**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**

# Bayesian Specification Analysis in Econometrics

John Geweke
University of Iowa

October 2, 2001

**Abstract**

This article describes the process of Bayesian specification analysis using state of the art simulation methods. It distinguishes between predictive specification analysis (comparison of model predictions with actual outcomes) and post-predictive specification analysis (comparison of predictions of a replication of an experiment with actual outcomes). The methods are illustrated using a specific example, evolving volatility in financial returns.

Partial financial support from National Science Foundation grants SBR-9731037 and SES-9996332 is gratefully acknowledged. George Chang provided help with the computations and useful discussions. The usual disclaimers apply.

Keywords: Decision making, Posterior simulation, Predictive

JEL classification: Primary C52; secondary C11



## 1 Introduction

All econometric models are wrong, but some are useful. They mirror only certain aspects of reality, and these imperfectly. But in complex situations they can provide structure and clarity that improves decisions. The process of discovering the reliability with which various aspects of reality are accommodated in an econometric model is known as specification analysis. It is a vital step in learning about the properties of a given model, in determining whether to use a model in actual decision making, and in improving models and, thereby, decisions.

This article outlines the essentials of Bayesian specification analysis, as practiced using state of the art simulation methods. This approach is especially pertinent to models used for decision making, because Bayesian inference is the econometric cornerstone of decision making within the expected utility theory on which virtually all modern economics is constructed ([6], [22]). It is attractive in practical work because it is straightforward to apply and can be used to

study the congruence of any complete econometric model with relevant features of the data.

This topic has received close attention in the Bayesian mathematical statistics literature but is less well known among practicing applied economists. There are two approaches, sometimes combined in various ways. The first is to ask: having expressed a model, what are its predictions for observables, before using data for inference? How consistent are these predictions with the actual data? The classic exposition of this approach is [4], and a recent treatment using modern computational techniques is [11]. Combining the terminology of this literature and that of econometrics, we use the term *predictive specification analysis* for this approach here.

The second approach is to ask: having expressed a model, and having used the observed data for inference about its parameters, what would we predict would happen in an independent replication of the observables? (In a prediction problem this is similar to asking what would happen in the next $T$ observations, where $T$ is the size of the original sample.) The event that the replications are quite different from what was actually observed, for some interesting aspect of the data, constitutes the notion of surprise: this idea has been developed in a series of important studies including [14], [16], [15], [20] and [2]. We use the term *post-predictive specification analysis* for this approach.

Section 2 of this article briefly reviews the conceptual foundations of Bayesian inference in econometrics. Since the ideas treated in this paper are best first understood by means of examples, Section 3 sets out some models of changing volatility in financial returns and introduces a data set for their application. Predictive specification analysis is taken up in Section 4, which includes an application to evolving volatility. Section 5 turns to the formal comparison of alternative models. The process of post-predictive specification analysis is set out in Section 6 and applied to evolving volatility models. Section 7 concludes.

# 2 Bayesian analysis

The concept of a complete model is central to Bayesian specification analysis. A complete model specifies the distribution of a $T \times 1$ vector of observable random variables, $\mathbf{y} \in Y$, by means of a $k_A \times 1$ vector of unknown parameters $\theta_A \in \Theta_A$. Conditional on $\theta_A$ the probability density of the observables is

$$p\left(\mathbf{y} \mid \boldsymbol{\theta}_A, A\right) = \prod_{t=1}^{T} p\left(y_t \mid y_1, ..., y_{t-1}, \boldsymbol{\theta}_A, A\right) \tag{1}$$

This function is familiar from non-Bayesian analysis - after $\mathbf{y}$ is observed, and the observed value $\mathbf{y}^o$ replaces the argument $\mathbf{y}$, (1) becomes the likelihood function.

A complete model includes a prior density for the unobservable parameters, $p\left(\boldsymbol{\theta}_A \mid A\right)$. The combination of the prior density and the density for observables (1) provides the joint density of parameters and observables:

$$p(\boldsymbol{\theta}_A, \mathbf{y} \mid A) = p(\boldsymbol{\theta}_A \mid A)\, p(\mathbf{y} \mid \boldsymbol{\theta}_A, A). \tag{2}$$

This joint distribution is the key to Bayesian inference and specification analysis. The marginal density for $\mathbf{y}$,

$$p(\mathbf{y} \mid A) = \int_{\Theta_A} p(\boldsymbol{\theta}_A, \mathbf{y} \mid A)\, d\boldsymbol{\theta}_A, \tag{3}$$

is the model predictive density of the observable $\mathbf{y}$. Conditional on observed $\mathbf{y} = \mathbf{y}^o$,

$$p(\theta_A \mid \mathbf{y}^o, A) = p(\theta_A, \mathbf{y}^o) / p(\mathbf{y}^o \mid A), \tag{4}$$

the posterior density of the unknown parameter vector $\boldsymbol{\theta}_A$.

The third and final component of a complete model is a vector of interest $\boldsymbol{\omega}$. Elements of $\boldsymbol{\omega}$ may include transformations of parameters $\boldsymbol{\omega} = \mathbf{g}(\boldsymbol{\theta}_A; A)$, for example the value of returns to scale in a production function. They also include observables whose values are not yet known, for example $\boldsymbol{\omega} = (y_{T+1}, y_{T+2}, y_{T+3})'$ in a forecasting problem. A complete model provides the density $p(\boldsymbol{\omega} \mid \boldsymbol{\theta}_A, \mathbf{y}, A)$. Then the model predictive density of $\boldsymbol{\omega}$ is

$$p(\boldsymbol{\omega} \mid A) = \int_{\Theta_A} \int_{Y} p(\boldsymbol{\theta}_A, \mathbf{y})\, p(\boldsymbol{\omega} \mid \boldsymbol{\theta}_A, \mathbf{y}, A)\, d\mathbf{y} d\boldsymbol{\theta}_A, \tag{5}$$

and the posterior density of the vector of interest is

$$p(\boldsymbol{\omega} \mid \mathbf{y}^o, A) = \int_{\Theta_A} p(\boldsymbol{\theta}_A \mid \mathbf{y}^o, A)\, p(\boldsymbol{\omega} \mid \boldsymbol{\theta}_A, \mathbf{y}^o, A)\, d\boldsymbol{\theta}_A. \tag{6}$$

Expressions like (6) are not immediately useful in applications, because the integral on the right side typically cannot be obtained in closed form. In modern Bayesian analysis this problem is usually obviated by means of a posterior simulator. This is an algorithm that produces random vectors $\boldsymbol{\theta}_A^{(m)} (m = 1, 2, ...)$ whose distribution corresponds to the posterior density (4)[1]. Simulation from the other three distributions, corresponding to the prior density $p(\boldsymbol{\theta}_A \mid A)$, the observables density $p(\mathbf{y} \mid \boldsymbol{\theta}_A, A)$, and the vector of interest density $p(\boldsymbol{\omega} \mid \boldsymbol{\theta}_A, \mathbf{y}, A)$ is typically straightforward. (Some examples will be offered subsequently.) Taken together, these simulators make it possible to generate synthetic random vectors from each of these distributions above. For example, to simulate from (3), draw $\widetilde{\boldsymbol{\theta}}_A \sim p(\boldsymbol{\theta}_A \mid A)$ and then $\widetilde{\mathbf{y}} \sim p\left(\mathbf{y} \mid \widetilde{\boldsymbol{\theta}}_A, A\right)$; to simulate from (6), draw $\widetilde{\boldsymbol{\theta}}_A \sim p(\boldsymbol{\theta}_A \mid \mathbf{y}^o, A)$ and then $\widetilde{\boldsymbol{\omega}} \sim p\left(\boldsymbol{\omega} \mid \widetilde{\boldsymbol{\theta}}_A, \mathbf{y}^o, A\right)$. These capabilities are central to implementing the program of Bayesian specification analysis described in this article.

[1] These algorithms use varied methods, some quite sophisticated, to achieve this correspondence. Moreoever, the nature of the correspondence varies with the simulator. These details are beyond the scope of this article, but there are quite a few accessible surveys and texts, including [8], [5] and [10].

# 3 Some alternative data distributions

The modeling of returns to financial assets has emerged as a challenging problem of considerable practical importance in recent years. The importance stems from such decision problems as the pricing of financial derivatives like options and strategies for avoiding risk such as hedging. In all of these situations the distribution of asset returns, and in particular the conditional distribution of future asset returns, is central to rational decision making. The problem is interesting, because the distributions in question are clearly not normal and do not appear to be of any other simple form. There is also strong evidence that the spread of distributions, and possibly the shape, changes as conditioning information evolves.

A leading model that captures some of these characteristics for a single asset return is the generalized autoregressive conditional heteroskedasticity (GARCH) model introduced by Bollserslev [3]. Denote the observable asset return from period $t-1$ to period $t$ by $y_t$ and the variance of $y_t$ at time $t$ by $h_t$. Given $h_t$,

$$y_t \sim N(\mu, h_t) \tag{7}$$

and variance evolves as

$$h_t = \alpha + \gamma (y_{t-1} - \mu)^2 + \delta h_{t-1}. \tag{8}$$

So long as $\mid \gamma + \delta \mid < 1$ the return series $\{y_t\}$ is stationary and displays periods of both high volatility (large values of $(y_t - \mu)^2$) and low volatility (small $(y_t - \mu)^2$) relative to its unconditional variance of $\alpha / (1 - \gamma - \delta)$. The unconditional distribution of $y_t$ is non-normal - it displays excess kurtosis, and the accompanying "fat tails" in its unconditional probability density relative to the normal.

The GARCH model so formulated still tends to display an unconditional distribution with insufficient weight in its tails relative to data histograms. This defect is often obviated by substituting

$$y_t \sim t(\mu, h_t; \nu) \tag{9}$$

in place of (7). The Student-$t$ distribution is more leptokurtic than the normal distribution. As the degrees of freedom parameter $\nu$ falls, the tails of the distribution fatten. Conditional moments of $y_t$ exist only up to order $\nu$, so that $\nu > 2$ is required for finite variance, $\nu > 4$ for a finite fourth moment, and so on. The $t$-GARCH model, consisting of (9) and (8) is widely regarded as the best fitting simple model for financial asset returns.

An alternative model for financial returns is the Markov normal mixture model ([19],[21]). In this model there is a latent state $s_t$ associated with each period $t$, taking on one of the values $s_t = 1, ..., m$. If $s_t = i$, then

$$y_t \sim N\left(\mu_i, \tau_i^2\right) \tag{10}$$

The transition between states is governed by a first order discrete Markov chain,

$$P(s_t = j \mid s_{t-1} = i) = p_{ij}.$$

If $p_{ii} >> p_{ij}\,(j \neq i)$ for most $i$, and if there is a substantial difference in the $\mu_i$, or (particularly) in the $\sigma_i^2$ from state to state, then $y_t$ will display persistence in volatility.

The number of states in the Markov normal mixture model can be chosen to be as large as desired. As a consequence, the model shares the flexibility of many nonparametric methods. Nevertheless it imposes some constraints that are uncharacteristic of financial asset returns. The most important is that the return series $\{y_t\}$ is serially uncorrelated (a good approximation for many return series) if and only if $\mu_1 = ... = \mu_m \equiv \mu$, but then the distribution of $y_t$ is necessarily symmetric about $\mu$ (not so good). This and similar problems can be avoided, and greater flexibility achieved, by substituting a more general distribution for (10).

This extension is accomplished in the compound Markov normal mixture model [13], which is doubly indexed by a latent state vector $\mathbf{s}_t = (s_{t1}, s_{t2})'$. Given $\mathbf{s}_t = (i, j)'$,

$$y_t \sim N\left(\beta + \phi_i + \psi_{ij}, \sigma^2\sigma_i^2\sigma_{ij}^2\right). \tag{11}$$

The first index $s_{t1}$ evolves exactly as does the index $s_t$ in the Markov normal mixture model. The second index $s_{t2}$ depends only on the contemporaneous $s_{t1}$, with

$$P\left(s_{t2} = j \mid s_{t1} = i\right) = r_{ij}.$$

Thus the compound Markov normal mixture model is the same as the Markov normal mixture model, but with (10) replaced by a mixture of normals distribution.[2]

# 4 Predictive specification analysis

The GARCH and mixture distributions do not, alone, constitute complete models. It remains to specify prior distributions, as well as a vector of interest. The process of predictive specification analysis consists of three steps: choosing a vector of interest that summarizes interesting aspects of the data, $\boldsymbol{\omega}$; (2) selecting a trial prior distribution; and (3) examining the implications of the prior distribution for the vector of interest $\boldsymbol{\omega}$. Steps (2) and (3) may be repeated, experimenting with different prior distributions. This process is best described by illustration.

The outstanding statistical characteristics of returns to financial assets tend to be changing but persistent volatility, excessive leptokurtosis relative to the normal distribution, and the "leverage" phenomenon in which extreme negative returns are more likely to presage high volatility than are similarly extreme positive returns. For some return series, the distribution of asset returns is also skewed to the left.

These characteristics can all be captured through transformations $\boldsymbol{\omega}$ of the observable returns $\mathbf{y}$. The transformed observables used in this study, which

[2] A detailed discussion of this model, including the apparent identification problem in (11) is given in [13].

constitute the vector of interest, are detailed in Table 1. The first three functions capture persistence in volatility: $\omega_1$ is the first order sample autocorrelation coefficient of the series $\left\{y_t^2\right\}$, and $\omega_2$ is the 20th order sample autocorrelation coefficient. The function $\omega_3$ is the ratio of $\omega_2$ to $\omega_1$, which is one aspect the rate of decay of the sample autocorrelation function of $\left\{y_t^2\right\}$. The next two elements of $\boldsymbol{\omega}$ address leptokurtosis: $\omega_4$ is the sample excess kurtosis coefficient and $\omega_5$ is the ratio of the range statistic to the interquartile range. The sample skewness is $\omega_6$. Leverage phenomena are captured by $\omega_7$, the sample correlation between $y_t$ and $y_{t+1}^2$, and $\omega_8$, the sample correlation between $y_t$ and $y_{t-1}^2$. The last element of the vector of interest, $\omega_9$, is simply the sample standard deviation, whose importance will become clear shortly.

**Table 1**
Definition of vector of interest

Preliminary statistics:

$\overline{y}_T = \sum_{t=1}^{T} y_t/T \qquad s_T = \sum_{t=1}^{T} \left(y_t - \overline{y}_T\right)^2/T$

$\overline{y}_T^{(2)} = \sum_{t=1}^{T} y_t^2/T \qquad s_T^{(2)} = \sum_{t=1}^{T} \left(y_t^2 - \overline{y}_T^{(2)}\right)^2/T$

| | | |
|---|---|---|
| $\omega_1$ | First order volatility | $\sum_{t=1}^{T-1} \left(y_t^2 - \overline{y}_T^{(2)}\right)\left(y_{t+1}^2 - \overline{y}_T^{(2)}\right)/T$ |
| $\omega_2$ | 20th order volatility | $\sum_{t=1}^{T-20} \left(y_t^2 - \overline{y}_T^{(2)}\right)\left(y_{t+20}^2 - \overline{y}_T^{(2)}\right)/T$ |
| $\omega_3$ | Volatility decay | $\omega_2/\omega_1$ |
| $\omega_4$ | Excess kurtosis | $\sum_{t=1}^{T} \left(y_t - \overline{y}_T\right)^4/T \left(s_T\right)^2 - 3$ |
| $\omega_5$ | Quantile ratio | $\left(y_{(T)} - y_{(1)}\right) / \left(y_{(3T/4)} - y_{(T/4)}\right)$ |
| $\omega_6$ | Skewness | $\sum_{t=1}^{T} \left(y_t - \overline{y}_T\right)^3/T \left(s_T\right)^{3/2}$ |
| $\omega_7$ | Leverage ahead | $\sum_{t=1}^{T-1} \left(y_t - \overline{y}_T\right)\left(y_{t+1}^2 - \overline{y}_T^{(2)}\right)/T \left(s_T \cdot s_T^{(2)}\right)^{1/2}$ |
| $\omega_8$ | Leverage behind | $\sum_{t=2}^{T} \left(y_t - \overline{y}_T\right)\left(y_{t-1}^2 - \overline{y}_T^{(2)}\right)/T \left(s_T \cdot s_T^{(2)}\right)^{1/2}$ |
| $\omega_9$ | Standard deviation | $\left(s_T\right)^{1/2}$ |

To illustrate Bayesian specification analysis, we shall apply the GARCH and mixture models described in the previous section to the daily returns of the Standard and Poors 500 index used in [21]. It extends from January 3, 1928, through April 29, 1991, a total of 17,052 observations. Returns are formed as $y_t = \log\left(p_t/p_{t-1}\right)$, where $p_t$ is the daily index; see [21] for complete details. Table 2 provides, in the first column, the observed values $\boldsymbol{\omega} = \boldsymbol{\omega}^o$ for this period.

To interpret these values, consider a model for returns that is much simpler than those described in the previous section: $y_t \sim N\left(\mu, \sigma^2\right)$ i.i.d. The sampling distribution of $\omega_1, ..., \omega_8$ does not depend on the value of $\mu$ or $\sigma^2$. Therefore the model predictive distribution of each of $\omega_1$ through $\omega_8$, for a sample size of 17,052 (or any other size) will be the same no matter what the prior distribution of the unknown parameters $\mu$ and $\sigma^2$. Some quantiles of this distribution are also indicated in Table 2.

**Table 2**
Predictive distribution of vector of interest
Gaussian i.i.d. model

| | | Data | Median | (25%, 75%) | (1%, 99%) |
|---|---|---|---|---|---|
| $\omega_1$ | 1st volatility | .218 | .000 | (-.005, .005) | (-.018, .018) |
| $\omega_2$ | 20th volatility | .083 | .000 | (-.005, .005) | (-.018, .018) |
| $\omega_3$ | Volatility decay | .382 | .000 | (-1.00, .982) | (-31.3, 31.9) |
| $\omega_4$ | Excess kurtosis | 22.4 | -.001 | (-.026, .024) | (-.083, .090) |
| $\omega_5$ | Quantile ratio | 39.9 | 5.874 | (5.686, 6.089) | (5.301, 6.761) |
| $\omega_6$ | Skewness $\times$ 100 | -.373 | .000 | (-.010, .010) | (-.030, .030) |
| $\omega_7$ | Leverage ahead | .0312 | .0000 | (-.0052, .0053) | (-.0177, .0179) |
| $\omega_8$ | Leverage behind | -.0752 | .0000 | (-.0051, .0052) | (-.0180, .0176) |
| $\omega_9$ | Stan deviation | .0115 | | | |

The failure of the normal model is strikingly evident. None of the characteristics of the data captured in the functions of interest $\boldsymbol{\omega}$ are consistent with an i.i.d. normal model. The observed values of volatility $\left(\omega_1^o, \omega_2^o, \omega_3^0\right)$, the thickness of the tails of the distribution $(\omega_4^o, \omega_5^o)$, skewness $(\omega_6^o)$ and leverage $(\omega_7^o, \omega_8^o)$ are so improbable as to be impossible for most practical purposes.

In the GARCH model, as well as in the other models considered in this article, the model predictive distribution of $\boldsymbol{\omega}$ depends on the prior density $p\left(\boldsymbol{\theta}_A \mid A\right)$ as well as the data density (1). This is generally the case, as is evident from (2) and (5). It is therefore necessary to specify a proper prior distribution for the parameters $\mu$, $\alpha$, $\gamma$ and $\delta$ of the GARCH model. Given the prior density $p\left(\boldsymbol{\theta}_A \mid A\right)$, the simulation $\widetilde{\boldsymbol{\theta}}_A \sim p\left(\boldsymbol{\theta}_A \mid A\right)$ followed by $\widetilde{\mathbf{y}} \sim p\left(\mathbf{y} \mid \widetilde{\boldsymbol{\theta}}_A, A\right)$ followed by the computation of $\widetilde{\boldsymbol{\omega}}$ as indicated in Table 1 (with $\widetilde{y}_t$ in place of $y_t$) produces a single drawing from the predictive distribution with density (5). Repetition of this process many times provides quantiles for the predictive density $p\left(\boldsymbol{\omega} \mid A\right)$.

This exercise is useful for several purposes. First, it yields an informal indication of whether a GARCH model is capable of accounting for the observed values of the vector of interest $\boldsymbol{\omega}^o$ provided in column 3 of Table 2. If, after experimenting with alternative prior distributions, we found that some of these values were well outside (say) the centered 98% predictive interval, we would question the ability of any model with a GARCH data density to account for the salient observed features of the data. We might then carry out the exercise with alternative models, rather than invest resources in developing a posterior simulator (or even non-Bayesian estimation methods) for the GARCH model. Thus, this exercise can be an effective part of research strategy in applied econometrics.

**Table 3**
Predictive distribution of vector of interest
GARCH model

| | | Data | Median | (25%, 75%) | (1%, 99%) |
|---|---|---|---|---|---|
| $\omega_1$ | 1st volatility | .218 | .335 | (.154, .468) | (.003, .710) |
| $\omega_2$ | 20th volatility | .083 | .0013 | (-.004, 010) | (-.016, .188) |
| $\omega_3$ | Volatility decay | .382 | .004 | (-.123, .044) | (-.568, .849) |
| $\omega_4$ | Excess kurtosis | 22.4 | 1.49 | (.206, 10.9) | (-.043, 662) |
| $\omega_5$ | Quantile ratio | 39.9 | 11.03 | (6.89, 23.4) | (5.52, 151) |
| $\omega_6$ | Skewness | -.004 | .000 | (-.0003, .0003) | (-.037, .034) |
| $\omega_7$ | Leverage ahead | .0312 | .000 | (-.013, .013) | (-.164, .168) |
| $\omega_8$ | Leverage behind | -.0752 | .000 | (-.014, .013) | (-.194, .191) |
| $\omega_9$ | Stan deviation | .0115 | .013 | (.006, .031) | (.001, .245) |

The GARCH model does not fail in this way. Table 3 provides quantiles for the predictive density of $\boldsymbol{\omega}$ for a GARCH model with the prior distribution

$$\log(\alpha) \sim N(-10, 2.2), \tag{12}$$

$$(\gamma, \delta, 1-\gamma-\delta) \sim Beta(1,1,1). \tag{13}$$

The distribution (13) corresponds to a "flat" prior for $\gamma$ and $\delta$, defined on the unit simplex $\{\gamma > 0,\ \delta > 0,\ \gamma + \delta < 1\}$. Given this prior, (12) was chosen to provide a range of sample standard deviations ($\omega_9$) roughly centered at the observed value of .0115 but including a wide range of alternative values. Alternatives to (13)—for example, Beta distributions with somewhat different values of the parameters—had distinct but small effects on the predictive distribution of $\boldsymbol{\omega}$.

This exercise serves two other purposes as well. First, it provides a concrete interpretation of the model and an understanding of the prior distribution. Any model must be understood ultimately in terms of its implications for observables, and the construction of $p(\boldsymbol{\omega} \mid A)$ does precisely that. As changes in the prior are made, their effect on $p(\boldsymbol{\omega} \mid A)$ can be studied. This is essential for a serious subjective Bayesian, and a useful learning exercise for anyone interested in the properties of the model.

The other purpose served by the predictive density $p(\boldsymbol{\omega} \mid A)$ is to provide a common ground in setting prior distributions for alternative models. For example, suppose the prior distribution for $\alpha$ were $\log(\alpha) \sim N(0, 2.2)$ rather than (12). Then the GARCH model would suffer in comparison with other models, because the variance of the observable would be pushed toward unreasonable values by the prior distribution. As will be seen in the next section, it is important to guard against such situations, which can arise unwittingly if priors are not taken seriously.

For these reasons we repeat the exercise of drawing from the predictive density $p(\boldsymbol{\omega} \mid A)$ for the other three models under consideration. The prior distribution for the $t$-GARCH model begins with the same prior distribution (13)

for $\gamma$ and $\delta$. The prior distribution for the degrees of freedom parameter $\nu$ is $\nu - 4 \sim \chi^2(4)$. The restriction $\nu > 4$ ensures the existence of population conditional fourth moments; without this restriction, the sample measures of volatility and excess kurtosis, for $T = 17,052$, can become so large as to overflow computer floating point representation. The prior distribution of $\alpha$ was again chosen to bring $\omega_9^o$ well within its support; here, $\log(\alpha) \sim N(-12, 2.2)$.

**Table 4**
Predictive distribution of vector of interest
$t$-GARCH model

| | | Data | Median | (25%, 75%) | (1%, 99%) |
|---|---|---|---|---|---|
| $\omega_1$ | 1st volatility | .218 | .334 | (-.187, .459) | (.004, .718) |
| $\omega_2$ | 20th volatility | .083 | .001 | (-.002 .012) | (-.014, .244) |
| $\omega_3$ | Volatility decay | .382 | .004 | (-.007, .052) | (-.458, .935) |
| $\omega_4$ | Excess kurtosis | 22.4 | 17.2 | (3.23, 196) | (.634, 3607) |
| $\omega_5$ | Quantile ratio | 39.9 | 29.88 | (14.6, 113) | 7.98, 6.0 $\times 10^5$) |
| $\omega_6$ | Skewness | -.004 | .000 | (-.002, .002) | (-.204, .211) |
| $\omega_7$ | Leverage ahead | .0312 | .000 | (-.027, .026) | (-.343, .344) |
| $\omega_8$ | Leverage below | -.0752 | .000 | (-.028, .027) | (-.350, .352) |
| $\omega_9$ | Stan deviation | .0115 | .008 | (.003, .025) | (.001, 2266) |

The quantiles shown in Table 5 indicate that $t$-GARCH can also account for the observed $\omega_i^o$. Moreover comparison of Tables 4 and 5 provides a quick study in the properties of the $t$-GARCH model relative to GARCH: about the same scope for persistence in volatility, much thicker tails in the unconditional distribution of $y_t$, a somewhat greater capacity for skewness and leverage, and a much wider range of sample variance in $y_t$.

Whereas the GARCH model has only a few parameters, compound Markov normal mixture models can have many. The two instances studied here are $m_1 = 6$, $m_2 = 3$ and $m_1 = m_2 = 7$. (The next section sets forth the reason for these choices.) The former model has 96 free parameters, and the latter 203.

Following [13] the prior distributions in these models are symmetric across the states. In the transition matrix $\mathbf{P}$ the rows are independent, each with a Dirichlet (multivariate Beta) distribution. The parameters of the Dirichlet distribution for row $i$ are all the same ($r$) except for position $i$, where the parameter value is $r^*$. By choosing $r^* > r$, the prior incorporates the purpose of $\mathbf{P}$, which is to provide persistence in states. The prior is centered about the probability $(m_1 - 1)\, r / [r^* + (m_1 - 1)\, r]$ of the current persistent state changing each period; this corresponds to a mean duration time of $1 + r^* / (m_1 - 1)\, r$ periods in each state. The prior here has $r = 1$, $r^* = 40$, so when $m_1 = 6$ the mean prior probability of changing persistent states each period is 1/9.

The row of the probability matrix $\mathbf{R}$ for choices of transitory states are also independent in the prior, with all parameter values equal to $r$. Since $r = 1$, this is equivalent to a "flat" prior on the $m_1$-dimensional unit simplex. The variance parameters $\sigma^2$, $\sigma_1^2$ and $\sigma_2^2$ all have inverted gamma distributions: $.75/\sigma_i^2 \sim$

$\chi^2(.75)$, $.75/\sigma_{ij}^2 \sim \chi^2(.75)$, and the parameters of the prior distribution for $\sigma^2$ are chosen to make the predictive distribution for $\omega_9$ include the observed standard deviation, but also have a wide range – just as was done with the prior for $\alpha$ in the GARCH models. These prior distributions also identify $\sigma^2$, $\sigma_1^2$ and $\sigma_2^2$ in the data density (11) – that is, they render the posterior distribution proper. (A detailed discussion of these and related points is given in [13].) The prior distribution of the compound normal mixture model is completed with the independent priors $\phi_i \mid \sigma^2 \sim N\left(0, \sigma^2\right)$ and $\varphi_{ij} \mid \left(\sigma^2, \sigma_i^2\right) \sim N\left(0, \sigma^2\sigma_i^2\right)$.

**Table 5**
Predictive distribution of vector of interest
Compound Markov normal mixture model, $m_1 = 6$, $m_2 = 3$

| | | Data | Median | (25%, 75%) | (1%, 99%) |
|---|---|---|---|---|---|
| $\omega_1$ | 1st volatility | .218 | .100 | (.054, .155) | (-.001, .330) |
| $\omega_2$ | 20th volatility | .083 | .008 | (.001, .020) | (-.010, .071) |
| $\omega_3$ | Volatility decay | .382 | .085 | (.005, .191) | (-.752, 1.65) |
| $\omega_4$ | Excess kurtosis | 22.4 | 27.6 | (14.1, 58.4) | (2.52, 710) |
| $\omega_5$ | Quantile ratio | 39.9 | 189.9 | (67.5, 732) | $(12.9, 1.0 \times 10^5)$ |
| $\omega_6$ | Skewness | -.004 | .000 | (-.003, .003) | (-.041, .041) |
| $\omega_7$ | Leverage ahead | .0312 | .000 | (-.012, .012) | (-.056, .055) |
| $\omega_8$ | Leverage behind | -.0752 | .000 | (-.012, .012) | (-.056, .056) |
| $\omega_9$ | Stan deviation | .0115 | .048 | (.016, .195) | (.003, 29.9) |

**Table 6**
Predictive distribution of vector of interest
Compound Markov normal mixture model, $m_1 = 7$, $m_2 = 7$

| | | Data | Median | (25%, 75) | (1%, 99%) |
|---|---|---|---|---|---|
| $\omega_1$ | 1st volatility | .218 | .043 | (.018, .075) | (-.002, .213) |
| $\omega_2$ | 20th volatility | .083 | .001 | (-.002, .008) | (-.008, .043) |
| $\omega_3$ | Volatility decay | .382 | .028 | (-.071, .181) | (-3.34, 4.32) |
| $\omega_4$ | Excess kurtosis | 22.4 | 77.8 | (39.8, 171) | (9.15, 2041) |
| $\omega_5$ | Quantile ratio | 39.9 | 603 | (205, 2432) | $(35.7, 3.7 \times 10^5)$ |
| $\omega_6$ | Skewness | -.004 | .000 | (-.006, .006) | (-.089, .101) |
| $\omega_7$ | Leverage ahead | .0312 | .000 | (-.010, .010) | (-.046, .049) |
| $\omega_8$ | Leverage behind | -.0752 | .000 | (-.010, .010) | (-.048, .049) |
| $\omega_9$ | Stan deviation | .0115 | .1106 | (.036, 457) | (.005, 76.2) |

These prior distributions are best understood in terms of their implications for observables - the density $p\left(\boldsymbol{\omega} \mid A\right)$. Table 5 provides quantiles for these distributions in the case $m_1 = 6$, $m_2 = 3$, while Table 6 does so for $m_1 = m_2 = 7$. The observed persistence in volatility lies near the upper range of the distribution (especially $\omega_2$ in the $m_1 = m_2 = 7$ model) but are by no means effectively ruled out. The functions of interest corresponding to the shape of

the unconditional distribution ($\omega_4$, $\omega_5$, $\omega_6$) are all well within the support of the predictive distribution. These models have more difficulty in accounting for the observed leverage, in particular the correlation between $y_t$ and $y_{t+1}^2$.

This elucidation of the prior density $p(\boldsymbol{\omega} \mid A)$ corresponding to a prior distribution $p(\boldsymbol{\theta}_A \mid A)$ is the key in choosing prior distributions; indeed, it is hard to see how else subjective prior distributions can be elucidated.[3] It also indicates whether the model at hand (data density combined with prior density) can account for individual $\omega_i^o$. But note that it says nothing about whether the model is consistent with the *entire* observed vector of interest. To examine this question it is necessary to move to the posterior density (4) and the corresponding density of the vector of interest (6).

# 5 Formal model comparison

This section examines the question, how probable are these models relative to one another? The following section takes up the issue of studying deficiencies in each model without regard to specific alternatives.

Let the models in question be denoted $A, B, ...$ and extend the notation for model $A$, of Section 2, to models $B$ and beyond in the obvious way. The case of just two models, $A$ and $B$, will suffice to convey the essentials. If we assign prior probabilities to each model, $p(A)$ and $p(B)$ with $p(A) + p(B) = 1$, then we have in hand a joint probability distribution of models $A$ and $B$, parameter vectors $\boldsymbol{\theta}_A$ and $\boldsymbol{\theta}_B$, and observables $\mathbf{y}$. Then the posterior odds ratio in favor of model $A$ is

$$\frac{p(A \mid \mathbf{y}^o)}{p(B \mid \mathbf{y}^o)} = \frac{p(A)\, p(\mathbf{y}^o \mid A)\, / p(\mathbf{y}^o)}{p(B)\, p(\mathbf{y}^o \mid B)\, / p(\mathbf{y}^o)} = \frac{p(A)}{p(B)} \cdot \frac{\int_{\Theta_A} p(\boldsymbol{\theta}_A, \mathbf{y}^o \mid A)\, d\boldsymbol{\theta}_A}{\int_{\Theta_B} p(\boldsymbol{\theta}_B, \mathbf{y}^o \mid B)\, d\boldsymbol{\theta}_B} \tag{14}$$

$$= \frac{p(A)}{p(B)} \cdot \frac{\int_{\Theta_A} p(\boldsymbol{\theta}_A \mid A)\, p(\mathbf{y}^o \mid \boldsymbol{\theta}_A, A)\, d\boldsymbol{\theta}_A}{\int_{\Theta_B} p(\boldsymbol{\theta}_B \mid B)\, p(\mathbf{y}^o \mid \boldsymbol{\theta}_B, B)\, d\boldsymbol{\theta}_B}. \tag{15}$$

(The last equality in (14) invokes (3), and (2) is used in (15).) The first ratio in (15) is the prior odds ratio, the second ratio is the Bayes factor, and the numerator and denominator of the Bayes factor are respectively the marginal likelihoods of models $A$ and $B$.

Since $p(A \mid \mathbf{y}^o) + p(B \mid \mathbf{y}^o) = 1$, the posterior odds ratio determines the posterior model probabilities $p(A \mid \mathbf{y}^o)$ and $p(B \mid \mathbf{y}^o)$. Given additional models $C, D, ...$, the posterior odds ratios lead immediately to all the posterior model probabilities. The marginal likelihoods $p(\mathbf{y}^o \mid A)$, $p(\mathbf{y}^o \mid B), ...$ are the key elements in these computations. The required integrations (the numerator and denominator of the Bayes factor in (15)) cannot be performed analytically (textbook examples aside, and certainly for the models used in this article) but they can be approximated using methods that are similar in nature to posterior simulators; see [10] for an introductory discussion.

[3] There is a substantial literature on elicitation of prior distributions that builds on this fact: see for example [7], [17] and [18].

| Table 7<br>Log marginal likelihood values, alternative models | | | | | |
|---|---|---|---|---|---|
| Gaussian | 51,927 | Mix(2,2) | 56,677 | Mix(6,6) | 57,199 |
| GARCH | 56,641 | Mix(3,3) | 57,074 | Mix(7,7) | 57,245 |
| $t$-GARCH | 57,243 | Mix(4,4) | 57,097 | Mix(8,8) | 57,206 |
| | | Mix(5,5) | 57,192 | Mix(6,3) | 57,261 |

Table 7 provides the (natural) logarithm of the marginal likelihood value of eleven alternative models, using our data set of 17,052 returns. From these values and a given set of prior odds ratios, posterior probabilities and odds ratios can be computed. For example, if all prior model probabilities are 1/11, the posterior odds ratio in favor of the compound Markov normal mixture model with 4 permanent states and 4 transitory states within each permanent state ($m_1 = m_2 = 4$), versus the same model with $m_1 = m_2 = 3$, is $exp(57,097 - 57,074) = 9.7 \times 10^9$. Thus the model probability ratios implied in Table 7 are rather large. This stems from the fact that there are so many observations.

The Gaussian model has by far the lowest marginal likelihood, reflecting deficiencies previously noted. As additional states are added to the mixture models they become increasingly flexible, and their marginal likelihoods rise. The compound Markov normal mixture model with $m_1 = m_2 = 2$ provides the least flexibility of those considered; its marginal likelihood is similar to that of the GARCH model. The incorporation of additional states increases the marginal likelihood, with by far the largest increase being from $m_1 = m_2 = 2$ to $m_1 = m_2 = 3$.[4] The mixture model with $m_1 = m_2 = 7$ has a marginal likelihood comparable to $t$-GARCH. There are other models, for which $m_1 \neq m_2$, that have even larger marginal likelihood values, like the one for $m_1 = 6$, $m_2 = 3$ shown in Table 7.

These findings provide a formal comparison of models, but they yield no insight into why (for example) posterior odds are overwhelmingly against the GARCH model, or why they favor (though not as strongly) some of the mixture models over the $t$-GARCH model. There are several useful by-products of the posterior distribution that can provide a wealth of such information. One is the decomposition of the marginal likelihood and Bayes factors observation by observation, as discussed in [12]. Another is to repeat the specification analysis carried out in Section 4, but using the posterior rather than the prior.

[4] Unlike log-likelihood statistics from maximum likelihood estimation, marginal likelihood values do not necessarily increase for nesting models: compare $m_1 = m_2 = 7$ with $m_1 = m_2 = 8$ in Table 7. The intuition is that as the prior distributions $p(\theta_A \mid A)$ is spread over more and more dimensions, $p(\mathbf{y}^o \mid A)$ will be reduced unless there is a compensating (or greater) increase in the average value of the data density. Thus, there is an inherent penalty for large numbers of parameters in the marginal likelihood, much as in Akaike's information criterion (AIC) and Schwarz's Bayesian information criterion (SBIC). In fact, SBIC is a transformation of an approximation to log marginal likelihood.

# 6 Post-predictive specification analysis

Consider the following conceptual experiment. We have observed data $\mathbf{y}^o$, and the corresponding vector of interest $\boldsymbol{\omega}^o$, in an experiment that can be repeated. Then given the complete model $A$ and the data, the predicted distribution of the vector of interest $\boldsymbol{\omega}$ over future independent experiments is

$$p(\boldsymbol{\omega} \mid \mathbf{y}^o, A) = \int_{\Theta_A} p(\boldsymbol{\theta}_A \mid \mathbf{y}^o, A)\, p(\boldsymbol{\omega} \mid \boldsymbol{\theta}_A, A)\, d\boldsymbol{\theta}_A. \tag{16}$$

Note that (16) differs from (6), in that here $p(\boldsymbol{\omega} \mid \mathbf{y}^o, \boldsymbol{\theta}_A, A) = p(\boldsymbol{\omega} \mid \boldsymbol{\theta}_A, A)$, reflecting the independence of the future expereiments and the recorded experiment that produced the data $\mathbf{y}^o$. The observed $\boldsymbol{\omega}^o$, in the context of this distribution, tells us much about the model $A$. In particular, $\boldsymbol{\omega}^o$ may be implausible – $p(\boldsymbol{\omega}^o \mid \mathbf{y}^o, A)$ is quite small, or $\boldsymbol{\omega}_i^o$ lies in the extreme tails of $p(\boldsymbol{\omega}_i \mid \mathbf{y}^o, A)$; the two are usually equivalent.[5] Such an outcome is a surprise, in the sense of Good ([14], [15]): the probability of the observed event occurring again in a great many repetitions of the experiment is quite low.

Of course time series are not repeated experiments. But for a long stationary time series with $T$ observations, a nearly equivalent conceptual experiment is to ask about the next $T$ observations instead of the next experiment, and the $T$ observations after those, and so on.

Carrying out a post-predictive analysis requires little additional effort, given the draws $\widetilde{\boldsymbol{\theta}}_A^{(m)} \sim p(\boldsymbol{\theta}_A \mid \mathbf{y}^o, A)$ from a posterior simulator. We simply repeat the exercise of Section 4, using these $\widetilde{\boldsymbol{\theta}}_A^{(m)}$ in place of the draws from the prior distribution.

The results of this exercise for the GARCH model are shown in Table 8. From the post-predictive quantiles of $\omega_1$, $\omega_2$ and $\omega_3$ we see that the slow rate of decay in the autocorrelation function of $y_t^2$ is inconsistent with the GARCH specification. (Note that the first autocorrelation is predicted to be higher, the twentieth lower, and the ratio of the twentieth to the first is predicted to be much lower.) From the post-predictive quantiles of $\omega_4$ and $\omega_5$ it is evident that the GARCH model implies tails in the unconditional distribution of $y_t$ that are in general too thick. (The distribution of the excess kurtosis lies well above that observed, and the observed quantile ratio is in the bottom 1% quantile of the post-predictive distribution.) Observed skewness and leverage are well within the support of the post-predictive distribution.

[5] The distinction is far from innocuous, however; see [2].

**Table 8**
Post-predictive distribution of vector of interest
GARCH model

| | | Data | Median | (25%, 75%) | (1%, 99%) |
|---|---|---|---|---|---|
| $\omega_1$ | 1st volatility | .218 | .549 | (.468, .620) | (.221, .787) |
| $\omega_2$ | 20th volatility | .083 | -.001 | (-.002, .001) | (-.007, .079) |
| $\omega_3$ | Volatility decay | .382 | -.001 | (-.004, .003) | (-.013, .164) |
| $\omega_4$ | Excess kurtosis | 22.4 | 165 | (88.1, 448) | (31.1, 3997) |
| $\omega_5$ | Quantile ratio | 39.9 | 87.2 | (66.0, 132) | (39.0, 865) |
| $\omega_6$ | Skewness | -.004 | -.001 | (-.014, .015) | (-.178, .231) |
| $\omega_7$ | Leverage ahead | .0312 | .002 | (-.082, .080) | (-.432, .414) |
| $\omega_8$ | Leverage behind | -.0752 | -.002 | (-.098, .102) | (-.466, .418) |
| $\omega_9$ | Stan deviation | .0115 | .0034 | (.0031, .0040) | (.0026, .0159) |

**Table 9**
Post-predictive distribution of vector of interest
$t$-GARCH model

| | | Data | Median | (25%, 75%) | (1%, 99%) |
|---|---|---|---|---|---|
| $\omega_1$ | 1st volatility | .218 | .474 | (.350, .586) | (.101, .777) |
| $\omega_2$ | 20th volatility | .083 | .000 | (-.0004, .003) | (-.002, .132) |
| $\omega_3$ | Volatility decay | .382 | .000 | (-.001, .007) | (-.005, .256) |
| $\omega_4$ | Excess kurtosis | 22.4 | 1312 | (712, 2388) | (184, 7683) |
| $\omega_5$ | Quantile ratio | 39.9 | 1031 | (558, 2124) | (217, 22940) |
| $\omega_6$ | Skewness | -.004 | -.003 | (-.086, .082) | (-.522, .468) |
| $\omega_7$ | Leverage ahead | .0312 | -.001 | (-.155, .146) | (-.534, .510) |
| $\omega_8$ | Leverage behind | -.0752 | .006 | (-.178, .164) | (-.544, .553) |
| $\omega_9$ | Stan deviation | .0115 | .028 | (.018, .054) | (.010, .504) |

The post-predictive distributions of the $\omega_i$ from $t$-GARCH (Table 9) are all substantially more diffuse. For the measures of tail thickness in the unconditional distribution ($\omega_4$ and $\omega_5$) the distributions are also shifted upward – well beyond the observed values $\omega_i^o$. The GARCH model places the observed standard deviation $\omega_9^o$ near the high end of the post-predictive distribution, while the $t$-GARCH model puts it at the low end.

The analysis in Tables 8 and 9 presents a series of specification problems with GARCH and $t$-GARCH. (A full exploration of these problems is beyond the scope of this article.) An important clue is the unsuccessful effort of these models to accommodate the slow decay in volatility correlations. This leads to high values of $\gamma + \delta$: the post-predictive interquartile range for the $t$-GARCH model is (.9948, .9975), and the centered post-predictive 98% intervals is (.9913, .9998). Such models are close to being integrated GARCH (IGARCH) models; in these models the observed value of $y_t$ eventually collapses about its unconditional

mean, but in the intervening period very large values of $|y_t|$ typically arise before they collapse [9]. This characteristic is evident in the post-predictive distribution of the largest absolute return in the series: it is .228 in the data set, whereas for the $t$-GARCH model the median of the post-predictive distribution of the largest absolute return is .929, the interquartile range is (.480, 1.60), and the 98% centered interval is (.298, 87.0). Some resolution of these difficulties is provided by fractionally integrated GARCH (FIGARCH) models [1], whose consideration is beyond the scope of this article.

**Table 10**
Post-predictive distribution of vector of interest
Compound Markov normal mixture model, $m_1 = 6$, $m_2 = 3$

| | | Data | Median | (25%, 75%) | (1%, 99%) |
|---|---|---|---|---|---|
| $\omega_1$ | 1st volatility | .218 | .147 | (.120, .172) | .037, .233) |
| $\omega_2$ | 20th volatility | .083 | .088 | (.068, .108) | (.019, .163) |
| $\omega_3$ | Volatility decay | .382 | .601 | (.506, .711) | (.282, 1.14) |
| $\omega_4$ | Excess kurtosis | 22.4 | 16.8 | (13.9, 21.1) | (8.64, 71.1) |
| $\omega_5$ | Quantile ratio | 39.9 | 31.3 | (28.0, 35.8) | (21.2, 60.1) |
| $\omega_6$ | Skewness | -.004 | -.002 | (-.004, .004) | (-.011, .010) |
| $\omega_7$ | Leverage ahead | .0312 | -.019 | (-.033, -.005) | (-.007, .030) |
| $\omega_8$ | Leverage behind | -.0752 | -.022 | (-.035, -.006) | (-.068, .027) |
| $\omega_9$ | Stan deviation | .0115 | .012 | (.011, .012) | (.009, .144) |

A similar post-predictive specification analysis of the compound Markov normal mixture models appears in Tables 10 and 11. For both models, almost all of the observed functions of interest $\omega_i^o$ lie well within their predictive intervals. The only possible point of difficulty is leverage: the observed combination of leverage ahead and leverage behind is near the limit of what these models can accommodate.

**Table 11**
Post-predictive distribution of vector of interest
Compound Markov normal mixture model, $m_1 = 7$, $m_2 = 7$

| | | Data | Median | (25%, 75%) | (1%, 99%) |
|---|---|---|---|---|---|
| $\omega_1$ | 1st volatility | .218 | .138 | (.103, .166) | (.013, .236) |
| $\omega_2$ | 20th volatility | .083 | .079 | (.057, .101) | (.006, .161) |
| $\omega_3$ | Volatility decay | .382 | .590 | (.478, .708) | (.214, 1.36) |
| $\omega_4$ | Excess kurtosis | 22.4 | 18.6 | (14.8, 25.3) | (9.27, 22.9) |
| $\omega_5$ | Quantile ratio | 39.9 | 33.2 | (29.0, 39.5) | (21.4, 90.0) |
| $\omega_6$ | Skewness | -.373 | -.001 | (-.004, .001) | (-.023, .016) |
| $\omega_7$ | Leverage ahead | .0312 | -.017 | (-.032, -.041) | (-.071, .036) |
| $\omega_8$ | Leverage behind | -.0752 | -.021 | (0.034, -.079) | (-.074, .023) |
| $\omega_9$ | Stan deviation | .0115 | .012 | (.011, .012) | (.009, .015) |

The contrast between the post-predictive analyses of the GARCH and mixture models is striking. The mixture models make *much* more precise statements about the characteristics of observables in the future, and these characteristics (with the possible exception of leverage) are consistent with past observation: there are few, if any, surprises. The GARCH models lead to quite vague predictions, that are nonetheless essentially inconsistent with what has been observed. This strong contrast between the models is rooted in the way they cope with the observed pattern of persistence in volatility. In the GARCH models, this phenomenon can only be accommodated by a near-unit root, which has unpleasant consequences for long-run behavior. The mixture models, on the other hand, can incorporate quite persistent movements in volatility, without linking them to the long run.

# 7 Conclusion

These examples illustrate how Bayesian specification analysis can be used to capture the implications of models for observables. The goal of this analysis is to highlight inconsistencies between models and observed data, thus increasing our understanding of models and sowing the seeds for the development of better models and improved decision making.

# References


[1] Baillie, R.T., T. Bollerslev and H.O. Mikkelsen, 1996, "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics* 74:3-30.

[2] Bayarri, M.J. and J.O. Berger, 1999, "Quantifying Surprise in the Data and Model Verification," in J.M. Bernardo, et al. (eds.) *Bayesian Statistics 6: Proceedings of the Sixth Valencia International Meeting,* 53-82. Oxford: Oxford University Press.

[3] Bollerslev, T., 1986, "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics* 31: 307-327.

[4] Box, G.E.P., 1980, "Sampling and Bayes Inference in Scientific Modeling and Robutness," *Journal of the Royal Statistical Society Series A,* 143: 383-430.

[5] Chib, S. and E. Greenberg, 1996, "Markov Chain Monte Carlo Simulation Methods in Econometrics, *Econometric Theory* 12: 409-431.

[6] Friedman, M. and L.J. Savage, 1948, "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy* 56: 279-304.

[7] Garthwaite, P.H. and J.M. Dickey, 1988, "Quantifying Expert Opinion in Linear-Regression Problems," *Journal of the Royal Statistical Society Series B,* 50: 462-474.

[8] Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin, 1995, *Bayesian Data Analysis.* London: Chapman and Hall.

[9] Geweke, J., 1986, "Comment on 'Modelling the Persistence of Conditional Variances'," *Econometric Reviews* 5: 157-61.

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

[11] Geweke, J., 1999, "Simulation Methods for Model Criticism and Robustness Analysis," in J.M. Bernardo, et al. (eds.), *Bayesian Statistics 6: Proceedings of the Sixth Valencia International Meeting*, 53-82. Oxford: Oxford University Press.

[12] Geweke, J. 2001, "Bayesian Econometrics and Forecasting," *Journal of Econometrics* 100:11-15.

[13] Geweke, J. and G. Amisano, 2001, "Compound Markov Mixture Models with Applications in Finance," University of Iowa working paper.

[14] Good, I.J., 1956, "The Surprise Index for the Multivariate Normal Distribution," *Annals of Mathematical Statistics* 27: 1130-1135.

[15] Good, I.J., 1971, "The Probabilistic Explication of Information, Evidence, Surprise, Causality, Explanation and Utility. Twenty seven Principles of Rationality," in V.P. Godambe and D.A. Sprott (eds.), *Foundations of Statistical Inference*. Toronto: Holt, Rinehart and Winston, 108-141.

[16] Guttman, I., 1967, "The Use of the Concept of a Future Observation in Goodness-of-fit Problem," *Journal of the Royal Statistical Society Series B*, 29: 83-100.

[17] Kadane, J.B., 1994, "An Application of Robust Bayesian Analysis to a Medical Experiment," *Journal of Statistical Planning and Inference* 40: 221-228.

[18] Kadane, J.B. and L.J. Wolfson, 1998, "Experiences in Elicitation," *Journal of the Royal Statistical Society Series D* 47: 3-19.

[19] Lindgren, G., 1978, "Markov Regime Models for Mixed Distributions and Switching Regressions," *Scandinavian Journal of Statistics* 5: 81-91.

[20] Rubin, D.B., 1984, "Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician," *Annals of Statistics* 12: 1151-1172.

[21] Ryden, T., T. Terasvirta and S. Asbrink, 1998, "Stylized Facts of Daily Return Series and the Hidden Markov Model," *Journal of Applied Econometrics* 13: 217-244.

[22] Savage, L.J., 1962, "Bayesian Statistics," in R.E. Machol and P. Gray (eds.), *Recent Developments in Information and Decision Processes*. New York: Macmillan and Co.