FACULTY PAPER SERIES

FP 01-10 October 2001

Two Approaches to the Model Specification Problem in Econometrics

George C. Davis, Editor
gdavis@tamu.edu

Department of Agricultural Economics
Texas A&M University
College Station, Texas 77843-2124

The Model Specification Problem from a Probabilistic Reduction Perspective: Discussion

George C. Davis

Introduction

Anyone serious about the model specification problem knows that Aris Spanos and Anya McGuirk (SM) are on the short list of influential econometricians writing on the subject. In this latest installment, they have the difficult task of summarizing over fifteen years of writing on the probabilistic reduction (PR) approach to the problem. Just as it is difficult to summarize the PR approach in a short paper, it is equally difficult to fully discuss it in a few pages.

Given the title of the paper, my preconception was that the paper would focus on providing lucid discussions on two questions: i) what are the identifiable steps in implementing the PR approach? (ii) what are the advantages and disadvantages of the PR approach? SM answer these questions within the historical development of the PR approach. Here I will focus on how well the PR approach answers these questions. In making this assessment it seems unfair to me to judge the approach solely on the limited space SM were given, so I will draw freely from some other cited references as well.

Philosophical Foundations and Implementation Steps in the PR approach

For me, the PR approach clearly rests on a tripod of philosophical beliefs about econometric modeling. Leg one: The *theory-data gap* is real and explicitly recognized. As startling as it may seem, economic theories *do not* purport to explain or describe *observed data*, rather they describe *theoretical data* (See Haavelmo for a good discussion). Because of this gap, the PR approach is one of the few, if not the only econometric methodology, that explicitly distinguishes between three types of processes or systems: (i) the actual *data generating process* (DGP); (ii) the *theoretical model* that attempts to approximate the DGP by concentrating on some subset of variables; (iii) the *statistical model* that is viewed as a consistent set of probabilistic assumptions relating to the observable random variables underlying the data chosen (Spanos 1995, p. 209). Leg two: *specification begins with a joint distribution* for the chosen observable variables and the conditional distribution is derived from the joint distribution. This is in contrast to the 'traditional econometric' approach, which bridges the theory-data gap by assuming a disturbance term with a certain distribution and then deriving the conditional distribution by the method of transformations. Leg three: *statistical distribution assumptions must be satisfied* before any 'within model' hypothesis testing begins. That is, obtaining an 'adequate' statistical model, which summarizes the 'systematic' probabilistic aspects of the data, is the goal.

The key implementation steps in the PR approach are presented schematically in figure 1 and here I want to focus on the three most important, which are defined very concisely elsewhere:

(i) *Specification*: choosing the appropriate statistical model, in view of the information available at the outset.
(ii) *Misspecification*: testing the assumptions underlying the statistical model.
(iii) *Respecification*: choosing an alternative statistical model when the original model is found to be misspecified [do this until a *statistically adequate* model is found](Spanos 1995 p. 210).

In the *specification* stage, 'the information available at the outset' comes mainly from economic theory, which suggest a vector of variables $\boldsymbol{Z_t}$. The researcher begins with some joint distribution $D(\boldsymbol{Z}_1, \boldsymbol{Z}_2,\ldots, \boldsymbol{Z}_T; \phi)$, with ϕ being a parameter vector. The reduction then seeks to reduce this distribution without loosing relevant systematic statistical information. The relevant regularities can be captured by the three broad categories stated: (D) *Distribution*; (M) *Dependence*; and (H) *Heterogeneity*, so initially some assumption must be made about these categories. Following Spanos (1995), suppose we assume the joint distribution is NIID with $\mathbf{Z}_t = (y_t, \mathbf{X}'_t)'$, then

$$(1) \qquad \begin{pmatrix} y_t \\ \boldsymbol{X}_t \end{pmatrix} \sim N\left[\begin{pmatrix} m_y \\ \boldsymbol{m}_x \end{pmatrix}\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \Sigma_{22} \end{pmatrix}\right]$$

and the conditional mean and variance take the form $E(y_t \mid \mathbf{X}_t = \mathbf{x}_t) = \beta_0 + \beta' \boldsymbol{x}_t$ (linearity) and $\mathrm{Var}(y_t \mid \mathbf{X}_t = \mathbf{x}_t) = \sigma^2 = \sigma_{11} - \sigma_{12}\Sigma_{22}^{-1}\sigma_{21}$ (homoskedasticity) and parameter invariance, where $\beta_0 = m_y - \sigma_{12}\Sigma_{22}^{-1}m_x$ and $\beta = \Sigma_{22}^{-1}\sigma_{21}$. "The model is specified exclusively in terms of probabilistic assumptions, with the parameters having a purely statistical interpretation; no theoretical interpretation is given at this point."(Spanos, 1995 p.212).

Having specified (1), attention turns toward *misspecification* testing to determine if the NIID assumptions are satisfied. As Spanos and McGuirk indicate in the empirical modeling section, this can be done "using graphical techniques" such as t-plots and scatter plots and "supplemented with a battery of judiciously chosen misspecification tests," such as tests for normality, linearity, homoskedasticity, independence, and parameter constancy. These tests are viewed as being based on a Fisher notion of 'pure significance testing.' If the underlying probabilistic assumptions are not satisfied then one returns to the "joint distribution of the observable variables involved $D(\boldsymbol{Z}_1, \boldsymbol{Z}_2,\ldots, \boldsymbol{Z}_T; \psi)$ and imposing a different set of reduction assumptions in order to take account of any systematic information excluded by the original assumptions" (Spanos 1995, p. 214). This process of *respecification* and *misspecification* testing continues until an 'adequate statistical model' is discovered. Adequate means that the underlying probability assumptions are satisfied and all the systematic information is accounted for in the model. Once a statistically adequate model is found, then one can test the theory 'within' that model.

Advantages, Disadvantages, and Lingering Questions of the PR Approach

For me, the major advantages of the PR approach are two of its philosophical beliefs about econometric modeling. First, by emphasizing the theory-data gap, the PR approach keeps distinct the data generating process, the theoretical model and variables, and the statistical model and observable variables. This is much in line with Haavelmo's original conception of econometric modeling and, if understood and kept in mind, can save the modeler from many false and misleading inferences (see Davis 2000 for more discussion). Second, if one follows the PR approach, then the underlying statistical assumptions will have been validated and faulty inferences should not be due to inappropriate statistical assumptions.

The major disadvantage I see in the PR approach is its lack of specificity in the implementation steps. Others I suspect will see this as an advantage but because lack of specificity is the model selection problem, I consequently do not believe it does much to ease the model specification problem. In the *specification* stage, 'the information available at the outset' comes mainly from economic theory or as Spanos states elsewhere (1995, p. 209), "The choice of the statistical model is influenced by the theory in so far as it is required to allow the modeler to consider the theoretical question of interest in its context." I believe this gives most economic theories proper more credit than they deserve. Note from equation (1) even if an economic theory proper were to give a complete specification of **X** (very unlikely), three other pieces of information are also needed: (i) the form of the distribution; (ii) the form of the mean conditional function (assuming normality is not implied); and (iii) the form of the variance. Many, if not most, economic theories proper say very little about these four components, so these then amount to auxiliary assumptions to which most economic theories proper are immune. Consequently, the change of emphasis from the error term to a joint distribution of *observables* I believe undermines the initial and important emphasis placed on the theory-data gap by the PR approach. It is understandable from a modeling standpoint why one would want to focus on distributions of observables, but most theories are not theories about observables, they are theories about unobserved counterfactuals. Ignoring this fact and focusing on the joint distribution of observables tends to provide a false sense of inferential security. Let me hasten to add however that I do not view this as an econometric problem but an economic theory problem. More on this later.

I applaud the *misspecification* step as it seems good scientific practice to try to test all underlying assumptions, but I am not convinced it is as easy as SM make it seem. I am very skeptical of looking at plots for modeling for two reasons: (i) As has been well established by psychologists, interpreting pictures is heavily theory laden and subject to numerous interpretations; (ii) and related to (i), plots can at best represent 3 dimensions simultaneously, and if the problem of the model is due to some higher dimensional causes, then mapping $k > 3$ dimensions into 2 or 3 dimensions can be misleading as well. It is for this reason I think more of the burden should placed on formal tests but then this raises other obvious questions: How many tests must be passed before the specification is deemed adequate? 50%? 75%? 95%? Why? Would not a joint test be more likely to reject the null? Of course, decision ease may be inversely proportional to the number of misspecification tests. I get the impression that most of these misspecification tests (secondary tests) are being conducted in order to validate small sample statistics used for other primary tests (economic hypotheses) but if the secondary tests are themselves asymptotic or based on other unsubstantiated assumptions why should we believe these tests over an asymptotic argument for the primary tests?

The respecification step is perhaps the most controversial. The most obvious criticism here is that of 'pre-test bias', which is not even mentioned here but has been addressed elsewhere: "[W]hen one distinguishes between the statistical and theoretical models, and the former is interpreted as an adequate summary of the 'probabilistic information' in the data, the 'pre-test' bias problems does not arise. This is because the probabilistic information in the data is invariant to how long one looks at the data or how many regressions one estimates." (Spanos 1995, p. 196). I am not sure how this answers the technical pre-test bias problem but my pre-test concern is somewhat more general. Analogous to the theory underdetermination problem discussed in the philosophy of science literature there is a statistical model underdetermination problem: multiple statistical models can explain a single data set. So where SM may see an ARCH or STAR type model, someone else may see omitted distributional effects (see e.g., Buse). For this reason I do not believe that 'the probabilistic information' is invariant to how long one looks at the data. In fact, in talking about misspecification testing Spanos (1998 p. 126) makes a similar point, "... there is usually *no unique way* (emphasis added) to specify the negation of the null hypothesis: non-normality, non-independence and t-variance can take *numerous forms* (emphasis added)." This then leads to the next question. How does one decide between equally adequate statistical models? This issue is not addressed, but it is a shortcoming of the PR approach.

I believe more fundamental than the pre-test problem is the relationship between the economic theory proper, initial statistical specification, the respecified model, and the final statistically adequate model. Based on the earlier discussion of the specification stage, the initial data $\boldsymbol{Z}_t$ seems to be fixed according to the economic theory. Yet in respecifying the statistical model in hopes of obtaining 'a statistically adequate model' often it seems the approach is to add variables in order to first help satisfy the statistical assumptions and second account for all the 'systematic information' in the data and second.

On the first count, strictly speaking if the data in the specification stage is truly fixed by the theory then should not one search for the distribution that fits the data rather than the data, the expanded data which means the data is not fixed, that fits the distribution? It seems clear from the applications of the PR approach I have seen that practitioners do not really take the data as fixed in the specification stage, though a strict interpretation of the PR approach would. Most of the applications of the PR approach I have seen have been in a time series context and violations of distributional assumptions are usually addressed by adding lagged variables. Now the argument for this appears to be in the PR approach that the statistical model is designed to account for all the 'systematic information' in the data. However, I know of no theory that claims to explain or capture all of the systematic information in some variables, rather there is always an explicit or explicit *ceteris paribus* phrase attached. The notion of requiring a statistical model to explain all systematic variation is I believe an auxiliary experimental design assumption (borrowing Haavelmo's terminology) that is not implied by the economic theory proper. Perhaps this notion of systematic variation is a way for accounting for the ceteris paribus clause by controlling for other items but this is not clear. If this is the purpose of adding the other variables, as is stated in other econometric methodologies, that is fine, but this then opens up the thorny question of how should the additional variables and their parameter estimates be interpreted? Are all variables then placed on equal footing with respect to their interpretation and inference? This would seem odd as strictly speaking the theory was silent on some and not silent on others. Some criterion or discussion of this distinction is needed and it has been made elsewhere. For example Pratt and Schlaifer have written on the distinction that is to be made in statistical models between factors and concomitants. They lay out conditions under which statistical laws can be established when concomitants are involved. To me the lack of a distinguishing factors from concomitants and the implications is a deficiency in the PR approach that could likely be easily overcome by just following the lead of Pratt and Schlaifer. A good example of this more progressive approach of not relying on concomitants but rather returning to a more complete theoretical structure is the work of McGuirk, et al.

It is also claimed that in the specification stage, "The model is specified exclusively in terms of probabilistic assumptions, with the parameters having a purely statistical interpretation; no theoretical interpretation is given at this point."(Spanos, 1995 p.212). While I think it is conceptually useful and possible to separate the theoretical model from the statistical model, I do not believe this is possible in practice. Spanos does appear to recognize the implication in one direction, "[E]ven though the probabilistic assumptions are not part of the theoretical model, one cannot separate the estimation and the testing of the theory from the validity of the probability assumptions underlying the statistical model adopted." (Spanos, 1995, p. 203). I believe this is a biconditional relationship, not just a conditional relationship. That is, and one cannot separate the validity of the probability assumptions underlying the statistical model from the theoretical model. How can the model be 'specified *exclusively* in terms of probabilistic assumptions' when it was the theory that suggested the initial specification?

At a more abstract and general level, the problem appears to be in the PR approach the use of monotonic logic where non-monotonic logic is required (see e.g., Nolt). Monotonic logic allows one to ignore conjunctive premises (by the law of simplification) that may change or be false in proving validity. In non-monotonic logic if a single conjunctive premise changes, then validity of the entire system may be compromised. In economic jargon, if the changing of a ceteris paribus condition does not change the conclusions of a closed deductive system (e.g., a theory) then monotonic logic can be applied. However, if the changing of a ceteris paribus condition does change the conclusions of a closed deductive system then monotonic logic is inappropriate and non-monotonic logic must be applied. It would seem to me that most of the time changing ceteris paribus conditions will affect our conclusions, and therefore the appropriate logic for economic and econometric analysis is not monotonic but non-monotonic logic.

It is also not clear exactly what the loss function is that applies in the PR approach. That is, there seems to be no recognition or concern about the tradeoff between model fit and parameter precision. The only literature cited with respect to the model selection problem is Fisher and Pearsan, but there have been numerous articles written on the model selection problem and goodness of fit measures since that time (e.g., Akaike, Schwarz). Is there nothing of use in this work?

One area that seems to require further analytical work in the PR approach is to determine if it is a model consistent procedure. In the model selection literature, a model selection procedure is deemed model consistent if as the sample size goes to infinity, the procedure will converge on the true model with probability one. Many may understandably view this as providing little comfort in a finite world but for me it is better to have this information than not have it. It also seems that the likelihood information criteria are attractive because they may be tied to the Kullback-Leibler entropy measure, which by definition is a measure of the distance between a true unknown density and an estimate of that density (see Akaike). Is not this what the PR approach is trying to minimize?

The empirical examples given by SM are very encouraging and demonstrate how useful the PR approach can be in *experienced* hands. These encouraging results are analogous to the encouraging results found by Hoover and Perez, who follow the closely related general-to-specific methodology of Hendry. However, this provides no evidence that the procedure is a reliable methodology. To be a reliable methodology it must be able to deliver the goods not only in the hands of those who advocate its use, but also in the hands of those who have no vested interest in the procedures. A methodology that is only useful when placed in the hands of a specific person or group and is not completely transferable to others is not a scientific methodology but an art. If this art can be learned by many (transferred) and yield the same results then it is science. This admonition just echoes that of Keynes's Septuagint analogy.

> It will be remembered that the seventy translators of the Septuagint were shut up in seventy separate rooms with the Hebrew text and brought out with them, when they emerged, seventy identical translations. Would the same miracle be vouchsafed if seventy multiple correlators were shut up with the same statistical material? (Keynes, pp.155-56).

I would like to propose a Septuagint type challenge. Let us set up a single blind experiment where someone plays the role of Laplace's demon and creates a DGP. Give the data to different modelers representing the PR approach and perhaps the Bayesian approach and see how well they fare. The PR approach tends to imply that if all are well trained in the PR approach all should get the same answer whereas the Bayesian approach allows for differences.

More General Methodological Problems for Econometrics

The lingering questions posed here are based on four fundamental methodological problems that all econometric methodologies must face: theory-laden observations (the statistical model specification is not independent of the underlying theory), Duhem's thesis (there is no way to specify and, therefore test, the statistical model independent of the theory), the theory underdetermination problem (multiple models can explain the same phenomenon), and metaphysical skepticism (how do we know when we have found the 'true' model). Presenting the PR approach with less certitude while acknowledging and attempting to address these fundamental issues would I believe strengthen the PR approach significantly.

In closing, what we are searching for is an econometric methodology that is coherent and has a recipe type structure that is agreed upon such that we are like Keynes's Septuagint translators: if 70 people are given the same data we will all end up with the same model. Unfortunately, we are not at that point. Many of our auxiliary assumptions are not tied to economic or statistical theory proper and are designed only to help bridge the theory-data gap in a specific application. Laudably, SM are attempting to bridge the theory-data gap by classifying and bringing the auxiliary assumptions in this nether region in to the fold of statistical theory or claim that statistical theory does have something to say about these auxiliary assumptions. While I applaud their efforts I am skeptical *any* econometric methodology can ease the model specification problem in isolation simply because I think the problem lies not with econometrics but with economic theory.

Consider what we do know about empirical economic analysis. We know our domain of empirical discourse will always involve analyzing multiple moments of probability distributions and yet most economic theories only make some vague statements about the first moment. Consequently, no matter how sophisticated the econometric methodology, because most theories say very little about the higher moments (e.g., variance) they are technically immune to arguments and refutations based on higher moments. Others have made this observation (see Davis 2000 for discussion and references) and for this reason, I think it is more productive to build the bridge from the theory side rather than the econometric side by requiring economic theories to say more about more moments of probability distributions.

References

Akaike, H. "Statistical Predictor Identification." *Annals of the Institute of Statistical Mathematics*. Vol. 22. (1970): 203-17.

Buse, A. "Aggregation, Distribution, and Dynamics in the Linear and Quadratic Expenditure Systems." *Rev. Econ. Stat.* Vol. 74. No. 1(February, 1992): 45-53.

Davis, G. C. "A Semantic Interpretation of Haavelmo's Structure of Econometrics." *Econ. and Phil.* Vol. 16. No. 1(October, 2000):205-228.

Haavelmo, T. "The Probability Approach in Econometrics." Supplement. 12. *Econometrica*.(1944): 1-115.

Hoover, K. and S. Perez. "Data Mining Reconsidered: Encompassing and the General-to-Specific approach to Specification Search." *Econometrics Journal*. Vol. 2. No. 2(1999):167-91.

Keynes, J. M. "On the Method of Statistical Business-Cycle Research: A Comment." *The Economic Journal*. Vol. 5. No. 197(March 1940): 154-156.

McGuirk, A., P. Driscoll, J. Alwang, H. L. Huang. "System Misspecification Testing and Structural-Change in the Demand for Meats." *J. Agr. Res. Econ* 20(July, 1995):1-21.

Nolt, J. *Logics*. Belmont, Ca: Wadsworth Publishing. 1997

Pratt, J. W. and R. Schlaifer. "On the Interpretation and Observation of Laws." *J. Econometrics*. Vol. 39. No. 1(January, 1988): 23-52.

Spanos, A. "Econometric Testing," in *The Handbook of Economic Methodology*. J. B. Davis, D. W. Hands, and Uskali Mäki, Eds., Northampton, MA: Edward Elgar. 1998.

________. "On Theory Testing in Econometrics: Modeling with Nonexperimental Data." *J. Econometrics*. Vol. 67. No. 1(May, 1995):189-226.

Schwarz, G. "Estimating the Dimensions of a Model." *Annals of Statistics*. Vol. 6. (1978):461-64.

“Comments on Geweke and McCausland, ‘Bayesian Specification Analysis in Econometrics’ ”

Walter N. Thurman[1]
North Carolina State University

Professors Geweke and McCausland (2001) gives us a lucid discussion of Bayesian model checking and a valuable lesson by example in its application. I learned much from both.[2] In the time I have to discuss their paper I will focus on four issues: a reason to be enthusiastic about the Bayesian methodology that they propound, the role of the likelihood function in model checking, the frequentist interpretability of the Bayesian methods, and the substantive lessons learned from the t-GARCH analysis.

1. Bayesian model checking promotes a focus on the economically relevant aspects of a model

Common treatments of specification testing suggest a battery of standard tests that routinely should be applied to models to test their adequacy. While there is an undeniable interest by practitioners in a standard set of diagnostic tests, this approach begs the question of the purpose of the model. As Geweke and McCausland (G&M) point out: all models are wrong, but some are useful. Further, their usefulness can only be gauged with respect to economic criteria, not statistical criteria. The important specification question is: what is the inferential purpose of the model and is its purpose compromised by particular aspects of lack of fit?

As G&M explain, Bayesian model checking begins with the notion of a complete model: the joint distribution of the observable data and the unobservable parameters, which is the product of the conditional density of the data and the prior density of the parameters. In their notation:

$$\mathbf{p(y, \theta_A \mid A) = p(y \mid \theta_A , A) \cdot p(\theta_A \mid A)}$$

The joint distribution displays all that we know regarding both parameter uncertainty and sampling variability of the data. All densities are conditional on A (the assumptions) of the model’s specification.

The complete specification of the joint distribution of y and θ_A allows any economically important function of the data to be simulated, conditional on the model. If one simulates synthetic samples in this way, one can calculate whatever lack-of-fit measure one is interested in, tabulate its probability distribution, and compare the distribution to the one realized value of the

[1] I thank George Davis, Dale Graybeal, and Matt Holt for useful discussions about Bayesian model checking.

[2] See also a more expansive review of simulation-based Bayesian methods in Geweke (1999).

statistic that comes from the sample.[3] The freedom afforded by this approach is truly liberating. In their example, G&M identify eight quantities of interest–interesting from a financial economics point of view–with which to assess a time series model of stock returns. They include measures of volatility, persistence, leverage, and distribution shape. This is a freedom not offered in general by a frequentist approach, which has no consistent recipe to follow to account for parameter uncertainty. Model checking in a frequentist framework is constrained by the search for diagnostic statistics that are pivotal quantities–statistics whose distribution does not depend upon unknown parameters.

The ability to focus on the economically relevant features of the model, instead of oftentimes arbitrary statistical measures of goodness of fit (e.g., R^2, Durbin-Watson statistics, and other largely residual-based diagnostics) allows an approach to modeling that is consistent with McCloskey's critique of significance testing in econometrics. She reminds us (1985 and elsewhere) that a p-value only tells how likely a test statistic was to have been observed by chance due to sampling variation. It doesn't reflect on the economic importance of the measured discrepancy from the null hypothesis. Ideally, Bayesian model checking involves a statistic the importance of which has been established and comparing its magnitude with what values a model might plausibly yield, given parameter and sampling uncertainty. It is specification testing with a focus on economic significance, not statistical significance.[4]

2. The dual role of the data in posterior predictive specification analysis

Model checking procedures require a quantity of interest and a probability model from which to simulate. The Bayesian method provides two possibilities for the simulation distribution: the prior predictive distribution and the posterior predictive distribution. The prior predictive distribution combines information from two sources only: the model specification and the prior distribution on the model parameters. It is uncontaminated by the data and so allows a

---

[3] Another possibility is to simulate a lack-of-fit measure that is a function of both observable data and unobservable parameters. Bayesian methods can deal with lack-of-fit measures that depend upon unknown parameters, but G&M's choice of quantities of interest are more conventional and only depend upon the data.

[4] Gelman et al. also suggest using the posterior predictive distribution to critique a model. They emphasize that one should choose quantities of interest to highlight important potential failings of the model. They suggest two criteria: "[1] Ideally, the test quantities T will be chosen to reflect aspects of the model that are relevant to the scientific purposes to which the inference will be applied. [2] Test quantities are commonly chosen to measure a feature of the data not directly addressed by the probability model..." (Gelman *et al.*, p. 172) Their second suggestion highlights the role of predictive distribution analysis in model checking as opposed to hypothesis testing. If one wishes to use a model directly to test a hypothesis, then one should construct a nested model that can account for the tested effect. Model checking, on the other hand, is directed toward testing the limits of the model in important directions.

clean comparison with the data; it is the methodology espoused by Box (1980). The posterior predictive distribution, on the other hand, optimally combines the information from the model, the prior distribution of model parameters, and the data via the likelihood. While the posterior predictive distribution is the best choice for the purposes of prediction conditional on the model, it seems less well suited to the task of model criticism. In the case of posterior predictive specification analysis, first the model is best fit to the data and then the fitted model is used to assess its own reasonableness. To this practitioner's eye this double counts the data. G&M perform both types of model checks but do not tell us which is preferred or how their results are to be reconciled should they conflict, as they do in the t-GARCH analysis.

Gelfand, Dey, and Chang (1992) also suggest using posterior predictive distributions for model checking, but modify the procedure in the following way: the quantities of interest they study are model residuals and the model-checking distribution they use is the posterior predictive distribution, conditional on all data points except the one for which the residual is calculated. Thus, in assessing how surprising the i^{th} residual is, one tabulates its predictive distribution conditional on all data points except the i^{th}. This procedure, akin to cross-validation, avoids the double reference to the data described above but is only directly applicable to quantities of interest that are functions of single observations, or subsets of observations. The quantities of interest analyzed by G&M are calculated from all observations and the Gelfand *et al.* logic is not directly applicable.

3. The role of pivotal statistics in prior predictive model checking

George Box (1980) proposed the use of the prior predictive distribution of y for model checking. He argued, as above, that the joint distribution of $(y, \theta_A|A)$ expressed the totality of uncertainty given the model. He laid out a methodology of science that iterated between the two activities of estimation and criticism. The posterior distribution $p(\theta_A|y, A)$ is the foundation of estimation; the prior predictive distribution $p(y|A)$ is the foundation of criticism. The product of the two is $p(y, \theta_A| A)$. For criticism, he argued that one should focus on the degree to which the observed value of a quantity of interest is surprising in light of hypothetical sampling from the prior predictive distribution. To Box, the explicit representation of parameter uncertainty in the form of a prior distribution, $p(\theta_A|A)$, is a necessary step to assess the reasonableness of the model indexed by A.[5]

For a Bayesian who is willing to specify a prior distribution explicitly, prior predictive model checking is natural. It is not so for a frequentist, who views parameters as unknown but fixed. But a frequentist can feel at home with prior predictive model checking in one case: that

---

[5] Box's approach is interesting because it apportions to Bayes' theorem the job of model estimation and to (an augmented) sampling theory the job of model criticism. Bayesians, who teach that performance in hypothetical repeated sampling is irrelevant when it comes to parameter estimation, adopt just this criteria when it come to assessing the reasonableness of a model.

in which the quantities of interest are pivotal–statistics that are functions only of the data and with distributions that do not depend upon unknown parameters. Because parameter values are irrelevant, so is the specification of a prior distribution. This is the case of G&M's first example, in which they model daily stock returns as i.i.d. drawings from a normal distribution. In this case, and as they point out, their quantities of interest (volatility, decay, leverage, etc.) are pivotal. Because their distributions do not depend on μ and σ^2, uncertainty over them plays no role in determining the extent to which observed values of the quantities of interest are surprising. Only sampling variability matters. As a practical matter, one simulates from the prior predictive distribution by fixing the values of μ and σ^2 at arbitrary values and simulating the data as i.i.d. normal random variates. As a methodological principle, one is simply simulating the data generating process, a procedure entirely consistent with sampling theory given known parameters.

When the amicable coexistence of Bayesian and sampling theory techniques breaks down is when G&M introduce the t-GARCH model as an alternative to the i.i.d. normal. At that point, the quantities of interest lose their pivotal qualities and their distributions depend upon the values taken by the expanded set of t-GARCH parameters. A frequentist practitioner who was nodding along with the normal simulations will not be so sanguine about drawing simulated parameters from the log-normal, beta, and adjusted chi-square distributions, which are G&M's choices of priors. The frequentist practitioner might wonder if there isn’t some circularity involved in picking prior parameter values, as G&M do, to fit the simulated model to match the sample median of the data.

But regardless of the practitioner’s unease, if the model does not imply that the quantities of interest are pivotal, then one must adopt the explicitly Bayesian point of view with informative proper priors. To contrast the situation with posterior predictive model checking, the priors there need not be informative. Because one simulates from the posterior distribution, the likelihood function forces the posterior to be proper and improper priors can be used. To the extent that a sampling theorist is more comfortable with diffuse priors than informative priors, he will be more comfortable with posterior predictive model checking than he will be with prior predictive model checking.

4. On interpreting the stock return data

What should one conclude when posterior predictive analysis calls the model into question in important ways and prior predictive analysis does not? This is a reasonable interpretation of G&M’s analysis of the t-GARCH model. In the prior analysis, all nine sample quantities of interest lie within the 98% prior probability intervals; half lie within the 50% prior probability intervals. From these results one might reasonably conclude that none of the checked features of the data are grossly at odds with the complete model. The situation is different with the posterior predictive distribution, where three of the nine sample values lie outside of the 98% posterior probability intervals and six of the nine lie outside the 50% intervals. The posterior intervals much more strongly disagree with the sample quantities of interest and the strongest rejections of the model concern the volatility decay and kurtosis. As G&M point out, posterior analysis of the model implies too little volatility persistence and too much leptokurtosis. But the prior predictive analysis provides no such rejection with respect to leptokurtosis and only weaker rejection with respect to volatility decay.

Another interesting comparison between the prior and predictive analysis of the t-GARCH model is that the posterior probability intervals for volatilities and volatility decay are narrower than the prior intervals, a situation consistent with the intuition that the posterior distributions result from a fitting to the data. But counter to that intuition, the posterior intervals are wider than the prior intervals for the kurtosis, skewness, and leverage measures. My intuition fails me here.

A last question I would raise concerns the interpretation of the leverage ahead measure offered in the paper. G&M say that one of the outstanding statistical characteristics of financial returns is "the 'leverage' phenomenon in which extreme negative returns are more likely to presage high volatility than are similarly extreme positive returns" (Geweke and McCausland [2001], p. 6). This result can also be found in Nelson (1989). But the sample value of the leverage ahead measure, which is a sample correlation between today's return and tomorrow's squared return, is positive (.0312). This implies that a below average return today forecasts a below average squared return tomorrow: the opposite of the leverage effect described in the text.

5. Conclusion

There is much that applied economists can learn from Geweke's and McCausland's paper and the references they cite, and I predict that we will. The Bayesian world view is appealing to most applied economists and increases in computing power continue to lower the relative cost of Bayesian methods. As the practical advantage tilts toward Bayesian methods more generally, Bayesian specification analysis is bound to become a standard method by which we assess and discuss econometric models.

References

Box, George E.P. (1980). "Sampling and Bayes' Inference in Scientific Modeling and Robustness," *Journal of the Royal Statistical Society*. *Series A (General)* 43(4): 383-430.

Gelfand, A.L., D.K. Dey, and H. Chang (1992). "Model Determination Using Predictive Distributions with Implementation via Sampling-Based Methods," in *Bayesian Statistics 4*. Cambridge: Oxford University Press.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin (1995). *Bayesian Data Analysis*. London: Chapman & Hall.

Geweke, John and William McCausland (2001). "Bayesian Specification Analysis in Econometrics," paper prepared for presentation at the 2001 meetings of the American

Agricultural Economics Association, Chicago, Illinois.

Geweke, John (1999). "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication" (with discussion and rejoinder), *Econometric Reviews* 18: 1-126.

McCloskey, Deirdre N. (1985). "The Loss Function Has Been Mislaid–The Rhetoric of Significance Tests," *American Economic Review* 75(2): 201-205.

Nelson, Daniel B. (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica* 59(2): 347-370.