Reporting Econometric Results: Believe It or Not?

Rod F. Ziemer

Econometric results—indeed a term that for many economists conjures up horrifying visions of well-meaning but perhaps marginally skilled, likely nocturnal, individuals sorting through endless piles of computer printouts. One "final" print-out is then chosen for seemingly mysterious reasons and the rest discarded to be recycled through a local paper processor and another computer printer for other "econometricians" to repeat the process ad infinitum. Besides supplying a lucrative business for the paper recyclers, what useful output, if any, results from such a process? This question lies at the heart of the so-called "science" of econometrics as currently applied, a practice which has been called "data mining," "number crunching," "model sifting," "data grubbing," "fishing," "data massaging," and even "alchemy," among other less palatable terms. All of these euphemisms describe basically the same process: choosing an econometric model based on repeated experimentation with available sample data.

The present concern, if not frustration, with the data-dependent approach to econometric modelling is evidenced in recent papers by Hendry (1980), Hocking and Pendleton (1983), Leamer (1983), Lovell (1983), and Zeller (1982). In this paper, the problem is reviewed and possible solutions, or perhaps more appropriately approaches, are discussed. An operational compromise solution is offered along with a *speculation* concerning its possible acceptance as an alternative to the "provincial" solution presently in widespread use.

THE PROBLEM

The problem commonly faced by econometric practitioners can be described as follows. A dependent variable, say y, is thought to be a function of a group of independent variables, the x's, and a random disturbance, say e. To make the problem tractable, let's assume, as is commonly done, that our function is linear in the x's and additive in e such that in matrix notation: $y = X\beta + e$, where β is a conformable parameter vector to the matrix of independent variable observations X. Let's also make the common assumption that the vector e has all the usual "nice" properties, i.e., normally distributed with mean vector zero and scalar identity covariance matrix, $\sigma^2 I$, and that x is not ill-conditioned, i.e., our x's are not highly collinear.

The situation usually faced with such a model is that some of the x's or columns in X, which we might denote by a matrix X_1 are known with high probability to actually belong in our model, while the rest of the x's say X_2, are only suspected to belong in our model. For example, theory and convention almost unquestionably suggest that price and income should appear in a demand function while variables such as age, sex, and other socioeconomic consumer characteristics are less supported by theory and only inconsistently supported by previous studies and a researcher's personal prejudices and beliefs. Our *problem* is: Given X_1, how do we use the information among the sometimes large set of

Land Economics, Vol. 60, No. 1, February 1984
0023-7639/84/001-0122 $1.50/0

Assistant professor, Department of Agricultural Economics, Texas A&M University. The author would like to thank Ronald C. Griffin and Glenn S. Collins for helpful comments. Texas Agricultural Experiment Station Article 18754.

variables in X_2 to estimate the parameters in our model?

THE PROVINCIAL SOLUTION

The most popular and apparently widely accepted approach to the problem is to estimate a large number of alternative models with different subsets of the suspect variables in X_2, and then choose one as being appropriate based on some seemingly reasonable criteria.[1] The result is often called a "final" model in the sense that it represents some hopefully meaningful end to an often long and vigilant search process. This provincial model is then reported usually as if all the trial models had never been estimated. And, even if the reporter states that alternative models (i.e., hypotheses) were rejected for whatever reasons, the reader still has little idea concerning the nature of the rejected models' estimates.

What degree of credence can the reader place in the researcher's final or provincial model? First, from a statistical perspective, it is well known that usual estimator properties essentially evaporate when one repeatedly experiments or "pretests" the data. In particular, least squares is no longer unbiased and minimum variance, and usual hypothesis testing criteria such as standard t-tests are unreliable. In addition, the statistical properties of multiple or sequential pretest estimators are generally unknown. Thus, if one "peeks" at the data, a sacrifice is made regarding usual estimator properties.[2]

Besides statistical problems, which are well-documented, another and perhaps more pernicious problem exits with the provincial approach. This problem has to do with interpretation. Specifically, how is one to interpret the validity of reported econometric results and conclusions based on the researcher's subjectively selected final model? This question is often difficult if not impossible to meaningfully address since the criteria by which the provincial model is chosen are usually not discussed by the researcher. In some cases, authors candidly report such practices as maximizing, $\bar{R}^2$, deleting statistically insignificant variables, or choosing the model that yields results most consistent with a priori expectations which are based on theory or previous studies. However, even in these cases, only the provincial model is usually reported so that the reader has little if any notion of actual results produced by the rejected models. Hence, the provincial approach forces upon the reader the search criteria (and thus the model) employed by the researcher without the luxury of actually seeing, or perhaps even knowing, the process by which the reported results were generated. Indeed, even the researcher may be ignorant of intermediate search results that led to a final model, as in the case of "canned" computer stepwise regression procedures.

For many, the provincial approach is distasteful if not unacceptable, as evidenced by what presently seems to be rampant and justifiable criticism of econometric analysis and reporting, evoking humorous metaphors such as: "There are two things you are better off not watching in the making: sausages and econometric estimates" (Leamer 1983). A number of alternatives to the provincial approach have been suggested but appear to have been largely ignored by econometric practitioners, likely because of the complexity or difficulty involved in their understanding and/or application. In the remainder of this discussion, some of these alternatives are reviewed and an estimation approach representing a compromise between desirable properties and operational ease is suggested.

SHRINKAGE ESTIMATORS

An alternative to the provincial solution, which involves one or more repetitions of a process that chooses one model as being preferred to another, is offered by what is sometimes called the family of shrinkage estimators. For the most part, these estimators shrink the ordinary least squares estimator to-

[1] A number of sequential ad hoc model choice criteria have been proposed; recent reviews can be found in textbooks by Chow 1983 and Judge et al. 1980.

[2] See Wallace 1977 for further discussion of the consequences of pretest estimation.

ward some point in the parameter space chosen by the investigator, often the null vector. For example, suppose we have reason to doubt that any of the suspect variables in X_2 belong in our model but are not quite sure. We might estimate β by least squares based on all the x's, yielding $\hat{\beta}=(X'X)^{-1}X'y$, and then re-estimate β by least squares based on only the variables in X_1 (i.e., omitting all suspect or doubtful variables), yielding $\beta^*=(X_1'X_1)^{-1}X_1'y$. The estimator β^* can also be obtained by estimating β based on all the x's subject to the restrictions that the coefficients associated with the suspect variables in X_2 are equal to zero.

The provincial approach suggests we choose either $\hat{\beta}$ or β^* usually based on a statistical test, resulting in a traditional pretest estimator. Alternatively, the shrinkage approach suggests we shrink the least squares estimator $\hat{\beta}$ toward zero, or some other prior point in the parameter space such as β^*, based on a data-dependent weighting formula. If we choose our prior to be zero, a general class of shrinkage estimators is defined as: $\tilde{\beta}=W\hat{\beta}$, where W is a weighting matrix or scalar which is a function of the data and the *loss function* chosen by the investigator.[3] Shrinkage estimators in this class include the Stein-Rule (see Judge and Bock 1978) and the Generalized-Ridge Rule (see Strawderman 1978). If we choose our prior to be β^*, a generalized Stein-rule may be defined as: $\hat{\beta}_*=W(\hat{\beta}-\beta^*)+\beta^*$ where $0 \leq W \leq 1$. In contrast to the traditional pretest estimator, which is equal to either $\hat{\beta}$ or β^*, $\hat{\beta}_*$ is a convex combination of $\hat{\beta}$ and β^*. Thus, loosely speaking, our final estimates can be thought of as some "weighted average" of the least squares estimator $\hat{\beta}$ and our prior notions about β inherent in β^*.

In general, estimators in the shrinkage family provide a means of combining information in the sample data with our prior notions concerning model parameter values (and thus, included or excluded variables) without having to exclusively choose between seemingly conflicting estimates. Under appropriate conditions, these estimators, in contrast to conventional pretest estimators, are unambiguously superior to ordinary least squares under commonly adopted loss functions.[4] At first glance, one might wonder why these and other related estimators, which have been available for some time, have only rarely been applied. The reasons are likely many but a lack of understanding and operational ease are certainly worthy of mention. Many students of economics have not been exposed to shrinkage estimators which are usually only a topic covered in advanced econometrics courses. Perhaps more important however, is that no existing popular computer packages provide shrinkage estimates, so it is not easy to obtain such estimates relative to, say, least squares.

Other reasons relate to the properties of the shrinkage estimators themselves. First, no conventional tests of hypotheses (such as standard t-tests) are possible; indeed, these estimators represent a complete alternative to pretesting (that is, choosing a model based on a hypothesis test). Second, many of the shrinkage estimators are highly sensitive to certain design-related conditions, such as multicollinearity, and often appropriate conditions for their use are not met in typical economic data (see Hill and Ziemer 1982). In sum, the shrinkage approach, although promising, is at present less than a generally operational tool for the data-minded econometrician.

THE BAYESIAN APPROACH

Thus far, we have only considered "classical" solutions to our econometric modelling problem. The Bayesian paradigm offers a complete alternative to classical econometrics. One important difference between the traditional classical and Bayesian approaches is that, to a Bayesian, β and other parameters that appear in our model are not fixed un-

---

[3]Expected loss or risk is defined by the function:

$$R=E[(b-\beta)'Q(b-\beta)],$$

where b is an estimate of β and Q is a symmetric positive definite matrix. If $Q=I$, R is expected squared error loss. Other popular choices of Q are $(X'X)$ and $(X'X)^2$.

[4]See Judge and Bock 1978 for an extensive discussion of the properties of shrinkage as well as pretest estimators.

known constants but rather *random* variables. Therefore, Bayesian analysis naturally admits prior notions about β through specification of a probability distribution. Just as the classical econometrician specifies a distribution for the random variable e, usually normal, the Bayesian specifies additional *prior* distributions for β and other relevant model parameters. Then, giving the priors for e, β, and other random variables in the model along with the sample data, Bayes Rule is applied to obtain a *posterior* distribution for β. It is this posterior distribution that is used to make inferences about β.

Let the probability density function $g(\beta)$ summarize the researcher's information or hypotheses about β. Let $l(\beta|y)$ be the sample likelihood function which summarizes the information in the sample data. Then by Bayes Rule we can define a posterior density function for β as: $f(\beta|y) \propto l(\beta|y) g(B)$ where $\propto$ denotes proportionality. The function $f(\beta|y)$ describes our notions about β after we have observed the sample data. Thus, we begin with prior knowledge about β expressed by the function $g(\beta)$ and then combine this knowledge with the sample information expressed through $l(\beta|y)$. The result is the function $f(\beta|y)$ which represents our revised feelings about β after observing the sample data. As an example, consider the normal linear regression model and suppose we choose the noninformative or diffuse prior $g(\beta) \alpha k$, where k is a constant (which implies complete ignorance about β, i.e.,$-\infty < \beta < \infty$). Then $f(\beta|y)$ has the form of a multivariate normal density function, a result consistent with the classical result that the least squares estimator $\hat{\beta}$ has a normal distribution. Specifically, $f(\beta|y)$ implies $\beta \sim N[\hat{\beta}, \sigma^2(X'X)^{-1}]$. Thus, from a Bayesian perspective, the classical estimator $\hat{\beta}$ has a clear interpretation in the normal linear regression model with a diffuse prior; it is the posterior mean of β.[5]

To approach our suspect variable problem, the Bayesian might formulate the prior $g(\beta)$ as normal with all coefficients associated with the doubtful variables in X_2 having mean equal to zero. Variances of these coefficients could be chosen to be large for associated variables that were highly doubtful or suspect and relatively smaller for other less suspect variables (for example, see Leamer 1982). Of course, many other approaches are also possible. Briefly, Bayesians first summarize their prior notions about β through a probability density function (the prior) and then allow the sample data to alter this function resulting in another probability density function (the posterior) which is used to make inferences about β.

Upon first exposure, many classically trained individuals are critical of the Bayesian approach, asserting that the choice of priors is purely subjective. This assertion is for the most part true; however, as most Bayesians would quickly point out, classical econometrics as currently practiced is no less subjective. Indeed, the classical provincial solution is the result of prior notions concerning the "appropriate" signs of model parameters, the magnitudes of estimated coefficients and standard errors, etc. Thus, the primary difference between the classical and Bayesian approaches is the manner in which prior information is used, rather than whether or not priors are deemed relevant.

The Bayesian approach has yet to "catch on", probably because posterior distributions are usually cumbersome to derive and interpret. It seems likely that few individuals would have trouble accepting a Bayesian approach conceptually. Rather, it is operational ease that appears to be the culprit responsible for the lack of acceptance of Bayesian methods as well as for the success of the classical provincial approach.[6]

A COMPROMISE SOLUTION AND SPECULATION

A thorough understanding of the classical

[5]It is important to keep in mind the interesting distinction that to a Bayesian, β is a normally distributed random variable which has mean $\hat{\beta}$ given a diffuse prior for β, while from a classical perspective, $\hat{\beta} \sim N[\beta, \sigma^2(X'X)^{-1}]$ so that β is a fixed unknown constant which is the mean of the random variable $\hat{\beta}$, the least squares estimator.

[6]The reader is referred to Zellner 1971 for a thorough introduction to Bayesian econometrics.

shrinkage or Bayesian approaches would likely be beyond the interest of the average econometric practitioner. It is unrealistic to expect many classically trained individuals (the majority of the user's of econometric procedures) to learn and use Bayesian methods or even complex shrinkage estimation procedures. In addition, few individuals have access to computer programs for applying such techniques. Thus, even for the "informed" individual desiring alternatives to the provincial approach, operational constraints may limit their consideration. Yet, there appears to be a growing justifiable reluctance, if not a refusal, to accept econometric results that are the result of the provincial approach.

The root of concern with the provincial solution seems to lie in the latency of the researcher's subjective priors, that is, his or her adopted model choice criteria. In particular, the reader or client of provincial econometric results often has little if any idea of how the so-called final model was produced and is forced into a polar "believe it or not" situation, i.e., believe the researcher's prior; and thus the model, or not.[7] A reasonable, and more importantly, an operational compromise solution to this apparent dilemma would be for the researcher to simply reveal model estimates obtained under alternative priors. As standard practice, if the number of suspect variables is small, one should present results for *all* possible models obtained by including and excluding the suspect variables. Why? First, "suspect" variables must be suspect because the researcher's priors are confounded with what he or she perceives the priors of others to be. Thus, presenting results for all possible suspect models allows the reader to choose a prior or priors consistent with his or her own preferences. Second, such an approach allows the reader, as well as the researcher, to determine how *robust*, or stable, parameters of interest are to a variety of different subjective priors.

If the number of suspect variables is large, the desire for parsimony may preclude presenting all possible suspect models. In such cases, one should at least present upper and lower bounds obtained during the search process for the parameters of interest, that is, parameters associated with variables well supported by theory and other rigorous foundation or parameters which will be used in some postmodel analysis.[8] If upper and lower bounds on parameters of interest are tight for alternative priors, one could more confidently suggest policy implications based on a post-model analysis or make inferences regarding behavioral hypotheses.

As an illustration of why parameter estimates from suspect models are valuable, consider the following example. Suppose a parameter estimate appeared negative in one suspect model, zero in another, and positive in yet another. Such a result suggests that the data do not consistently support a hypothesis concerning the sign of such a parameter or the inclusion or exclusion of its associated variable. In addition, one would likely be leery of simulation, control, or other posteconometric analyses based on an estimate chosen from such conflicting candidates. Finally, policy implications arising from such analyses should certainly be appropriately tempered. In sum, the reader should not be denied the opportunity of knowing such information if he or she is to intelligently judge the validity of hypothesis tests or a postmodel analysis.[9]

---

[7]This problem can be compounded by journal reviewers and editors who, in the interest of parsimony, request authors to delete discussions of model selection and results obtained under alternative priors. Thus, reviewers should be wary of denying the eventual reader information concerning the researcher's own subjective prior(s) or imposing upon the reader their own "edited" prior(s).

[8]This approach is applied by Leamer 1983 and has a formal Bayesian interpretation in that the resulting bounds lie on a function known as the "curve decolletage" (Dickey 1975). In nontechnical terms, this function represents a locus or intersection of the researcher's prior about β and the sample data. Computationally, the curve decolletage is defined by the 2^k linear regressions resulting from all possible combinations of the k independent variables in the model (Chamberlain and Leamer 1976).

[9]Given a wide range for a parameter of interest for different suspect models, it would seem desirable to perform any postmodel analysis using the upper and lower estimate bounds. Such a "sensitivity" analysis could be reported in addition to results and implications based on a researcher's chosen (i.e., personally preferred) parameter estimate.

Of course, there is no substitute for good theory and well-specified models, and one is likely much better off using theory and careful contemplation to reduce the contents of a basketful of suspect variables than a statistical search procedure. Even so, given the nature of economics, we will always be faced with suspect variables in our econometric models, although good theory should continue to reduce their numbers. In the interim we need computationally tractable alternatives to the generally unacceptable provincial approach. It is my *speculation* that the compromise solution of simply more complete reporting of numerical particulars from an adopted search procedure, other than just the final model, will generate increased credibility and acceptance of reported econometric results. At least we will all be better able to judge what to believe, or perhaps more fitting, what *not* to believe.

References

Chamberlain, Gary, and Leamer, Edward E. 1976. "Matrix Weighted Averages and Posterior Bounds." *Journal of the Royal Statistical Society B* 38(1): 73–84.

Chow, Gregory C. 1983. *Econometrics*. New York: McGraw-Hill.

Dickey, J.M. 1975. "Bayesian Alternatives to the F-test and Least-Squares Estimates in the Normal Linear Models." In *Bayesians Studies in Econometrics and Statistics*, ed. S.E. Fienberg and A. Zellner. Amsterdam: North-Holland.

Hill, R. Carter, and Ziemer, Rod F. 1982. "The Application of Generalized Ridge and Stein-Like General Minimax Rules to Multicollinear Data." *Communications in Statistics B* 11(6): 623–38.

Hendry, David F. 1980. "Econometrics—Alchemy or Science?" *Economica* 47(Nov.): 387–406.

Hocking, R.R., and Pendleton, O.J. 1983. "The Regression Dilemma." *Communications in Statistics B* 12(5):497–527.

Judge, George G.; and Bock, M.E. 1978. *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*, Amsterdam: North-Holland.

———; Griffiths, W.E.; Hill, R.C.; and Lee, T.C. 1980. *The Theory and Practice of Econometrics*, New York: Wiley.

Leamer, Edward E. 1982. "Sets of Posterior Means With Bounded Variance Priors." *Econometrica* 50 (May): 725–36.

Leamer, Edward E. 1983. "Let's Take the Con out of Econometrics." *American Economic Review* 73 (Mar.):31–43.

Lovell, Michael C. 1983. "Data Mining." *Review of Economics and Statistics* 65(Feb.):1–12.

Strawderman, William E. 1978. "Minimax Adaptive Generalized Ridge Regression Estimators." *Journal of the American Statistical Association* 73 (Sept.): 623–27.

Wallace, T. Dudley. 1977. "Pretest Estimation in Regression: A Survey." *Amerian Journal of Agricultural Economics* 59 (Aug.): 431–43.

Zellner, Arnold 1971. *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley.

———. 1982. "Basic Issues in Econometrics: Past and Present." *American Economist* 26(1): 5–10.