Rat. Max.	i	3.135	1.018	.18	3.03	1.399	1.40	.28	3.61
Eye-Eye	i	2.278	2.160	.21	4.54	2.969	2.97	.19	5.36
Bleed. Heart	d	∞	0.320	.00	1.48	∞	13.23	.00	11.50
Passion	d	∞	0.133	.00	.57	∞	5.49	.00	7.41

Table 8
Test Statistics

Probability of execution

Prior	Status	χ^2_γ	Est(∞)	t	Est(0)	t(0)	$\lvert t_H \rvert$
Rt. Wing	i	68.99	-12.22	-1.37	-9.66	-0.81	0.55
Rat. Max.	i	25.14	-12.22	-1.37	-3.97	-0.39	2.49
Eye-Eye	i	127.37	-12.22	-1.37	-11.59	-0.78	0.13
Bleed. Heart	d	40.37	-12.22	-1.37	-	-	-
Passion	d	6.30	-12.22	-1.37	-	-	-

Note: χ^2_γ tests the joint significance of the doubtful variables.
Est(∞) is the estimate of the issue with all variables included.
Est(0) is the estimate of the issue with doubtful variables excluded.
t is the t statistic of the issue when all variables are included.
t(0) is the t statistic of the issue when doubtful variables are excluded.
t_H test the "Hausman" hypothesis that the constrained estimate is unbiased.

Table 9

Estimates and t-values

	ALL	RWING	RATMX	EYE	BLEED	PASSION
Intercept	47.55	15.51	7.37	7.90	-2.23	29.55
	(0.622)	(7.07)	(0.51)	(4.13)	(-0.16)	(0.40)
PC	-8.78	-15.81	-9.79	-9.51	-	-
	(-2.21)	(-3.04)	(-2.15)	(-1.52)		
PX	-12.22	-9.66	-3.97	-11.59	-	-
	(-1.37)	(-0.81)	(-0.39)	(-0.78)		
T	-0.0089	-0.046	-0.025	-	-	-
	(-1.01)	(-4.86)	(-2.75)			
W	0.0006	-	-0.0073	-	-0.009	-0.0003
	(0.23)		(-4.37)		(-5.40)	(-0.11)
X	26.88	-	-9.94	-	0.823	40.44
	(1.21)		(-0.59)		(0.05)	(1.86)
U	-0.310	-	-22.25	-	-25.09	-27.42
	(-0.009)		(-0.72)		(-0.88)	(-0.99)
LF	14.21	-	40.82	-	48.60	13.20
	(0.62)		(1.53)		(1.70)	(0.55)
NW	10.21	-	-	-	-	12.16
	(1.90)					(2.23)
AGE	46.99	-	-	-	-	38.25
	(0.61)					(0.47)
URB	-0.93	-	-	-	-	3.5
	(-0.17)					(0.65)
MALE	-85.10	-	-	-	-	-66.35
	(-1.19)					(-0.93)
FAMHO	-22.35	-	-	-	-	-19.34
	(-0.36)					(-0.31)
SOUTH	4.93	-	-	-	-	5.15
	(3.03)					(3.45)