Bayesian Prediction, Entropy, and Option Pricing
In the U.S. Soybean Market, 1993-1997

F. Douglas Foster
Professor of Finance, University of New South Wales

Charles H. Whiteman
Director, Economic Research Institute
And
C. Woody Thompson Professor of Economics
The University of Iowa

October, 2004

Abstract

This paper studies the performance of the Foster-Whiteman (1999) procedure for using a Bayesian predictive distribution for the future price of an asset to compute the price of a European option on that asset. A technical contribution of the paper is the description of a sequential importance sampling procedure for implementing an informative prior that reflects and rewards past option-pricing success. The risk-neutralization of the predictive distribution is accomplished by Stutzer’s (1996) constrained KLIC-minimizing change of measure. The procedure is used in weekly pricing of July and November options on Soybeans on the Chicago Board of Trade from 1993-1997, and produces option prices that mimic market prices much more closely than those of the Black model or those produced by risk-neutralizing a nonparametric predictive.

1 The authors thank Logical Information Machines for their extensive data support. Discussions with Tom Smith and Michael Stutzer substantially improved the paper. Earlier versions have been presented in seminars at APFA 2001, Iowa State University, the Melbourne Business School, The University of Melbourne, The University of Pennsylvania, Purdue University, The University of Queensland, and the University of Technology at Sydney; comments received at these sessions are greatly appreciated.
Bayesian Prediction, Entropy, and Option Pricing
In the U.S. Soybean Market, 1993-1997

I. Introduction

Options pricing techniques have been an important part of finance for some time. Most approaches specify a particular stochastic process to represent the price dynamics of the underlying asset and then derive an explicit pricing model. While this may be acceptable for standard financial assets, it may be problematic for commodities. Many commodities have significant seasonals and require a far more elaborate time-series specification of the price dynamics of the underlying asset. Hence, it becomes difficult at best to derive explicit pricing formulae. Further, with the additional complexity of a rich time-series specification, estimation risk becomes a genuine concern. Finally, not all predictive information need be drawn from the historical time-series, so allowing for an explicit incorporation of non-sample beliefs could be important.

In this paper we suggest an alternative approach. We use numerical Bayes techniques to build a predictive density for the price of the underlying asset (for the example in this paper we need to predict the soybean cash and futures prices at the option’s expiration). Bayesian techniques allow for two very important additions. First, they enable us to integrate out any estimation risk. Second, they allow us to incorporate properly any non-sample information that we may have through an informative prior. Indeed, our informative prior is built sequentially by rewarding past option-pricing successes. Once the predictive density has been computed, we use a procedure proposed by Stutzer (1996) to translate this density to its risk-neutral form. With the risk-neutral density, pricing European options is very straightforward.

To illustrate this technique we consider prices of options on soybean futures traded on The Chicago Board of Trade. We start with a simple vector autoregressive specification for the spot and futures prices. We enrich this predictive model to include weather data as well as futures market
trading activity as evidenced by trading volume and open interest. We compare this procedure with traditional approaches as well as with a non-parametric procedure advocated by Stutzer (1996).

The paper is organized as follows. The next section describes the construction of a predictive distribution for the futures price at the time of expiration of the contract. Section III uses a procedure due to Stutzer (1996) to change the probabilities implicit in the numerical predictive distribution in such a way that they can be thought of as “risk-neutral” probabilities. Using this transformed distribution, options can be priced by computing a partial expectation and discounting at the risk-free rate. Section IV describes a study of the performance of the procedure in week-by-week pricing of soybean options from 1993-1997. Section V concludes.

II. Building the Predictive

The options to be priced are on soybean futures contracts, so the predictive distribution that will be needed is for the soybean futures price on the day the option contract expires, \( F_e \), where \( F \) denotes the futures price and the \( E \) subscript refers to the future’s expiration day. To predict the futures price we use a multivariate model describing the evolution of the (log of the) spot and futures prices, which are related via the cost of carry relation:

\[
\ln(F_t) = \ln(S_t) + b_t,
\]

where \( S \) denotes the spot price and \( b \) is the basis. The basis represents the percentage cost of carrying the spot commodity forward in time to the future’s expiration date (which is not the same as the options expiration date).
The model we utilize is a kind of vector autoregression (VAR) for the spot and futures prices. This structure is convenient for our study, but it will become clear below that in principle we could use any model that can be simulated to produce a predictive distribution. Let \( y_t \) denote the 4 \( \times \) 1 vector containing the futures price, the spot price, and open interest \( (O_t) \) and volume \( (V_t) \) in the futures market:

\[
(2) \quad y_t = (\ln(F_t), \ln(S_t), \ln(O_t), \ln(V_t))^\prime.
\]

The trading volume and open interest variables and equations are included to pick up any volume / volatility relations.\(^2\) Because reported trading volume gives the summed absolute value of trade sizes (it ignores whether the trade is buyer or seller initiated) we also include open interest. Open interest can be thought of as a signed summation of past trading volumes.

In addition to the variables in \( y \), we take as exogenous a set of \( \delta \) variables \( d_t \) which include seasonal dummy variables and variables describing the weather, including rainfall and temperature data at various locations around the soybean-growing region of the Midwestern United States. The model is a near-VAR because it does not describe the evolution of these variables. The near-VAR can be written

\[
(3) \quad y_t = C + D d_t + A(L) y_{t-1} + v_t, \\
\quad v_t \sim iid \ N(0, \Sigma),
\]

where \( C \) and \( D \) are vectors of constants and \( A(L) \) is a vector of \( \lambda \)-degree polynomials in the lag operator. Henceforth, we shall refer to the parameters in \( C, D, \) and \( A(L) \) as \( \mu, \) and write \( \theta = (\mu, \Sigma). \)

\(^2\) See for example, Foster and Viswanathan (1993) for an explicit derivation of a volume / volatility relation in the context of a microstructure model.
Our approach to prediction is Bayesian. Thus we treat the unknown parameters $\theta$ as random, and we condition our analysis on the observed data. We first need to describe how observed data modifies our subjective views about the unknown parameters through the posterior distribution, and then how the posterior is used to build a predictive distribution for future values of the spot and futures prices. Given the simple structure of our model, this is quite standard, and readers familiar with such derivations may wish to skip to the description of the informative prior case.

*Flat prior posterior.* Our description of the posterior distribution under a flat prior follows Foster and Whiteman (1999) closely. Let $Y$ denote the $T \times n$ matrix with $t$-th row given by $y_t'$, and let $X$ denote the $T \times (2 + n \lambda)$ matrix with $t$-th row given by $(1, d_t', y_{t-1}')$. Using the independence of the $v_t$'s and noting that the Jacobian of the transformation from $v$ to $y$ is unity, the sampling density of $Y$ conditional on $\lambda$ initial values, is

$$
(4) \quad P(Y|\theta) \propto \prod_{t=\lambda+1}^T \left| \Sigma \right|^{-\frac{1}{2}} \exp(-\frac{1}{2} v_t' \Sigma^{-1} v_t)
$$

or

$$
(5) \quad P(Y|\theta) \propto \left| \Sigma \right|^{-(T-\lambda-1)/2} \exp\left( \frac{1}{2} \text{tr} \left( YXB \right) (YXB) \Sigma^{-1} \right)
$$

where $\text{tr}$ denotes the trace operator. That is, the VAR can be seen to be a version of the standard multivariate regression model:

$$
(6) \quad Y = XB + V,
$$

where the $(I + \delta + n \lambda) \times n$ matrix $B$ contains the VAR coefficients, and the rows of $V$ are iid $N(0, \Sigma)$. 

- 5 -
To begin, we adopt an “uninformative” prior. There are many interpretations that can be
given to the term “uninformative”; we use the standard “flat” prior:

\[ P(B, \Sigma) \propto |\Sigma|^{(n+1)/2} \]  

(7)

(see Zellner, 1971). The posterior distribution of the parameters is the product of the likelihood and
the prior, or

\[ P(B, \Sigma|Y,X) \propto |\Sigma|^{(T-k+1)/2} \exp\left\{ -\frac{1}{2} \text{tr} \left( (Y-XB)(Y-XB)\Sigma^{-1} \right) \right\}. \]

(8)

Analysis of this expression is simplified by rewriting using the least squares estimate of \( B \), \( \hat{B} \), and the
sum of squares matrix,

\[ S = (Y-X\hat{B})(Y-X\hat{B}), \]

(9)

yielding

\[ P(B, \Sigma|Y,X) \propto |\Sigma|^{(T-k+1)/2} \exp\left\{ -\frac{1}{2} \text{tr} \left( S + (B-\hat{B})'X'(B-\hat{B})\right)\Sigma^{-1} \right\}. \]

(10)

This expression can be factored to reveal that

\[ P(B, \Sigma|Y,X) \propto P(B|\Sigma,Y,X)P(\Sigma|Y,X), \]

(11)

where

\[ P(B|\Sigma,Y,X) \propto |\Sigma|^{k/2} \exp\left\{ -\frac{1}{2} \text{tr} (B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1} \right\}. \]

(12)
is the normal distribution, \( k = 1 + \delta + n \lambda \), and

\[
(13) \quad P(\Sigma|Y,X) \propto |\Sigma|^{\nu/2} \exp\left\{ -\frac{1}{2} \text{tr} S \Sigma^{-1} \right\}
\]

is the “inverse-Wishart” distribution \( (\nu = T \cdot \lambda \cdot k + n + I) \).

**Predictive distribution.** Suppose the prediction horizon is \( h \), so that \( E = T+h \). Given \( \theta \) and data on \( y_t \) and \( x_t \) for \( t = 1, 2, \ldots, T \) denoted by \( Y_T \) and \( X_T \), standard Gaussian regression arguments yield the sampling distribution for \( y_{T+h} \):

\[
(14) \quad P(y_{T+h}|\theta, Y_T, X_T).
\]

Then the joint distribution of the future values and the unknown parameters is

\[
(15) \quad P(y_{T+h}, \theta|Y_T, X_T) = P(y_{T+h}|\theta, Y_T, X_T) P(\theta|Y, X).
\]

The predictive distribution is obtained by integrating out the uncertain parameters,

\[
(16) \quad P(y_{T+h}|Y_T, X_T) = \int P(y_{T+h}|\theta, Y_T, X_T) P(\theta|Y, X) d\theta.
\]

Analytical treatment of this distribution is difficult under the best of circumstances; were an informative prior embedded in the posterior distribution, such analysis is generally not possible. But numerical analysis is straightforward. To generate a random sample from the predictive distribution, one simply generates a random sample from the posterior distribution, \( \{\theta_i\}, i = 1, \ldots, N \), and for each element of this sample, one generates a sample from \( P(y_{T+h}|\theta_i, Y_T, X_T) \). This is straightforward. In particular, to sample from the posterior distribution, simply sample from the appropriate inverse-Wishart, use this drawing of \( \Sigma \), \( \Sigma_i \), to condition the normal, and draw a \( \beta_i \). For each drawing of \( \beta \) and \( \Sigma_i \), a drawing from the predictive is calculated as follows: first make \( h \) drawings from the
$N(0, \Sigma, \iota)$ to generate realizations of shocks $v_{T+1}, v_{T+2}, ..., v_{T+h}$. Then starting from the last $\lambda$ sample data points, perform a dynamic simulation of the VAR using the previously drawn $B_i$ and the newly drawn shocks. The only complication here stems from the fact that the future values of the $d$ variables are not known at date $T$. We make the process fully operational by replacing these future values for the weather variables in $d$ by historical averages using data through 1992.

**Informative prior.** In our application, we also experimented with various informative priors in place of the “flat” prior expression (7). In particular, in place of (7) we used

$$(17) \quad P(B, \Sigma) \propto |\Sigma|^{(n+1)/2} f(B)$$

where $f(B)$ is a density. The density we used embodies the notion that values of $B$ that price options “well” are more likely; this density will be described once the option pricing procedure is introduced in the next section. To generate a sample under the informative prior, we turned our flat prior posterior distribution into an “importance sampler” (see Geweke, 1989). That is, we sample from the flat prior posterior distribution, and assign each such drawing a “weight” of $\frac{f(B_i)}{\sum_{i=1}^{N} f(B_i)}$. Thus, for example, posterior moments under the flat prior posterior would be calculated by $N^{-1} \sum_{i=1}^{N} B_i$, while posterior moments under the informative prior would be calculated by $\sum_{i=1}^{N} \left[ \frac{f(B_i)}{\sum_{j=1}^{N} f(B_j)} \right] B_i$

Under mild conditions (see Geweke, 1989), estimates computed in this way converge almost surely to the population values. (See DeJong and Whiteman (1995) for an application of this procedure.)

### III. The Risk-Neutral Density and Option Prices
Once we have computed the predictive density we need to risk-adjust the probabilities to form the risk-neutral or pricing density. To do this we use a procedure advocated by Stutzer (1996). This procedure uses the maximum entropy principle of information theory to transform the predictive density to its risk-neutral form. This section describes his basic approach.

Using the Monte Carlo sample from the predictive density for $F_E$ we compute a futures return factor, $R^i(E-T)$, for each draw, $i = 1, 2, \ldots, N$:

$$F^i = F^i R^i(E-T) \tag{18}$$

where $T$ denotes the end of the sample (the “current” date) and $E$ is the options expiration date. In the flat-prior case, each of these drawings is assigned an equal weight of $1/N$, whereas in the informative prior case the weights will differ according to (17) and the ensuing discussion. We now need to transform the Monte Carlo probabilities for each draw, $\hat{\pi}(i)$ so that the resulting estimated risk-adjusted density, $\hat{\pi}^*(i)$ prices the futures contract properly. That is, we require the true risk-adjusted density to satisfy the following:

$$\sum_i \pi^*(i) R(E-T) = 1 \tag{19}$$

Equivalently, under the probabilities $\pi^*$, the expected value of the futures price at $E$ is the current futures price. Thus $\{\pi^*(i), i = 1, \ldots, N\}$ is the “equivalent martingale measure” associated with arbitrage-free price system. (See Huang and Litzenberger, 1988, Chapter 8.).

Of course, there are many choices of the $N$-component vector $\pi^*$ satisfying expression (19). Following Stutzer (1996), we use an estimate, $\hat{\pi}^*$, satisfying expression (19) that is chosen to minimize the Kullback-Leibler Information Criterion (KLIC) distance between the risk-adjusted
probabilities and those from the predictive density formed with our numerical Bayes procedure. This optimization is of the form:

\[
(20) \quad \pi^* = \arg \min_{\pi^*} \sum_{i=1}^{N} \pi^*(i) \ln \left( \frac{\pi^*(i)}{\pi(i)} \right), \quad \text{given: } \sum_{i=1}^{N} \pi^*(i) R'(E-T) = 1.
\]

When the prior is flat and the weights \( \pi(i) = N^{-1} \) are equal for all \( N \) Monte Carlo draws, the constrained optimization in expression (20) is identical to a constrained maximization of Shannon entropy, \(-\sum_{i} \pi^*(i) \ln(\pi^*(i))\). In the general case, using the Lagrange multiplier method gives the Gibbs canonical distribution:

\[
(21) \quad \pi^*(i) = \frac{\pi(i) \exp(\gamma R'(E-T))}{\sum_{i} \pi(i) \exp(\gamma R'(E-T))}, \quad \forall i = 1, 2, ..., N,
\]

whose Lagrange multiplier, \( \gamma^* \), is found by solving:

\[
(22) \quad \gamma^* = \arg \min_{\gamma} \sum_{i} \pi(i) \exp(\gamma R'(E-T) - 1).
\]

Finally, to price a European option on a future contract, use the risk-neutral density to compute the discounted expected value at the option’s expiration. For a call option with a strike price of \( X \) we have:

---

3 See Hobson (1971), Maasoumi (1993), and Stutzer (1995) for a more detailed review of the KLIC. For its use in deriving well-known parametric option pricing models, see Gerber and Shiu (1994).
where \( r^{(E-T)} \) is the risk-free discount factor from the current time, \( T \), to the option expiration date \( E \).

IV. Data and Implementation

For our analysis we use the soybean futures contract traded on the Chicago Board of Trade (CBOT) from 1993-1997. Each futures contract entitles the long to receive 5,000 bushels of soybeans. The deliverable grade is No. 2 Yellow at par and there can be delivery substitutions at differentials established by the CBOT (e.g., #1 soybeans at par + $0.06 per bushel). During the sample period, delivery sites were terminal elevators in the Chicago region. (Beginning in 2000, delivery may be made at any point along a 403-mile stretch of the Illinois and Mississippi rivers from Chicago to St. Louis, with location differentials per bushel established by the CBOT.) Prices for the futures are quoted in cents and quarter cents per bushel, with a tick size of $0.0025 ($12.50 per contract). The daily price limit is $0.30 per bushel above or below the prior day’s price, with no price limit in the spot month. The contract year starts trading in September and each year there are futures that expire in September, November, January, March, May, July, and August. During the sample period, the last delivery day was the last business day of the delivery month; for 2000 and beyond, this has been changed to the 2\(^{nd}\) business day following the last trading day, now the business day prior to the 15\(^{th}\) calendar day of the contract month.

Our histories of the soybean futures and cash prices extend back to January 1959. Figure 1 shows a plot of two key contract months, July and November. July is the last contract before the new crop is harvested, while November is the first futures contract where the details of the new crop are known. For these reasons we concentrate our efforts on pricing these two contract
months. Note that there are a number of occasions where the two series diverge, indicating, in part, traders’ beliefs about the quality of the incoming crop. Each series represents a joining of the underlying contract years. In producing this graph we have a simple rollover policy; use the price from the next year when a contract expires. In the predictive we use a dummy variable and adjust explicitly for the change in the contract year.

We also examine the cash price and the simple difference between the cash and futures prices (the basis). This is done in Figure 2. The cash price series follows the same general pattern as the two futures. The bases have a saw-tooth pattern that is due to the rollover and subsequent decay to expiration as the futures and spot converge. Secondly, the November basis is more variable than the July, supportive of the view that the November basis is more driven by the uncertainty in the growth of the new crop.

The soybean options on futures contract also trades on the CBOT. The options are American style and upon exercise the option holder (for a call) receives a futures contract for the appropriate month. From the CBOT we get the volume and open interest for each contract and strike price each day. We also get the closing (settlement) put and call prices for each contract and strike price. Because there are many strike prices offered and most trading occurs in the at-the-money contracts, we concentrate on nine strike prices; the at-the-money strike price, the four nearest in-the-money (call) strike prices, and the four nearest out-of-the-money (call) strike prices. For these options the early exercise value of the option is quite small, so our use of European option pricing technique is a less onerous assumption. The expiration day of the futures option is the last Friday with at least seven business days remaining in the month before the futures contract expires.

In addition to market data we have data on weather in upper Midwest of the United States. Knowing that Illinois, Indiana, and Iowa are all significant soybean growing states we use the

---

4 Logical Information Machines (LIM) provided all data other than option prices through their XMIM system. The authors are very grateful to LIM for providing this significant data support. The Chicago Board of Trade.
weather in Springfield Illinois, Indianapolis Indiana, and Des Moines Iowa to help predict future soybean cash and futures prices.

Figure 3 shows the minimum and maximum daily temperatures averaged across the three centers. As one would expect there is a significant annual seasonality. Note that temperature changes are likely to have a greater impact on soybean prices at certain times of the year (post planting and prior to harvest, for example).

Another fundamental weather factor is precipitation. In Figure 4 we give the maximum number of days since precipitation across the Des Moines, Indianapolis, and Springfield sites. The larger this value, the more drought-like are conditions. This series shows considerable variation and we note that in some years there can be a significant period without rain in at least one of the cities.

The last information source we use is for US Treasury rates. We use the current 3-, and 6-month bill prices to compute the risk-free rate. For options expiring in less than 20 weeks, we discounted using the 3 month rate; for options 20 weeks or more from expiration, we used the six month rate.

Preliminary specification and estimation. In our implementation, we selected Friday data and estimated a weekly VAR using the most recent 1000 observations through November 6, 1992. In addition to a constant and three lags of the log futures, spot, open interest, and volume, we included four weather variables and a set of deterministic variables including eleven monthly dummies and a contract rollover dummy. The weather variables were the average minimum and maximum temperatures over the three sites, and the minimum and maximum number of days since the last measurable precipitation across the sites. Each of the four equations of the near VAR had provided all option price histories.
Informative prior. In addition to the flat-prior procedure described above, we implemented an informative prior procedure that embodied a type of “learning” over the contract year. In our analysis, we valued calls and puts for July and November contracts for 1993-1997. For each contract, we priced the option on 26 consecutive Fridays beginning 27 weeks prior to expiration. In each case, with the passage of each week, we added that week’s data to the data set, and updated the posterior distribution. Updating the posterior distribution associated with the flat prior thus incorporates a very crude sort of learning—the additional data is used as it becomes available, but in precisely the same way (i.e., via a flat prior) as the initial sample. A more natural type of learning would respect the success of the option pricing procedure itself. Indeed, our informative prior was built iteratively: in each contract year, we would begin 27 weeks prior to expiration with the basic near VAR specification and the data, but no other prior information. After pricing options in that first week with the flat prior specification, we would also determine the KLIC-closest predictive distribution satisfying the no-arbitrage condition that also priced three options correctly: the at-the-money, and just-in and just-out of the money options. The reweighting of sample values of the parameters implicit in this calculation was used as a prior distribution for the subsequent week’s calculation.

Implementing this prior requires special care because of its nonconjugate nature. To see why, consider the situation at week t, somewhere in the midst of the 26-week prediction period for

---

5 Some of the details of the specification (the number of lags, the set of deterministic variables used, etc.) were chosen after a preliminary specification search involving a smaller two variable system that omitted volume and open interest data and pseudo-real-time pricing experiments on the 1993 data. This effort was primarily aimed at developing a prior distribution for the model parameters that would improve option pricing. None of the prior distributions so developed survived the transition to the larger four variable system. We believe the only important vestige of the initial specification search is that the logarithms of both the spot and futures data enter the VAR directly. This is because an initial system involving the change in the spot rate together with the basis, a natural implementation of a cointegrated VAR specification, was a poor performer in pricing options. Thus
a particular contract. Suppose that the current sample from the posterior is \{\theta_i\} for \(i = 1, \ldots, N\). Suppose further that the posterior distribution can be represented by a set of probability density values \(\omega_{i,t} = f(\theta, t)\). After constructing the sample from the predictive density and risk-neutralizing (reweighting) via the Stutzer procedure, we also computed a different reweighting that priced the three options correctly. This reweighting can be represented by \(\hat{\omega} \ast (\hat{i})\), which is calculated from \(\omega_{i,t}\) in the same manner as \(\hat{\pi} \ast (\hat{i})\) is calculated from \(\pi(i)\), except that in addition to the no-arbitrage constraint, three additional constraints are imposed to ensure that the three nearest-the-money options are priced correctly in week \(t\). Entering week \(t+1\), the \(\hat{\omega} \ast (\hat{i})\) distribution plays the role of the prior for that week.

The difficulty with this procedure is that the mapping from \(\theta_i\) to \(\hat{\omega} \ast (\hat{i})\) is unknown. Thus it is not possible to sample from the posterior distribution given by the product of the likelihood for week \(t+1\) and the prior just derived the week before. Fortunately, this difficulty can be finessed because it is possible to construct an importance sample for which it is necessary only to know the weights \(\hat{\omega} \ast (\hat{i})\) themselves rather than the mapping from \(\theta_i\) to \(\hat{\omega} \ast (\hat{i})\). This in turn requires that the values of the drawings \{\theta_i\} not change from week to week. To carry this out, we used the flat prior posterior from the first week of our pricing exercise (for each of the 5 years) as an importance sampler. That is, in week 1 of each contract year, we drew a very large number (\(N=50,000\)) of \(\theta_i\). These drawings were then used in each of the subsequent 25 weeks. When used in week \(t\), for example, the drawing \(\theta_i\) was associated with a “weight” not, of course, equal to \(1/N\), but rather equal to the product of the \(\hat{\omega} \ast (\hat{i})\) calculated at the end of week \(t-1\) and an additional weight.

---

\(\hat{\omega} \ast (\hat{i})\) calculated at the end of week \(t-1\) and an additional weight

6不像Foster和Whiteman（1999），在这篇论文中我们不进行第一差分，也不对基础关系施加限制。

6我们使用了50,000个对称产生。平-先验后验的B是正态分布的，所以样本大小是通过将奇数编号的抽样反射到奇数编号的抽样来获得的。这种对称加速的优点是使得蒙特卡罗估计的后验均值精确。蒙特卡罗误差在图中大约是五分之一的便士。计算一周的期权价格大约需要30分钟在800 MHz Pentium IV个人电脑上。所有计算都在RATS中完成。
proportional to the likelihood of $\theta$, using data up to and including week $t$. As the weeks passed, the prior weights and likelihood values changed, and the priors and posteriors updated sequentially.$^{7,8}$

**Alternative predictive.** As one standard for comparison, we also produced a nonparametric predictive like that used by Stutzer (1996). In particular, at each date $t$ we determined the (nonparametric) empirical distribution of $(E-T)$-period growth rates of futures prices using the most recent 1000 weekly observations. This empirical distribution was used with the then-current futures price to produce a predictive distribution of futures prices at expiration which was then risk-neutralized using the procedure outlined in Section III. The resulting distribution was used to price options using a formula analogous to (23).

V. **Pricing Results**

We valued calls and puts for July and November contracts for 1993-1997. As noted above, for each contract, we priced the option on 26 consecutive Fridays beginning 27 weeks prior to expiration. In each case, with the passage of each week, we added that week’s data to the data set, and updated the posterior distribution.$^9$ Thus each Friday we computed the price of the option on the *ensuing* Friday, using the information that would have been available to the actual options traders in real time. Each week we priced every option at every strike price that traded at least once

---

$^7$ Experimentation with other types of more conventional informative (i.e., “shrinkage”) priors in the smaller two-variable system did not produce alternative prior specifications that produced superior option prices.

$^8$ It is important in this sort of analysis that the importance density not be too unlike the posterior of interest. Here, the two are quite similar: they differ only in that the posterior at week 26 of our experiment is based on 1026 observations and the entropic transformation associated with correctly pricing past options, whereas the importance density uses the initial 1000 observations and does not reflect any entropic transformation. By the standards of the importance sampling literature, both effects were small in our implementation. See DeJong, Ingram, and Whiteman (2000) for an example in which a substantially less well-tailored importance density was used successfully.

$^9$ The natural way to do this for the flat prior case is to re-estimate the VAR each period to obtain the moments to be used in the Normal-Wishart posterior. In practice, we actually did this in parallel with the informative prior calculation described in the previous section, using the same 50,000 drawings of the parameters, but weighting by the flat prior alone.
between 1992 and 1998. We used four pricing methods: the Black model (which uses the familiar log normal predictive), the Stutzer model (the risk-neutralized nonparametric predictive), and our flat- and informative-prior numerical Bayes procedures involving the VAR of Section II. In each case, we also computed a “constrained” option price. For the Black model, this involved using the at-the-money call price to determine an implied volatility for use in pricing other options. For the other three procedures, when risk-neutralizing we imposed one additional constraint, that the procedure correctly prices the at-the-money call. (In the constrained cases, in pricing the option at date \( t \) using other information dated \( t-1 \) and earlier, we therefore also used the price of one option at date \( t \) (the at-the-money call.).)

Results are presented for the at-the-money and eight adjacent strike prices in Figure 5. The figures graph mean absolute pricing errors for the calls, in cents per bushel, relative to market prices. We include an overall average for all years (1993-97) as well as year-by-year breakdowns for both the July and November option on futures contracts. Errors are generally smaller for the November contract than the July contract, perhaps reflecting greater uncertainty about the evolution of prices as inventories of the old crop dwindle during the course of the July contract. In terms of matching market prices, among the unconstrained procedures for pricing the July option, the informative-prior Bayesian approach dominates, followed by the flat prior, the Stutzer model, and the Black model. For the November option, the flat-prior Bayesian procedure is dominated by the Black model, the Stutzer model, and the informative-prior Bayesian procedure. Among the latter three, each one is more accurate than the others for at least one strike price shown. Pricing errors tend to be greater when the time value of the option is at its largest (when the option is nearer to the money).

An inspection of the graphs in Figure 5 shows that adding the constraints improves the pricing accuracy considerably. (That is, using contemporaneous market price information in addition yields option prices for the unconstrained strikes that are closer to market prices for options at those strikes.) Both the Stutzer model and the Bayesian approaches price the options at least as well as
the Black model. Given the complexity of the VAR, parameter uncertainty appears to be well managed. The Stutzer model seems to do particularly well when one pricing constraint is added.

Note that the use of the informative prior can improve dramatically the pricing error for the numerical Bayes procedures. This suggests that information in the change of measures from prior weeks that is not part of the sample futures and cash data is helpful in understanding the structure of the option prices. It illustrates the potential importance of using such informative prior calculations, yet the incremental computation cost is relatively minor.

Options traders often calibrate models for pricing using implied volatilities. This gives us another method of computing pricing error. We use the Black model to compute the volatility (standard deviation) of futures price returns implied by the observed market option price and contrast this with the volatilities implied (using the Black model) of each model. This calculation therefore summarizes the same information as in Figure 5, but measured in “volatility units” instead of cents per bushel. The results are reported in Figure 6, where we give the overall averages for the July and November contracts. We do not present the year-by-year breakdowns as the results are consistent with those of Figure 5. Computing errors in this fashion makes more noticeable the improvement of the informative prior for the Bayesian techniques, in the true one-week-ahead (unconstrained) calculations.

Figures 7 and 8 provide mean absolute pricing errors weighted by the volume and option interest from the option market each week, respectively. The intent of these figures is to see whether the largest observed errors occur for options that are actively traded. Overall averages for the both the July and November contracts are presented. Not surprisingly, the at- or near-the-money option each week tends to be the most actively traded, and has the largest open interest. Hence, we see that the weighted errors reflect a concentration of weight and absolute pricing error (as depicted in Figure 5) for at- or near-the-money options. The relative ranking of the various models does not change noticeably, however the improvement of the Bayesian model with an informative prior
becomes more noticeable -- the “success at historical option pricing” prior helps for the most actively traded options.

A final caveat to our implementation is that the options on soybean futures referred to are American-style and we have provided European options prices. Our investigations suggest that for the options that we consider, the additional value of the American early-exercise feature is very modest.

VI. Conclusions

In this paper we studied the real-time performance of a procedure for pricing derivative securities when the underlying asset has rich time-series properties. Many commodities are examples of such assets. We use some simple examples to show that relative to the standard Black (1976) model, as well as a non-parametric procedure advocated by Stutzer (1996), a procedure that makes use of numerical Bayes techniques to develop an underlying predictive density holds significant promise. The results are strongest when we adopt a prior that reflects and rewards historical option pricing success—a prior that can only be implemented numerically using an importance sampling procedure. That these techniques work well for complicated time series models (in our case, the model had 128 parameters) and the use of an informative prior information is particularly encouraging.
VII. References


Figure 1

Soybean Futures Price Histories

Figure 2

Cash Price and Bases
Figure 3

Temperature Data

Figure 4

Precipitation Data
Figure 5: Pricing Errors, 1993-1997

Overall Average for the July Contract, 1993-1997

Mean Absolute Pricing Errors

July 1993 Contract

Mean Absolute Pricing Errors
July 1996 Contract

Mean Absolute Pricing Errors

July 1997 Contract

Mean Absolute Pricing Errors
Overall Average for the November Contract, 1993-1997

Mean Absolute Pricing Errors

November 1993 Contract

Mean Absolute Pricing Errors
November 1996 Contract

Mean Absolute Pricing Errors

November 1997 Contract

Mean Absolute Pricing Errors
Figure 6: Pricing Errors Measured in Implied Standard Deviation, 1993-1997

Overall Average for the July Contract, 1993-1997

Overall Average for the November Contract, 1993-1997
Figure 7: Pricing Errors Weighted by Trading Volume, 1993-1997

Overall Average for the July Contract, 1993-1997

Overall Average for the November Contract, 1993-1997
Figure 8: Pricing Errors Weighted by Open Interest, 1993-1997

Overall Average for the July Contract, 1993-1997

Overall Average for the November Contract, 1993-1997