Conditional Forecasting Using Relative Entropy

John C. Robertson, Ellis W. Tallman and Charles H. Whiteman*

June 2002

ABSTRACT

The paper describes a relative entropy procedure for imposing moment restrictions on simulated distributions from a variety of models. Starting from an empirical distribution for some quantity of interest, the technique generates a new distribution that satisfies a set of moment restrictions. The new distribution is chosen to be as close as possible to the original in the sense of minimizing the associated Kullback-Leibler Information Criterion, or relative entropy. The technique is illustrated in several examples that indicate how to produce conditional forecasts from time series models with unknown or difficult to calculate conditional distributions; how restrictions from economic theory may be introduced into a forecasting model; and how to adjust the simulations from a dynamic general equilibrium model to introduce minimal consistency with actual data.

JEL Classification: E44, C53

Please address questions to: John C. Robertson, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street, NE, Atlanta, GA 30309-4470, 404-521-8782 (FAX 404-521-8956) email: john.c.robertson@atl.frb.org; or Ellis W. Tallman, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street, NE, Atlanta, GA 30309-4470, 404-521-8915 (FAX 404-521-8956), email: ellis.tallman@atl.frb.org; or Charles H. Whiteman, W380 PBB, University of Iowa, Iowa City, IA 52242-1000: email: whiteman@uiowa.edu.

---

* All errors are our own fault, and the opinions expressed are those of the authors, and do not reflect the opinions of the Federal Reserve Bank of Atlanta or the Federal Reserve System. We thank Will Roberds , Frank Schorfheide, and Tao Zha for helpful discussions. We also received helpful comments from the participants in the Atlanta Fed brown bag lunch series, and from participants in seminars at the Vanderbilt University Economics Department and the University of Georgia Department of Economics.

INTRODUCTION

One of the frustrations of macroeconometric modeling and policy analysis is that empirical models that forecast well are largely nonstructural, yet making the kinds of theoretically coherent conditional forecasts policymakers wish to see requires imposing structure which in turn takes the model far from empirical relevancy. In this paper, we address this frustration by providing a procedure that can be used for producing conditional forecasts in the "middle ground": the procedure can be used to produce conditional forecasts from any model capable of generating a predictive distribution. Have simulation? Will forecast conditionally. Thus it is possible to modify predictions of a time series model by imposing ever more of the restrictions implied by economic theory, or to modify predictions of an equilibrium model by imposing ever more realism from the data.

Our procedure, inspired by Stutzer (1996) and Kitamura and Stutzer (1997), involves a change of measure from an initial predictive distribution to a new one that satisfies modeler-specified moment conditions, but which incorporates the least additional information possible. That is, we minimize the relative entropy between the two distributions, subject to the restriction that the new one satisfies the given moment conditions. Stutzer (1996) used this idea to modify a nonparametric predictive distribution for the price of an asset to satisfy the martingale condition associated with risk-neutral pricing. Kitamura and Stutzer (1997) used the idea to provide an alternative to generalized method of moments estimation in which the moment or orthogonality conditions hold exactly relative to a new measure (but not, of course, in the data); the parameters are chosen to make the new probability distribution as close in the information-theoretic sense as possible to the original.

Conditional forecasting has been applied in a variety of settings. In the VAR literature, Doan, Litterman and Sims (1984) exploit the contemporaneous and inter-temporal variance-covariance matrix structure in a VAR to account for the impact of conditioning a forecast on one or more variables for one or more periods into the future. Waggoner and Zha (1999) extended this analysis to accommodate uncertainty in model parameters in a fully Bayesian setting. Our procedure can be viewed as an alternative to the Wagonner-Zha procedure that incorporates the conditioning information directly into the prior.

There are several non-controversial applications of conditional forecasting that are justified for forecasting conditional upon known values for other variables in a model. For example, when there are data release lags, the forecaster is subject to irregular patterns of data availability, and is forced to stagger the elements of the forecast data set. In circumstances when one data series is known before another contemporaneous data series is released, a forecaster will make predictions for the unknown value conditional on the known data. In these circumstances, the conditioning value would have no uncertainty unless there was revision error. In the case of financial markets data, like interest rates, when the conditioning values are known data, these conditions would require no error bounds for its values. In the case of estimated data such as GDP, the forecaster can utilize the distribution of revision errors as additional information.

In what follows, we first present the general theory of entropic conditional forecasting. We then turn to several examples. The first example involves incorporation of conditioning information associated with the availability of sub-period data as well as implicit financial market forecasts into a forecasting exercise with a vector autoregressive model. We then turn to the incorporation of conditions implied by economic theory. We conclude with an example that takes a step toward incorporating restrictions from real data into calibrations from a simulated general equilibrium model.

I. UPDATING PREDICTIONS USING RELATIVE ENTROPY

I.1 *Relative Entropy and Moment Conditions.* Our problem involves interest in the expectation of a function *h(X)* of the *M*-dimensional random variable *X* representing predictions implied by a model, *J*. To facilitate calculation of *E[h(X)]*, we have

1. A sample of *N* observations (draws) $\{x_i, i=1,\cdots,N\}$ on *X*, together with weights $\{\hat{\pi}_i, i=1,\cdots,N\}$, which ensure that each observation receives weight in the sample dictated by the predictive density. (For an iid sample from the predictive density itself, $\hat{\pi}_i = 1/N$ for all i.)

2. Other information about *X* or functions of *X* not used in the creation of the predictive density.

The expectation *E[h(X])* might represent the predictive mean or standard deviation, while the other information might involve knowledge of values of some elements of *X*, knowledge of some predictive moments of *X*, knowledge of expectations of functions of *X*, etc. The question is how to use this "new" information.

In the absence of the new information, a natural way to estimate the expectation of a function of interest *h(X)* from the predictive density is via the sample moment estimator

$$\sum_{i=1}^{N} h(x_i)\hat{\pi}_i .$$

Suppose the new information takes the form of knowledge that the expectation of another function of *X*, *g(X),* is equal to a known quantity, $\bar{g}$. In general,

$$\sum_{i=1}^{N} g(x_i)\hat{\pi}_i \neq \bar{g} ;$$

that is, expectation under the original probability measure will not satisfy the moment condition associated with the new information. This, of course, is what makes the information "new". Accommodating the new information requires modifying the beliefs embodied in the original probabilities $\{\hat{\pi}_i\}$. Following Stutzer (1996) and Kitamura and Stutzer (1997), we seek to do this by finding a new set of probabilities $\{\pi_i\}$ representing a new predictive density that is as close as possible to the original, in the information-theoretic sense, subject to the restriction that the new moment condition be satisfied. Following the notation of Soofi and Retzer (2002), the Kullback-Leibler Information Criterion (KLIC), or relative entropy of π to $\hat{\pi}$ is

(1) $$K(\pi : \hat{\pi}) = \sum_{i=1}^{N} \pi_i \log\left(\frac{\pi_i}{\hat{\pi}_i}\right) \quad .$$

Thus we seek new weights that minimize K, subject to the moment constraints[1]:

(2) $$\min_{\{\pi_i\}} \sum_{i=1}^{N} \pi_i \log\left(\frac{\pi_i}{\hat{\pi}_i}\right) \text{ s.t. } \pi_i \geq 0,\ \sum_{i=1}^{N} \pi_i = 1,\ \sum_{i=1}^{N} g(x_i)\pi_i = 0 .$$

There is a substantial literature in science and statistics motivating KLIC and demonstrating its successful application (see volume 107—spring 2002) of *The Journal of Econometrics* for examples). Solution of this problem is straightforward using the method of Lagrange (see Csiszar, 1975 for the solution in case of general probability distributions); the solution can be written

(3) $$\pi_i = \frac{\hat{\pi}_i \exp\left(\gamma' g\left(x_i\right)\right)}{\sum_{i=1}^{N} \hat{\pi}_i \exp\left(\gamma' g\left(x_i\right)\right)}$$

[1] The KLIC is a "directed divergence" between two probability distributions. Reversing the roles of π and $\hat{\pi}$ in the objective function would yield a different set of weights. In the estimation context, the formulation we have adopted leads to the "information-theoretic" estimator of Kitamur and Stutzer (1997); the alternative leads to the "empirical likelihood" estimator of Qin and Lawless (1994).

where γ is the vector of Lagrange multipliers associated with the moment constraints. Thus the initial weights $\hat{\pi}$ have been modified or exponentially "tilted" to generate the new weights π in much the same way that the state-price density modifies objective probabilities of payoffs to risk-neutral probabilities in contingent-claims asset pricing. Moreover, using the fact that $\sum_{i=1}^{N} \pi_i = 1$, and $\sum_{i=1}^{N} g(x_i)\pi_i = 0$, the "tilting parameter" γ can be computed as the solution to a minimization problem:

(4) $$\gamma = \underset{\tilde{\gamma}}{\arg\min} \sum_{i=1}^{N} \hat{\pi}_i \exp\left(\tilde{\gamma}'[g(x_i) - \bar{g}]\right).$$

Then, with the weights in hand, one can compute the updated expectation *h(X)* as

$$\sum_{i=1}^{N} h(x_i)\pi_i \,.$$

I.2 *A Gaussian Example.* To illustrate the tilting procedure, consider the problem of finding the KLIC-closest density g(y) to a bivariate normal f(y) = $N(\theta,\Sigma)$ subject to the restriction that $E(y_2) = \mu_2$ and $var(y_2) = \Omega_{22}$. Letting γ_1 denote the Lagrange multiplier associated with the mean restriction and γ_2 the multiplier associated with the variance restriction, the first order conditions lead to

(5) $$g(y) = c \cdot f(y) \cdot \exp\left\{\gamma_1 y_1 + \gamma_2 y_2^2\right\}$$

where *c* is the normalizing constant. The exponential tilt simply adds a linear and a quadratic term to the quadratic form in the Gaussian kernel. Upon completing the square, we find that $g(y) = N(\mu,\Omega)$ where μ_2 and Ω_{22} are as given, and

$$\mu_1 = \theta_1 + (\Sigma_{22})^{-1}\Sigma_{12}(\mu_2 - \theta_2)$$

$$\Omega_{12} = \Sigma_{12}(\Sigma_{22})^{-1}\Omega_{22}$$

$$\Omega_{11} = (\Sigma_{22})^{-1}(\Sigma_{11}\Sigma_{22} - \Sigma_{21}\Sigma_{21}) + \Omega_{22}((\Sigma_{22})^{-1}\Sigma_{12})^2 .$$

Thus the moment conditions lead to the usual formula for the conditional mean. If, in addition, the variance condition is $\Omega_{22} = 0$, we obtain the usual formula for the conditional variance-covariance matrix as well.

The example illustrates the general principle, apparent from (3) that for a random vector $y \sim f(y)$, the random vector y^* with given $E[T(y^*)]$ for some integrable function T has density given by

(6) $$g(y^*) \propto f(y^*)\exp\left\{\gamma' T(y^*)\right\}.$$

This also suggests a convenient way to sample from the density g, a subject we take up next.

I.3 *Relation to Importance Sampling.* Expression (6) suggests how to generate a sample $\left\{y_i^*\right\}$ from the density g using "importance sampling" (Geweke, 1989). If it is possible to sample directly from the density f, generate an iid sample $\left\{y_i\right\}$ for $i = 1, \ldots, n$ from $f(.)$. To each drawing y_i, assign the "weight"

$$w_i^* = w_i / \sum_{j=1}^n w_j ,$$

where

$$w_i = exp\{\ \gamma' y_i\ \}.$$

Then to compute the expectation of a function $h(y^*)$, simply calculate

$$\overline{h}_n = \sum_{j=1}^n w_j^* h(y_i).$$

Note that the *drawings* are those from the $f(.)$ density; the *weights* are adjusted to make these a set of drawings from the $g(.)$ density. Of course, for this procedure to make sense, the support of $f(.)$ and $g(.)$ must be the same.

More generally, Geweke (1989) gives several guides regarding when $f(.)$ is a “good” importance density for $g(.)$. In essence, what is required is that the weights w_i* must be well-behaved. For example, the new weights should not be “too far” from the uniform weights 1/n: that is, the new density $g(.)$ should not be too far from $f(.)$ in the KLIC sense. To monitor this, Geweke suggests keeping track of the fraction of total weight assigned to the drawing y_j* receiving highest weight. It is common experience amongst importance samplers to encounter a situation in which a poorly designed importance density yields largest weight nearly equal to unity. It is of course desirable to keep this largest weight small—much smaller, say than 1 percent. Another monitoring device advocated by Geweke is more sensitive to unequal weighting: this device compares the sum of the squares of the highest m weights relative to the sum of the squares of all of the weights from the importance sample, multiplied by a normalizing factor. In our applications, m = 10 is used.

Still another device for assessing the quality of an importance sampler introduced by Geweke involves the concept of “relative numerical efficiency”. This quantity is the ratio of the variance of h to the asymptotic variance of the (scaled) Monte Carlo estimator of E(h), $n^{1/2}(\overline{h}_n - E(h))$. Were it possible to sample from the tilted density directly, this ratio would be 1.0. The ratio differs substantially from 1.0 when the weight function can take on relatively large values. The relative numerical efficiency measure is the ratio of the number of draws necessary to achieve any given numerical standard error using the posterior density to the number required using the importance density. Small values suggest that the importance density is a poor one.

I.4 *Interpretation of Weight Function as a Prior Distribution.* In our applications, the distribution of interest is generally a predictive distribution. Such a distribution arises as follows. First, a parametric model (likelihood) for a random vector y given parameters θ is specified: $f(y|\theta)$. Similarly, a prior distribution for θ is specified as $p(\theta)$. By Bayes' rule, the posterior distribution for θ is proportional to the product of prior and likelihood,

$$p(\theta \mid y) \propto f(y \mid \theta) p(\theta) .$$

Given the data y and the parameters θ, the distribution of a future value of y, y' is given by $p(y' \mid y, \theta)$. Then the predictive distribution is

$$p(y' \mid y) \propto \int p(y' \mid y, \theta) p(\theta \mid y) d\theta .$$

To sample from the predictive distribution, one typically samples θ_i from the posterior $p(\theta|y)$ and then y_i from $p(y' \mid y, \theta_i)$. That is, we can think of the drawing of y' as being a function of the underlying parameter draw θ_i. Similarly, the re-weighted draw from a tilted density for y' can simply be thought of as a drawing from the predictive density associated with the original likelihood but a tilted *prior* $\exp\{\gamma' T(y'(\theta))\} p(\theta)$. Thus the moment condition used to modify the original predictive density can be thought of as part of the prior itself, a very natural way to incorporate non-sample information into the analysis.

II. EXAMPLES

In this section we present four examples that implement the entropic conditional forecasting scheme. In the first, we employ a simple Bayesian vector autoregression routinely used in forecasting at the Federal Reserve Bank of Atlanta to produce an alternative forecast conditional on information about the future course of the federal funds rate. Since it is possible (though cumbersome) to produce approximate conditional forecasts in such a model (see Waggoner and Zha, 1999), our procedure simply provides a computationally convenient substitute for existing methods. The second example shows how to impose moment conditions from economic theory on the predictions of a time series model. Specifically, we incorporate the consumption-based asset pricing model covariance restriction between the intertemporal marginal rate of substitution and returns into a predictive density computed from a simple two-variable vector autoregression for consumption growth and returns. Finally, our third example completes a progression from *atheoretic* conditional forecasting (example 1) to theoretically coherent empirical modeling by incorporating moment conditions from *data* into a simulation exercise involving a dynamic stochastic general equilibrium model. That is, the data inform the simulation exercise via a tilt of the model simulations that ensures that the simulated distributions share some features of actual data.

II.1 Forecasting the Federal Funds Rate Using Information from the Futures Market. In this example we employ a VAR model that embodies a Normal-Wishart prior of the type describe in Sims and Zha (1998) and as implemented in Robertson and Tallman (1999). The model variables are the federal funds rate, the log of real GDP (distributed monthly), the log of the CPI price index, the log of the price of West Texas Intermediate oil, the unemployment rate, and the log of the M2 monetary aggregate.

The VAR can be written as:

(7) $$y_t = v + B_1 y_{t-1} + \ldots + B_p y_{t-p} + u_t \,,\ t = 1,\ldots,T$$

where y_t denotes an $m\times1$ vector of current dated observations for period t on the m variables in the VAR; the B_i are $m\times m$ coefficient matrices; and b is an $m\times1$ vector of constant terms. The error term is defined as $u_t = A^{-1}\varepsilon_t$ where ε_t is assumed to be a Normal and independently distributed m×1 vector such that $E[\varepsilon_t | y_{t-s}, s > 0] = 0$, and $E[\varepsilon_t \varepsilon_t' | y_{t-s}, s > 0] = I$ for all t; and

A is a non-singular m×m matrix so that the covariance of u_t is $\Sigma = A^{-1} A^{-1'}$.

To compute out-of-sample h-period-ahead forecasts we solve equation (1) forward from the last *estimation period T* and express y_{T+h} as

(8) $$y_{T+h} = C_{h-1} b + J\boldsymbol{D}^h \boldsymbol{Y}_T + \sum_{i=0}^{h-1} \Psi_i \varepsilon_{T+h-i}\ ,\ h = 1,2,\ldots$$

where $C_0 = I$, $C_i = I + \sum_{j=1}^{i} B_j C_{i-j}$, $i = 1,2,\ldots$, with $B_j = 0$ for $j > p$;

$$\boldsymbol{Y}_T = \begin{bmatrix} y_T \\ y_{T-1} \\ \vdots \\ y_{T-p+1} \end{bmatrix} : (mp \times 1)\ ; \qquad \boldsymbol{D} = \begin{bmatrix} B_1 & \ldots & B_{p-1} & B_p \\ I & & 0 & 0 \\ & \ddots & \vdots & \vdots \\ 0 & \ldots & I & 0 \end{bmatrix} : (mp \times mp)\ ;$$

$\Psi_i = J\boldsymbol{D}^i J' A^{-1} : (m \times m)$ is the i-period-ahead matrix of impulse responses; and $J = \begin{bmatrix} I & 0 & \ldots & 0 \end{bmatrix}$ is the $(m \times mp)$ matrix such that $y_T = J\boldsymbol{Y}_T$ (see Lutkepohl, 1991). Given data, $Y = [y_1, \ldots, y_T]'$, and parameters, $\theta = \mathrm{vec}(B, \Sigma)$, where $B = [b, B_1, \ldots, B_p]'$, the first two terms in (8) sum to give the average value of y_{T+h}, that is $E[y_{T+h} | Y, \theta]$. The last term in (8) is the resulting forecast error $v_{T+h} = y_{T+h} - E[y_{T+h} | Y, \theta]$, and has conditional variance equal to $\sum_{i=0}^{h-1} \Psi_i \Psi_i'$.

The usual point forecast of y_{T+h} reported in VAR studies is an estimate of $E[y_{T+h} | Y, \theta]$. But θ is not completely known, even after observing the data and so some uncertainty exists about the point forecast. Moreover, even given θ it is frequently the case that we know more about average behavior of y_{T+h} than $E[y_{T+h}|Y,\theta]$ would imply. In particular, because of data release lags some, but not all elements of y_{T+j}, j=1,…,n may be known. For example, at the end of February 1999 data values for the federal funds rate, commodity prices, and perhaps the money aggregates for February would be known. In contrast, the most recent CPI and unemployment observations would be measurements for January, and the GDP data would contain measurements only as recent as the fourth quarter of 1998. Thus, the VAR could only be fitted using data through December of 1998. This description is especially relevant for our forecasting exercise.

To generate simulated out-of-sample realizations, we first calculated predictive densities for 24 months beginning in January 1992 using data from 1960:2 to 1991:12. Samples from the predictive densities include those for the variables for which we do not yet have data (like output, unemployment and the CPI) at the time the forecasts are constructed.

The entropy-based forecasts use six months of the federal funds rate futures market forecasts as extra information. That is, we calculate the futures market forecasts and use the relative entropy procedure to force the tilted forecast distributions to have the same mean. One can stop at this point, leaving the conditions "soft" in the terminology of Wagonner and Zha (1999). One may also restrict the variation around the mean forecast to be very small, meaning the conditions are "hard"—the traditional conditional forecast. Another possibility would be to restrict the variability of titled funds rate forecast to match the historical sample variance of the futures market forecast errors, thereby imposing the same precision as the futures market.

Notice that the three-month lag in the availability of quarterly GDP data means that, the forecast formed at the end of February, say, are for the 24 months including January, because there is still no real GDP data yet. For March, the forecast follows the same procedure, but there is clearly more "data" that can be used for conditioning the forecasts of January, February and March real GDP.[2] The process is followed for each quarter until we reach 1999:1, the last period for which we have 8 quarters of data to compare to 8 quarters of forecasts. This produces 36 overlapping quarterly forecasts to compare. Due to the overlapping forecast observations, we focus most attention of the relative magnitudes of the forecast error statistics.

Table 1 presents comparisons of the standard forecast accuracy measures from the unconditional forecast, the relative entropy (tilted) conditional forecast (with the mean and variance restricted to match the futures market data), and the hard conditional forecast computed using the methods of Waggoner and Zha (1999). The point forecast accuracy results are about the same for the hard and the relative entropy conditional forecasts.

The federal funds rate forecasts are more accurate when the forecasts condition on federal funds futures market information than when the model forecasts are unconditional. These results are consistent with those in Robertson and Tallman 2001, Evans and Kuttner 1998, and Rudebusch 1998. The relative RMSE of the conditioned federal funds forecasts for the one quarter ahead RMSE is 60 percent lower than the unconditional forecast error, indicating that there is substantial information content in the federal funds futures market forecasts in short-run forecasts of the federal funds rate. The reduction of RMSE persists, with still more than a 20 percent reduction for the funds rate forecast 4 quarters into the future, although the improvement dissipates by the end of the forecast horizon.

Despite notably improved forecast accuracy for the federal funds rate, there is no systematic evidence that the conditional forecasts contribute to a consistent improvement in the forecast accuracy of the other variables. Among the more notable results, the RMSE of the

[2] The tilting procedure could be readily adapted to take the advance and preliminary GDP estimates as mean estimates of "final" GDP, and use the historical variability of the revision errors as variance conditions.

unemployment rate forecasts are 10 percent smaller than those of the unconditional forecast. However, at that same four-quarter horizon, the conditional forecast errors of inflation are 10 percent larger. Also, the RMSE for the unemployment rate at the eight-quarter horizon are worse for the conditional than the unconditional forecasts.

Financial market data is not the only source of potentially useful forecast information. Suppose, for instance, that one was interesting in forecasting from a VAR model of the funds rate, inflation and the output gap, and that some form of dynamic Taylor-rule for the funds rate were available. Then, rather conditioning on the futures market forecast one might condition on the policy rule-based interest rate forecast. However, one shortcoming of the using only forecasts of a single variable as auxiliary information is that the co-movement between variables implied by the external model are not utilized in constructing the tilted forecast distributions. This suggests that richer results could be potentially obtained by looking at the correlation information contained in other forecasting models.

II.2 Forecasting Consumption and Returns by Incorporating Asset Pricing Model Information. Rather than using alternative forecasts as moment restrictions it is also possible to impose theoretically motivated moment restrictions directly on a forecast distribution by using the relative entropy procedure. In the following example, we use the Euler equation from a standard specification of the inter-temporal consumption capital asset pricing model (CCAPM) as a moment restriction in a forecast of consumption and interest rates. Specifically, we impose the prediction that the expected value of the product of the gross real return and the stochastic discount factor is unity; that is,

$$E\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\alpha} r_{it+1}\right]=1;\ \ i=1,\ldots,N,$$

where r_{it} is the gross real return to asset i at time t, c_t is the level of real consumption at time t, α is the constant relative risk aversion parameter [inherited from (1-α) in the utility function], β is the discount factor, and E is the expectation operator.

Empirical applications of this restricted specification of the CCAPM typically fit the data poorly. For our application, the characteristic bad fit of this model helps highlight how a variety of "goodness of fit" indicators can be used to assess the adequacy of the relative entropy procedures. In the present context, these indicators signal inadequacy of the procedure when one imposes restrictions that require radical distortions to the predictive distribution. If the restrictions do not alter greatly the empirical predictive density, then the KLIC value will not be far from zero. We use the statistics proposed by Geweke (1989), as described above, to indicate how closely the 'tilted' distribution mimics the original distribution.

For illustrative purposes, we use the nominal three-month Treasury bill rate as the nominal interest rate measure and the (annualized) percentage change in the CPI (average of monthly CPI levels over the quarter) as the inflation measure. To proxy the real rate of interest, we use the nominal three-month Treasury bill rate less the quarterly inflation rate measured by the CPI. For the real consumption growth rate, we add nominal consumption expenditures for services and nominal consumption expenditures for non-durable goods and then deflate that number by a geometric weighted-average of the relevant implicit deflators. All data series are quarterly measures over the time period 1960Q1 to 1998Q4.

We estimate a simple vector autoregression model for real consumption growth and the real interest rate with four lags and use a diffuse (Jeffrey's) prior to produce an empirical forecast distribution over a 7 quarter horizon. We then impose the CCAPM restriction on the forecast distribution. One can interpret the altered empirical distribution as the approximated predictive distribution conditional on the moment restriction.

Of course, one must choose values for the constant relative risk aversion (CRRA) parameter and the discount factor, β, to impose the Consumption CAPM restriction. In this application, we hold β fixed at .96 and choose α to be 2, implying a degree of risk aversion. We enforce the CCAPM restriction on the last forecast period (period 7) because at this point the forecast distribution is close to the unconditional distribution (less affected by initial conditions on the forecast). Table 2 presents the results. We find that restricting the furthest forecast period results in a noticeable adjustment to the empirical distribution. The indicators of adequacy – the KLIC value, the relative numerical efficiency (RNE), ratio of the sum of squares of the highest 10 weights (denoted as the "weight ratio"), and the Lagrange multiplier values – all suggest that the empirical distribution must be altered radically in order to satisfy the moment condition, and so results taken from the conditional density may not be reliable. Enforcing the restriction on earlier forecasts in addition to the final forecast exacerbates the distortion.

Nonetheless, economic interpretations of the forecasts values for the real interest rate and the growth rate of real consumption are consistent with the consumption CAPM restriction: mean forecasts of the real interest rate are increased and the forecasts of the real consumption growth rate are lowered relative to the respective original forecasts. These results from imposing the restriction on the final forecast period appear in Figure1.

It is notable that imposing the condition for only the final forecast period still has measurable effects on the forecasts of real consumption growth and the real interest rate for several periods prior to the restriction. However, it is only in the final period, and specifically in the consumption growth rate forecast that the conditional forecast falls outside the 2/3 error bounds of the original forecast. In order to satisfy the restriction, the consumption growth rate must fall close to zero (compared to the original mean consumption growth forecast of about 3.2) and the real interest rate mean forecast must rise to about 4.5 percent (versus the original mean interest rate forecast of about 2.8). For the real interest rate, the higher value is maintained over the prior periods of the forecast horizon.

We can use the relative entropy procedure to generate values of KLIC, RNE, and other indicators of fit to indicate the value of the CRRA parameter that causes least distortion of the empirical distribution. We range the risk aversion parameter from –2 (risk loving) to 2 in .125 increments. We produce a set of observations of fit indicators for each value of the risk aversion parameter after we impose the CCAPM moment restriction using the given risk aversion parameter value. In this way, we could use the relative techniques as an alternative estimation procedure. Figure 2 presents KLIC and ω_{10} plotted relative to the value of the CRRA imposed in the moment condition. In this case, we find that the empirical predictive density for the data is altered least when the parameter of risk aversion is -.375, consistent with displaying risk-loving behavoir, clearly in contrast with the value of 1 (a log specification) or of 2 as in most empirical applications of CCAPM. However, it is notable that the finding of risk loving behavior is consistent with the empirical results of Hansen and Singleton (1996). (See Neely, Roy, and Whiteman, 2001 for a demonstration that the risk-loving estimates can be traced to near non-identification of the model due to poor predictability of consumption growth and returns.)

II.3 Forecasting Using an Equilibrium Model by Incorporating Information in Actual Data. In this section we use the relative entropy procedure to impose data-based moment restrictions on the simulations of a dynamic stochastic general equilibrium model. Standard calibration exercises assess the empirical fit of the theoretical model by specifying values for the structural parameters of the model, simulating data from the model, and then comparing the second moment properties of the simulated data with those of actual data. Often, models perform adequately in some dimensions, but fail to fit the data well in other ways. In this example we take as a benchmark the model of Hansen (1985), which was calibrated to fit certain aspects of US data, and then impose additional data-based correlation restrictions on the simulated data. The measures of fit proposed by Geweke then provide a metric for assessing the adequacy of the restrictions. The analysis is also related to that in DeJong, Ingram and Whiteman (1996) in that we allow for prior uncertainty about the structural parameters.

The basic model involves a representative agent whose objective function is to maximize

$$E\left[\sum_{t=0}^{\infty}\beta^t\left(\log C_t - A\cdot N_t\right)\right]$$

subject to

$$C_t + K_{t+1} = Z_t K_t^{\theta} N_t^{1-\theta} + (1-\delta)K_t$$

where C_t is consumption; N_t is labor; K_t is the stock of capital; Z_t is a technology shock displaying the law of motion

$$Z_t = (1-\rho)\log\bar{Z} + \rho Z_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim i.i.d.N(0,\sigma^2);$$

and $\Gamma = \left(A,\delta,\theta,\rho,\bar{Z},\sigma,\beta\right)$ are parameters.

Given the specification of parameter values and the stochastic process for the technology shock, an approximate solution can be obtained by log-linearizing the resource constraint and the implied Euler equations. As a benchmark we consider an economy measured quarterly in which Γ = (0.33,0.04,0.36,0.95,1,0.712,0.99): the values used in Uhlig (1999). To allow for some degree of parameter uncertainty we assume a prior over Γ that specifies that each parameter is independent and normally distributed, with means given by the benchmark values and standard deviations (0.1,0.004,0.05,0.025,0,0,0), respectively. The 95 percent coverage intervals implied by the prior are narrow: For example, the interval for the capital share is [0.26,0.46]. The range for ρ captures the high degree of persistence in the technology shock that is assumed in most of the literature. Draws on $\rho \geq 1$ and $\delta \leq 0$ were discarded from the simulations.[3]

We focus attention on the contemporaneous correlation between output and hours and try to match this correlation with that in comparable US time series data. We simulate sample paths of output and hours for each draw from the prior on Γ and take a total of 10,000 parameter draws. As a preliminary experiment, we match the correlation restriction in the simulated output and hours time series. We proceed by restricting the standard deviations in one period to be equal to those in the simulated data, and require that the output-hours correlation be equal to 0.86 – a value representative of US time series data on quarterly GDP growth and hours. The computed entropy weights are then applied to the simulated distributions of data in all subsequent time-periods.

The correlation (across parameter draws) between the simulated output and hours data is close to 0.70 in each period. As expected, matching the hours and output correlation is problematic. The tilted distributions satisfy the moment restrictions by construction, but the Geweke measures suggest that the fit is relatively poor. The RNE is only 0.53 and the weight-ratio measure is relatively high at 11.3. As can be seen from Table 3, imposing the higher correlation between hours and output has only modest effects on the standard deviations of the other variables and the correlations between output and the other variables. However, it does substantially raise the correlation between interest rates and output (from 0.37 to 0.60). These

[3] No uncertainty is assumed in the real interest rate, productivity trend, or the standard deviation of the productivity shocks. These latter restrictions were imposed simply for convenience and could be easily relaxed.

effects also persist into the standard deviations and contemporaneous correlations in subsequent periods since the model is stationary.

It is also of interest to consider the contribution of parameter uncertainty to the previous results. Repeating the exercise with a prior that is a point mass at the benchmark parameter values reveals that the output-hours correlation in the simulated data is slightly higher at around 0.75, and as a consequence the fit of the tilted distribution is somewhat better. The RNE and the ratio of the weight-ratio are 0.80 and 2.6, respectively. Nonetheless, the effect of the correlation restriction on the contemporaneous second moment properties of the simulated data is similar to the case when there is parameter uncertainty. Again, the only large change is the increase in the output-interest rate correlation (from 0.4 to 0.61).[4]

III. CONCLUSION

This paper has described a relative entropy procedure for imposing moment restrictions on simulated distributions from a variety of models. The technique produces a set of weights that imply a distribution that is as close as possible to the original in the sense of minimizing the associated Kullback-Leibler Information Criterion, or relative entropy. The technique is illustrated by three examples that progress from *atheoretic* conditional forecasting, to imposing restrictions from a theoretical model on a forecast, and finally to incorporating moment conditions from *data* into the simulations from a dynamic stochastic general equilibrium model. The preliminary results from the application of the technique are encouraging, and the potential breadth of application seems to be large.

[4] The next step will be to extend this analysis to incorporate intertemporal correlation restrictions as well.

References

Csiszár, I., (1975). "*I*-Divergence Geometry of Probability Distributions and Minimization Problems," *The Annals of Probability* 3:146-158.

Doan, T., Litterman, R., and C. Sims (1984), "Forecasting and Conditional Projection Using Realistic Prior Distributions," Econometric Reviews 3:1-100.

Evans, C. L. and Kuttner, K. N., (1998), "Can VARs Describe Monetary Policy?" in *Topics in Monetary Policy Modeling*. Basle: Bank of International Settlements, 93-109.

Geweke, J., (1989). "Bayesian Inference in Econometric Models Using Monte Carlo Integration," Econometrica, Vol. 57, No. 6 (November), 1317-1339.

Hansen, G.D., (1985) "Indivisible Labor And The Business Cycle," Journal of Monetary Economics, Volume 16, no 3, pp 309-328.

Hansen, L.P. and K. Singleton, (1983) "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns." Journal of Political Economy, Vol 91, no 2, pp 249-265.

Hanesen, L.P., and K. Singleton (1996), "Efficient Estimation of Linear Asset Pricing Models With Moving Average Errors," Journal of Business and Economic Statistics 14:53-68.

Kitamura, Y., and M. Stutzer (1997), "An Information–Theoretic Alternative to Generalized Method of Moments Estimation," Econometrica 65:861-874.

Neely, C.J., Roy, A., and C.H. Whiteman, (2001), "Risk Aversion Versys Intertemporal Substitution: A Case Study of Identification Failure in the Intertemporal Consumption Capital Asset Pricing Model," Journal of Business and Economic Statistics 19:395-403.

Qin, J., and J. Lawless, (1994). "Empirical Likelihood and General Estimating Equations," Annals of Statistics, Vol. 22, No. 1. (March), pp. 300-325.

Robertson, John C. and Ellis W. Tallman. 1999. "Vector Autoregressions: Forecasting and Reality." Federal Reserve Bank of Atlanta *Economic Review*, First Quarter, 2-??.

Rudebusch, G. D. (1998), "Do Measures of Monetary Policy in a VAR Make Sense?" *International Economic Review*, 39, 907-31.

Sims, Christopher A. and Tao A. Zha. 1998. "Bayesian Methods for Dynamic Multivariate Models." *International Economic Review."* 39, 4: 949–968.

Stutzer, M., (1996). "A Simple Nonparametric Approach to Derivative Security Valuation," Journal of Finance, Vol. 51, December.

Stutzer, M., (1997) "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," (with Yuichi Kitamura), Econometrica, v.65, July.

Uhlig, Harald. (1999), "A Toolkit for analyzing Nonlinear Dynamic Stochastic Models easily," Published in Computational Methods for the Study of Dynamic Economies, Ramon Marimon and Andrew Scott (editors), Oxford University Press (February), pp.30-61.

Waggoner, D.F., and T. Zha, (1999), "Conditional Forecasts in Dynamic Multivariate Models," The Review of Economics and Statistics 81(4):639-651.

Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics, J. Wiley and Sons, Inc., New York.

Zellner, A., (1977). "Maximal Data Information Prior Distributions," in A. Aykac and C. Brumat (editors), New Developments in the Applications of Bayesian Methods, Amsterdam: North-Holland, 211-232.

Table 1: Root Mean Squared Error Relative to Unconditional Forecast (Ratio)

	Federal Funds Rate		CPI Inflation Rate	
Quarters Ahead	Tilted Mean and Variance	Hard conditioned	Tilted Mean and Variance	Hard conditioned
1	0.39	0.39	1.02	1.02
2	0.61	0.61	1.04	1.04
3	0.74	0.73	1.10	1.11
4	0.77	0.77	1.11	1.11
5	0.84	0.83	1.06	1.05
6	0.90	0.89	1.02	1.02
7	0.93	0.93	1.03	1.02
8	0.95	0.95	1.01	1.01

	Unemployment rate		Real GDP Growth Rate	
Quarters Ahead	Tilted Mean and Variance	Hard conditioned	Tilted Mean and Variance	Hard conditioned
1	0.95	0.95	0.97	0.97
2	0.95	0.95	0.97	0.98
3	0.93	0.94	0.98	0.99
4	0.90	0.91	1.05	1.04
5	0.93	0.93	1.02	1.02
6	0.98	0.98	1.03	1.03
7	1.02	1.03	1.05	1.05
8	1.06	1.07	1.05	1.05

Table 2: Indicators of Adequacy for the Importance Sampling

Example: Consumption Capital Asset Pricing Model Restriction Imposed on VAR Forecast

KLIC	Risk Aversion	RNE	ω_{10}	1	2	3	4	5	6
1.12	2.00	0.0128	4380.09	32.79	0.00	0.00	0.00	0.00	0.00
1.40	2.00	0.0032	7239.28	22.08	19.64	0.00	0.00	0.00	0.00
1.66	2.00	0.0033	7250.66	22.12	11.94	15.88	0.00	0.00	0.00
1.85	2.00	0.0024	7803.26	17.66	14.83	13.42	10.51	0.00	0.00
2.73	2.00	0.0027	8206.33	40.53	1.20	14.76	2.52	17.15	0.00
4.06	2.00	0.0005	8714.79	46.76	23.46	7.33	4.48	5.11	18.79

KLIC is the Kullback-Leibler Information Criteria, Risk Aversion is the parameter value used for imposing the CCAPM, RNE is relative numerical efficiency, ω_{10} is the ratio of the sum of the squares of the ten highest weights relative to the sum of the squares of all weights, and the numbers 1 through 6 are the values for the Lagrange multipliers for imposing the first constraint, first and second, etc, respectively.

Table 3: Statistics From the Simulated Economy
Standard Deviations and Correlation with Output

Parameter Uncertainty

	ORIGINAL		"TILTED"	
VARIABLE	ST. DEV.	CORR.	ST. DEV	CORR
CAPITAL	4.8	.82	4.1	.865
Consumption	3.8	.9	3.3	.934
Labor Hours	5.1	.7	*5.1	*.86
Output	2.3	1.0	*2.3	1.0
Interest Rate	.11	.38	.10	.6
Investment	11.2	.89	11.6	.94
Technology	2.66	.993	2.6	.995

Relative numerical efficiency = .53 Weights ratio measure = 11.26

No Parameter Uncertainty

	ORIGINAL		"TILTED"	
VARIABLE	ST. DEV.	CORR.	ST. DEV	CORR
CAPITAL	4.5	.77	3.7	.81
Consumption	3.24	.875	2.85	.904
Labor Hours	4.6	.75	4.6	*.86
Output	2.4	1.0	2.4	1.0
Interest Rate	.11	.4	.10	.61
Investment	10.7	.91	11.1	.95
Technology	2.3	.9993	2.3	.9996

Relative numerical efficiency = .80 Weights ratio measure = 2.6

Statistics derived from simulations of the Hansen (1986) model as implemented by Uhlig (1999). The "tilted" columns utilize the relative entropy procedure to impose the restriction that the correlation between output and labor hours is .86. We also impose the restrictions that the mean and the variance of the two series in the restriction are the same in the "tilted" distribution as in the original simulated data. The top involve parameter uncertainty, whereas the bottom has no parameter uncertainty

Figure 1A: Real Consumption Growth

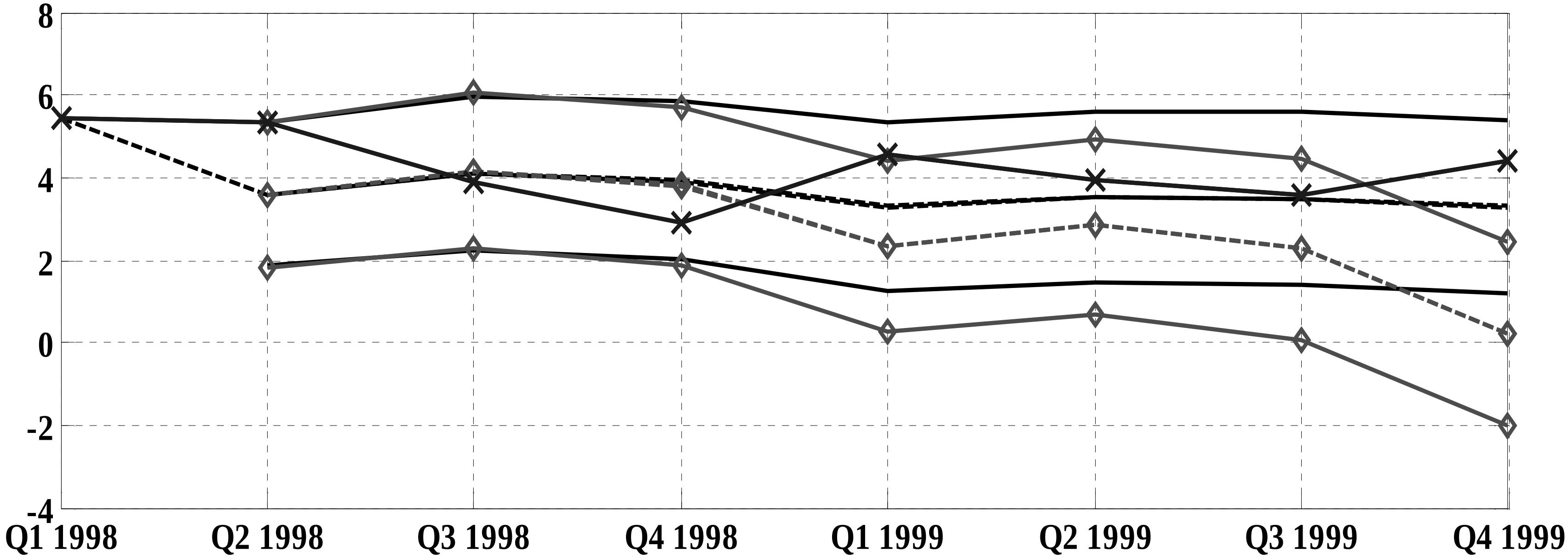

Figure 1B: Real T-Bill

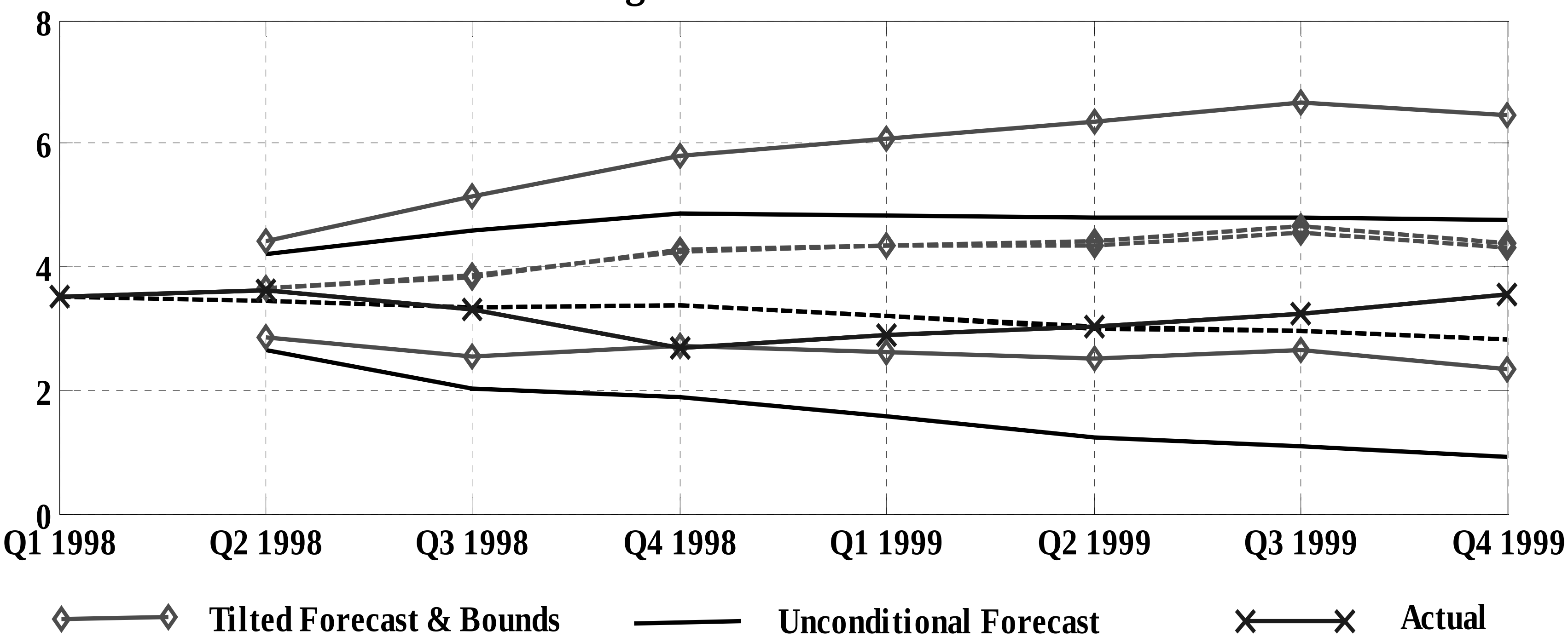

Figure 2: Indicators of Adequacy of Importance Sampling 7th Step