MODEL ERROR IN CONTINGENT CLAIM MODELS
(DYNAMIC EVALUATION)

by

Eric Jacquier
Robert Jarrow

7-96

RODNEY L. WHITE CENTER FOR FINANCIAL RESEARCH
The Wharton School
University of Pennsylvania
3254 Steinberg Hall-Dietrich Hall
Philadelphia, PA 19104-6367
(215) 898-7616

Model Error in Contingent Claim Models

(Dynamic Evaluation)

Eric Jacquier

Robert Jarrow*

October 1994

Revised, May 1996

*Johnson Graduate School of Management, Cornell University. Jacquier is the corresponding author: (on leave from Cornell), 2323 Steinberg - Dietrich Hall, Wharton School, University of Pennsylvania, Philadelphia, PA 19104. Email: jacquier@jacquier.wharton.upenn.edu. Phone: 215 - 898 4239. We gratefully acknowledge essential support from the C.I.R.A.N.O. (Centre Interuniversitaire de Recherche en Analyse des Organisations), Montreal, and the Kamakura Corporation for Jarrow. This paper has benefitted from discussions with Warren Bailey, David Bates, Peter Carr, Bernard Dumas, René Garcia, Eric Ghysels, Antoine Giannetti, Dilip Madan, Rob McCulloch, Nick Polson, Eric Renault, Peter Rossi, Peter Schotman, and the participants of the seminars at CIRANO, Cornell, and Wharton. We thank Carol Marquardt for letting us use her daily interest rate data.

Model Error in Contingent Claim Models: (Dynamic Evaluation)

Abstract

This paper formally incorporates parameter uncertainty and model error into the estimation of contingent claim models and the formulation of forecasts. This allows inference on functions of interest (option values, bias functions, hedge ratios) consistent with uncertainty in both parameters and models. We show how to recover the exact posterior distributions of the parameters or any function of the parameters. Exact posterior or predictive densities are crucial because a frequent updating setup results in small samples and requires the incorporation of specific prior information. Markov Chain Monte Carlo estimators are developed to solve this estimation problem. Within sample and predictive model specification tests are provided which can be used in dynamic testing (or trading systems) making use of cross-sectional and time series options data. Finally, we discuss several generalizations of the error structure.

These new techniques are applied to equity options using the Black-Scholes model. When model error is taken into account, the Black-Scholes appears very robust, in contrast with previous studies which at best only incorporated parameter uncertainty. We extend the Black-Scholes model by adding polynomial functions of its inputs. This allows for intuitive specification tests. Although these simple extended models improve the in-sample error properties of the Black-Scholes, they do not result in major improvements in out of sample predictions. The differences between these models are important, however, because they produce different hedge ratios and posterior probabilities of mispricing.

Model Error in Contingent Claim Models:

(Dynamic Evaluation)

1 Introduction

Three types of errors occur in the empirical investigation of contingent claim models. The first is the measurement error introduced via the noisy recording of prices (due to non-simultaneity, recording errors, etc...). The second is due to the noise involved in the estimation of parameters, e.g. the volatility. The third is model error which is due to the fact that no model perfectly explains prices. Models are approximate due to simplifying assumptions made about the structure of markets and trading. No existing papers integrate all three errors into the evaluation of contingent claim models. Doing so would provide a methodology which learns from misspecifications while using imperfect models. The purpose of this paper is to propose and implement such a method.

The existing literature, under the null hypothesis of a given model, usually does not incorporate model error in the investigation. Further the sampling error due to parameter uncertainty is often not fully accounted for. [1] To see this, note that typical option pricing studies generate estimates of option prices via direct substitution of point estimates for the underlying volatility. These model estimates are then compared to market prices. As such, these studies lack the ability to produce specification tests for either in-sample fit or out-of-sample prediction.

A different approach for estimating option prices is outlined in Rubinstein (1994). This approach ignores *any* possibility of error as it provides exact fitting. Although this is consistent with the no arbitrage framework at the time of calibration, the model is necessarily imperfect as it ignores market frictions and institutional features too hard to model. The estimation of such an imperfect model with no overidentifying restrictions can be a case of overfitting: the cross-section is fitted nearly perfectly at time t, but the parameter estimates may be useless at time t+1. A realistic requirement for an acceptable model is that it produces similar out- and in-sample performances. This insight may explain the results in Dumas, Fleming, and Whaley (1995) who fit a binomial

[1] See for example Macbeth and Merville (1979), Gultekin, Rogalski and Tinic (1982), Whaley (1982), and Rubinstein (1985). See Bates (1995) and Renault (1995) for discussions of the empirical literature.

model nearly perfectly, only to conclude to a very poor out-of-sample performance.

A few studies concentrate on the error introduced by parameter uncertainty by incorporating only the underlying data into the likelihood. [2] For the Black-Scholes model, Lo (1986) adopts a Maximum Likelihood (ML) setup which assumes that the option price estimator is normally distributed. The prediction error comes from the sampling variability of the ML estimator for the volatility. Lo provides an approximation for the sampling standard deviation of the option price estimator. [3] Lo soundly *rejects* the Black-Scholes model. That is, the confidence intervals generated by the option price prediction error do not cover the true values with the expected frequencies. Although this approach allows for parameter uncertainty and provides a specification test, it still has no model error. More realistically, one would want to admit a model error that: (1) is unrelated to observable model inputs, and (2) has distributional properties preserved out of sample. As in any econometric procedure, predictive intervals would be generated by both model error and parameter uncertainty. We show in our empirical results that when this is done, the Black-scholes appears much more robust than currently believed.

Many reasons are given in the academic literature for the observed departure of market call prices from the Black-Scholes model. Hull and White (1987), Renault and Touzi (1992), Heston (1993), and others attribute it to stochastic volatility. Bossaerts and Hillion (1994b) attribute it to the inability to transact continuously. Platen and Schweizer (1995) develop an equilibrium model with hedgers and speculators where departures from Black-Scholes are caused by the (time varying) demand for hedging out-of or in-the money options. Practitioners' common practice of gamma and vega hedging with the Black-Scholes model reveals an awareness of model error. Most trading systems, however, do not incorporate a formal analysis of this model error. The first contribution of this paper is a method for estimating contingent claim models which includes both model error and parameter uncertainty. This approach provides a representation for the uncertainty surrounding an estimate or prediction of any quantity of interest. This representation is also

[2]The incorporation of option prices in the likelihood which would bring the issue of the imperfect fit of the estimated model, is thus avoided.

[3]Jacquier, Polson, and Rossi (1995) show how to construct efficient option price predictors reflecting parameter uncertainty under stochastic volatility models.

needed for specification tests such as an analysis of residuals and predictive performance, hedging, dynamic model selection, and identification of likely mispricing. Imperfect models often result in time varying parameters which need to be implemented in a setup allowing for periodic reestimation. Although effective sample sizes are then not large, the results of a previous sample can serve as prior information for the current sample. Consequently, our Bayesian approach is designed: (1) to be relevant for small samples and (2) to allow the incorporation of prior information. In contrast, asymptotic methods based upon the normality of the estimator, e.g., methods of moments or maximum likelihood, are ineffective given these requirements.[4]

The estimation procedure should also produce the various posterior and predictive distributions desired (parameters, call prices, hedge ratios). Unfortunately, none of these densities can be written analytically for even the simplest model. We resolve this difficulty by using simulation based Markov Chain Monte Carlo (MCMC) estimators nesting Metropolis and Gibbs algorithms. This allows us to compute any characteristic (moment, quantile, confidence interval) of a distribution to any desired precision by generating sufficient draws. The crucial point is that convergence of the densities obtains in the sequence of draws, rather than in the length of the sample as for standard methods.

In addition, the flexibility of MCMC estimators allows us to generalize the model for heteroskedasticity, inputs observed with error, and for an intermittent mispricing where the error has a larger variance than usual. This last modification introduces an additional state variable for each option price observation, equal to one if there is an additional error, zero otherwise. This is more than a way to allow for fat tails in the model error. The state variable produces a quantity of economic interest, the probability of a quote being an outlier. We provide the exact sample estimator for this probability and other model parameters, using tools from Markov Chain estimation theory. This diagnostic may provide a first step in differentiating between *model error* and *market error*. In the least, learning more about pricing error is crucial because a market error may be the basis for a trading strategy whereas a model error is not. The MCMC framework can also be adapted to formulations where the stock price (or the interest rates) is unobservable.

[4]Bossaerts and Hillion (1994a) use GMM to fit panels of options data. The GMM method admits the existence of model error since the overidentificaton does not let the model fit perfectly. However the model error can not easily be extracted or diagnosed for specification or prediction purposes.

A second contribution of this paper is a method for extending a *basic* model, here the Black-Scholes, with terms derived from (Taylor series) expansions of the input variables. The motivation is as follows. Model error should have the properties desired for a well specified model, i.e., zero mean conditional on the information set (the model inputs). Otherwise, the model implies predictable abnormal returns from simple trading strategies, and/or the resulting price prediction leads to incorrect inference. [5] For example, well known biases for Black-Scholes, smiles and skewness, imply that it does not meet these error requirements. The econometrics of more sophisticated models designed to incorporate these biases, e.g., stochastic volatility, is at a very early stage, and their computation is more complex. In contrast, the *extended* models are as simple to implement as the Black-Scholes. Although illustrated with the Black-Scholes model, the relevance of the extended models is not limited to this case alone. For any implementation of a contingent claims model - the basic model - that is, the status-quo is well-understood, while the more complex models being entertained are significantly harder to learn, implement, and use. The extended model can always serve as a relevant bridge between basic and more complex models.

The final contribution of the paper is the empirical analysis where we implement the preceeding model. First, we show that tests of the Black-Scholes model which only account for parameter uncertainty are flawed. This is because they vastly underestimate the width of the predictive densities and are biased toward rejection. The evidence against the Black-Scholes model once model error is introduced, is much milder. Second, we use an extended model with powers and cross-products of moneyness and time to maturity to nest the Black-Scholes. These variables are justified for inclusion both as potential expansions of a more complex (unknown) model, and as Hutchinson, Lo, and Poggio (1993) argue, these extended models can approximate the Black-Scholes quite well. Here, we use the extended model as an approximation to the unknown model, over and above the Black-Scholes. We show that the pricing and hedging implications ot the extended model differs from Black-Scholes, and improves the in sample specification. However, this is less true out of sample.

The paper is structured as follows. Section 2 introduces the general model and the estimators.

[5]The introduction of model error removes the possibility of perfect arbitrage.

Section 3 details the implementation of the method in the case of both *basic* and *extended* Black-Scholes models. Technical issues are delegated to the appendix. Section 4 provides the empirical application of the method to stock options data. Section 5 discusses the updating implementation and generalizations of the error distribution. Finally, section 6 concludes the paper.

2 General Framework

This section provides the general framework for incorporating the three types of errors - observation, parameter uncertainty, model - into a dynamic estimation procedure.

2.1 Model and Error Structure

Consider a time series of discrete observations of a contingent claim's market price, C_t, for t $\in$ 0,1,2,...,N. For simplicity, we think of C_t as a limited liability derivative, like a call option. We assume that there exists an unobservable equilibrium or arbitrage free price for each observation t, denoted c_t. By definition,

$$C_t \equiv c_t \times e^{\epsilon_t} \tag{1}$$

where ϵ_t is an unobservable disequilibrium or mispricing component of the market price. Examine expression (1), if the market is always in equilibrium, then $\epsilon_t \equiv 0$. Otherwise, mispricings exist. The error structure is multiplicative. One may prefer this multiplicative error specification to the standard additive structure for two reasons. First, it insures the positivity of C_t, without requiring a (complicated) bounded distribution for ϵ_t. [6] Second, the multiplicative specification models have relative rather than absolute errors, insuring that contingent claims with low prices are not ignored in the diagnostic. Given a fixed investment in a strategy, one may argue that relative pricing errors are the relevant criterion. The difference between multiplicative and additive errors is ultimately an empirical question. Consequently, we conduct our empirical tests using both forms. For brevity, we only present the multiplicative error form in the text.

[6] This, however, does not address the issue of alternative lower bounds. For example, a call option must exceed the stock price less the present value of the strike price, S- PV(X).

Let $\mathcal{F}_t : i \in 0, 1, ..., N$ be the information set of the researcher studying the system. We assume that the current market price C_t is included in $\mathcal{F}_t$, but that both the equilibrium price c_t and the market error ϵ_t are not. We also assume that $\mathcal{F}_t$ includes observations of the underlying asset's price, denoted S_t. Let P(.) represent the objective probability measure associated with the system, and let E(.) denote the expectations operator.[7] If the equilibrium price is an unbiased estimator of the observed price, then $E(e^{\epsilon_t} \mid c_t, \mathcal{F}_t) = 1$.

The agents or economists formulate a model for the equilibrium price c_t. The model depends on vectors of observables x_t, and non-observables θ, i.e., $x_t \in \mathcal{F}_t$ and $\theta \notin \mathcal{F}_t$. [8] We now incorporate into our analysis the view that the model is an approximation, even though it was theoretically derived as being exact. This is done via the introduction of an unobservable error, η_t. Formally,

$$c_t = m(x_t, \theta) \times e^{\eta_t}, \tag{2}$$

where m(.) is the model. An unbiased model would have $E(e^{\eta_t} \mid \mathcal{F}_t, \theta) = 1$. This restriction is an implication of economic models with rational agents, operating under the knowledge of expression (2). Because of the necessary simplifying assumptions, typical zero error contingent claim models set $\eta_t \equiv 0$.

The introduction of a non-zero value for η_t can be justified on numerous grounds. First, simplifying assumptions on the structure of trading or the underlying stochastic process made to derive tractable models, often result in errors, i.e. a biased and non i.i.d error structure. For example, Renault and Touzi (1992), Taylor and Xu (1993), Engle, Kane, and Noh (1993), and Heston (1993) show this within the context of stochastic volatility option pricing models. Renault (1995) shows that even a small non-synchroneity error in the underlying price measurement can cause skewed

[7]Formally, there is a probability space $(\Omega, \mathcal{F}, P)$ and a filtration $\mathcal{F}_t : t \in 0, 1, ..., N$. C_t is $\mathcal{F}_t$-measurable, c_t and ϵ_t are not $\mathcal{F}_t$-measurable.

[8]For example, a contingent claim model has the following form. Consider a European call option on the stock, S_t, with exercise price K. Let time T be the maturity date of the option, and let r_t be the risk-free rate appropriate for option C_t. Assume no model or market error at maturity, i.e., $C_T = c_T = [S_T - K]^+$. Then, by a no arbitrage argument as in Harrison and Pliska (1981), there exists an equivalent martingale measure $\tilde{Q}$ with expectations operator $\tilde{E}(\cdot)$, which can depend on a vector θ of unobservable parameters, e.g., volatility, such that

$$m(x_t, \theta) \equiv \tilde{E}\left(C_T e^{\int_t^T r_s ds} \mid \mathcal{F}_t, \theta \right), \quad \text{where } x_t = (s_t, r_t).$$

Black-Scholes implied volatility smiles. Bossaerts and Hillion (1994b) show that the assumption of continuous trading leads to biases in the implementation while Platen and Schweizer (1995) develop a hedging model which results in time varying skewed smiles in the Black-Scholes model. In all of the above cases, the model errors are related to the inputs of the model. Second, the derivation of typical models has the rational agents unaware of either market or model error. Such models could easily be biased in the "larger system" consisting of expression (2). [9]

For many applications (as in Black-Scholes) and models, the error η_t in the *basic* model equation

$$\log c_t(x_t, \theta) = \log b_t(x_{1t}, \theta) + \eta_t, \tag{3}$$

is unlikely to be unbiased and i.i.d. In expression (3), b_t refers to the *basic* model used. In order to document and improve upon the possible misspecification of the basic model, we introduce the *extended* model

$$\log m_t(x_t, \theta, \beta) = \beta_1 \log b_t(x_{1t}, \theta) + \beta_2' x_{2t}. \tag{4}$$

The extended model m_t differs from the basic model b_t by the introduction of a coefficient β_1 and the linear combination $\beta_2' x_{2t}$. The variables x_{2t} may include functions of the observables x_{1t} or other relevant variables and an intercept term. The variables x_{2t} can also include one or several models competing with b_t, in which case the *extended* model equation allows an estimation procedure for nesting competing models. [10] The extended model is intended to capture the biases in $E(\eta_t \mid \mathcal{F}_t)$. It is justified as an approximation of a more general model (see Jarrow and Rudd (1982)) when the more general model is costly to implement. In contrast, the additional cost of the extended model is minimal and its intuition similar to the basic model. [11]

The combination of expressions (1) and (2) yields a general contingent claims valuation model:

$$\log C_t \equiv \log m_t(x_{1t}, \theta) + \eta_t + \epsilon_t,$$

[9]See for example Clément, Gouriéroux, and Montfort (1993). They address the issue of model uncertainty by randomizing the equivalent martingale measure, and show that this induces a non i.i.d. structure of pricing errors related to the model inputs.

[10]Competing models are likely to be highly correlated, causing quasi multicollinearity. Priors in a Bayesian framework resolve this problem. See Schotman (1994).

[11]In a trading system where the cost of changing a model is high, model (4) provides an inexpensive control of model error in hedging a trader's portfolio (book). Standard portfolio theory can be used to minimize the remaining model error risk. At that stage, a proper specification of the error covariance structure is important.

where the observed contingent claims market price is decomposed into three unobservables, the model value, a model error η_t, and a market error ϵ_t. Our goal is to identify the two errors η_t and ϵ_t, and estimate m_t. ϵ_t and η_t can not be separated without further assumptions. In section 6, we propose an error structure to identify outlying quotes which may originate from market error. Until then, we assume only one pricing error exists, η_t.

2.2 Estimation and Prediction with Monte Carlo Methods

This section discusses estimation and prediction using Monte Carlo simulation techniques. For simplicity, we assume that the information set $\mathcal{F}_t$ consists of past and current observations of the contingent claim's market price $C_{t-\tau}$, the underlying asset price $S_{t-\tau}$, and other relevant observables $x_{t-\tau}$, e.g., interest rates, time to maturity. Let $\underline{y}_t$ be this finite-dimensional vector of histories at each date t, then $E(\cdot \mid \mathcal{F}_t) = E(\cdot \mid \underline{y}_t)$.

We want to derive the posterior distributions of the parameters, of any function of the parameters, and of the error for each observation. We also want the predictive distributions of the model error and contingent claim values themselves. These distributions provide the input necessary for better pricing, hedging, and trading. For example they can provide estimates of (1) the probability that observed differences between model and market prices are due to model error, and (2) the probability that a given hedge (delta, gamma, vega) will lose money due to model error. This later quantity is crucial for risk management purpose.

We start the model with a prior distribution $P(\theta)$, where for convenience θ represents all the parameters including σ_η the standard deviation of the error η_t, and the coefficients β from expression (4). Next, the contingent claim model used and the distributional assumptions made for the error η_t, and for the process of the underlying asset yield the likelihood function $P(\underline{y}_t \mid \theta)$. By Bayes' theorem, the posterior is $P(\theta \mid \underline{y}_t) \propto P(\underline{y}_t \mid \theta)\, P(\theta)$. The specifics of the posterior and predictive distributions vary with the prior and likelihood functions and are discussed in the following sections. Here we outline the methodology for generating posterior and predictive densities given a sample of draws of $P(\theta \mid \underline{y}_t)$.

The general approach to estimate the posterior distributions is to (we will show how) simulate from the distribution $\mathrm{P}(\theta \mid \underline{y}_t)$. This yields a sample of random draws of the vector θ. Moments and quantiles are readily obtained from this sample. The main advantage of this Monte Carlo approach is that each draw of θ produces a draw of any deterministic function of θ by direct computation. So, we also have a sample of draws of the exact posterior distribution of the model value, m_t or b_t, or any hedge ratio. Neither θ nor functions of θ have closed form posterior densities. Fortunately, the Monte Carlo approach removes the need to perform numerical integrations.

Another important deterministic function of the parameters is the residual for each observation. Again, each random draw of θ implies a draw of the posterior distribution of η_t for each observation because the residual is simply $\log C_t - \log m(x_t, \theta)$. This residual estimate is the basis for within sample tests of model specification. We discuss these tests in section 4.

For predictive specification tests, one makes predictions and keeps track of their validity. Let η_f be the error of a quote C_f that was not used in the estimation. For example the density of $(\eta_f \mid \theta, \underline{y}_t)$ could be normal with mean 0 and variance $\sigma^2_{\eta_f}$. Next, the predictive density of $(\eta_f \mid \underline{y}_t)$ can be written as:

$$P\left(\eta_f \mid \underline{y}_t\right) = \int P\left(\eta_f \mid \theta, \underline{y}_t\right) P\left(\theta \mid \underline{y}_t\right) d\theta. \tag{5}$$

Again, the integration in equation (5) is readily performed by a Monte-Carlo simulation. The procedure is as follows. First, a sample of draws of $(\eta_f \mid \underline{y}_t)$ is obtained by making one draw of $(\eta_f \mid \theta, \underline{y}_t)$ for each draw of θ. We now have a sample of joint draws of $(\eta_f, \theta \mid \underline{y}_t)$. Second, for each such a draw, compute c_f as in equation(2) with either the basic or extended model. This yields a sample of draws of $P(C_f \mid \underline{y}_t, x_f)$, the predictive density desired. The mean or median of this predictive density yields a point prediction. Quantiles provide a model-error based uncertainty around the point prediction, i.e. a probabilistic method for determining to what extent the difference between a market quote C_f and $E(C_f \mid \underline{y}_t)$ is due to market error. We discuss the specification tests based on this predictive distribution in section 4.

2.3 Markov Chain Algorithms

The previous section outlined the general method for determining posterior and predictive distributions of interest given that draws of the posterior distribution of θ had already been obtained. Here we discuss the intuition underlying two algorithms required to produce these random draws of θ. The technical details of these methods are in section 3 and the appendix.

Let θ be a vector of parameters. We need to make random draws from the posterior density of θ. The crucial requirement for this procedure is an analytic expression for the kernel of this joint density. Given this, the standard approach is to rewrite the joint posterior as a product of conditional densities, from which one can randomly draw. For example, in a standard linear regression, a draw of the slopes β and noise standard deviation σ is generated as follows: (1) draw σ from the inverted gamma, then (2) draw $\beta \mid \sigma$ from the normal. These random draws of β follow the standard Student-t distribution.

The model here is more akin to a non linear regression even in the simplest case. Therefore, it is impossible to use the standard approach to draw from θ. The solution is facilitated by the incorporation of the Gibbs sampling algorithm. [12] This algorithm solves the following problem. Let σ be the standard deviation of the underlying asset's return. Recall that σ_η is the standard deviation of the model error. In our case, we want to but can not draw from the joint density for (σ, σ_η). We can, however, draw from the two conditional densities $(\sigma \mid \sigma_\eta)$ and $(\sigma_\eta \mid \sigma)$. Under mild regularity conditions, draws from the chain $(\sigma_{\eta,0} \mid \sigma_0), (\sigma_1 \mid \sigma_{\eta,0}), (\sigma_{\eta,1} \mid \sigma_1), \ldots, (\sigma_n \mid \sigma_{\eta,n-1}), (\sigma_{\eta,n} \mid \sigma_n)$ converge in distribution to draws of the joint distribution σ and σ_η. This is the Gibbs sampling algorithm. It applies to any number of conditionals. It is easy to check that this procedure is invariant to the initial values.

We would like to use the Gibbs algorithm, but there is one remaining problem. Although we can draw from the distribution of $(\sigma_\eta \mid \sigma)$, we can not draw directly from $p(\sigma \mid \sigma_\eta)$. This is partly because the kernel of p has an analytical expression but the integration constant of p does not. The Metropolis algorithm solves this problem. In fact, with this algorithm, we do not need to know

[12]See Casella and George (1992) for an introduction.

the integration constant of p. [13] First, we select a *blanketing* density q with shape similar to p, from which it is easy to make direct draws. Then, we only need to know the shapes of the two distributions p and q. This knowledge is available from the kernel of p. For each draw made from q, the Metropolis algorithm is a probabilistic rule for acceptance or rejection of that draw. The rule takes into account the shape difference between p and q. This acceptance or rejection rule results in a sample of draws of q which converges in distribution to a sample of draws from p. Even if the shape of q is not "close" to that of p, the algorithm still converges. The closer the shapes of q and p however, the faster the algorithm will converge and generate informative draws on σ. We discuss the Metropolis algorithm in more details in section 3.3

The combination of these two algorithms constitutes the Markov Chain Monte Carlo (MCMC) estimator which we use. As argued, the draws converge in distribution to draws of the posteriors of the parameters, under very mild and verifiable conditions. We now apply the MCMC estimator to the Black-Scholes model.

3 Application to the Black-Scholes Model

3.1 Models and Data

In the Black-Scholes economy, the underlying asset's price S_t follows a lognormal distribution, i.e., $R_t = \log(S_t/S_{t-1}) \sim N(\mu, \sigma)$. The underlying asset is a stock. In our application, the last stock return used predates the first panel of option price data. We can therefore assume a zero correlation between the stock's return and the model error. In fact, we view the stock return data as the basis for the prior distribution on σ. Next, we assume that the stock price and the risk free rate $r_{f,t}$ are observeds without error. These assumptions could be relaxed. We discuss this issue in section 5.2.

Our price data come from the Berkeley options database. Quotes for call options on the stock TOYS'R US, are obtained from December 89 onward. These quotes will be used in the analysis

[13]This is essential for an algorithm to be feasible. It is theoretically possible but practically infeasible to use a standard method such as the inverse CDF method. Neither the CDF nor its inverse have an analytical expression. Each draw of $\sigma \mid \sigma_\eta$ would require an optimization, each step of the optimization requiring a numerical integration.

for the remainder of the paper. TOY is an actively followed stock traded on the NYSE. It does not pay dividends. This allows the use of a European option model for pricing the calls. There are commonly between 80 and 300 quotes on TOY calls daily. We will estimate the models over a day, a week, and months. We collect all quotes whatever their maturity and moneyness because: (1) we want global model diagnostics, and (2) we analyze extended models including bias functions. To build the prior on the volatility σ, we also collect TOY stock returns from the CRSP database for the period leading to November 30^{th}, 89.

The basic model $b(\sigma, x_t)$ is the Black-Scholes. So x_t includes the stock price S, the time to maturity τ, the appropriate interest rate $r_{f,\tau}$, and the exercise price X. The extended model is:

$$\log c_t = \beta_1 b(\sigma, x_{1t}) + \beta_2' x_{2t} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta).$$

θ includes the stock's volatility σ, the model error standard deviation σ_η, and the coefficients β. In this application we restrict the extended model variables to expansions of moneyness and maturity. The moneyness variable z is the logarithm of the ratio $S/Xe^{-r_{f,\tau}}$, of the stock to the present value of the exercise price. Renault and Touzi (1992) show that under a stochastic volatility framework, the Black-Scholes exhibits parabolic biases in z, of intensity decreasing in time to maturity τ. The second variable is τ, the maturity in days (centered around its sample mean). In the empirical analysis, we refer to the following models numbered 0 to 4.

Extended Models Considered

Model	Extended Variables	Number of Parameters
B-S	none	2
0	intercept β_0	3
1	β_0, slope coefficient β_1	4
2	$\beta_0, \beta_1, \tau, z, z^2$	7
3	$\beta_0, \beta_1, \tau, z, z^2, \tau z, \tau z^2$	9
4	$\beta_0, \beta_1, \tau, z, z^2, \tau z, \tau z^2, \tau^2, z^3, z^4$	12

The logic for these models is straightforward. Model 2 allows a linear maturity effect and a moneyness smile. Model 3 lets the smile depend on the maturity. Model 4 introduces higher powers of τ and z. As we do not know the functional form of the better parametric model, it is unclear how extended the expansion of the relevant variables needs to be. Model 4 can be seen as a test of the necessity of further expansions beyond the models 2 and 3. One could also consider other variables. For example, liquidity considerations might suggest the bid-ask spread of both the option and the stock as posssible right hand side variables. We did not include these other variables.

3.2 Priors and Posteriors

This section discusses in more details the estimation of the posterior distribution of θ. The extended models with proper priors nest the basic model with diffuse priors. Indeed, diffuse priors can be obtained by increasing the variance of the proper priors. We use the proper priors

$$\begin{aligned} p(\sigma, \sigma_\eta, \beta) &\propto p(\sigma)p(\sigma_\eta)p(\beta \mid \sigma_\eta) \\ &\sim \mathrm{IG}(\sigma : \nu_0, s_0^2)\ \mathrm{IG}(\sigma_\eta : \nu_1, s_1^2)\ \mathrm{N}(\beta : \beta_0, \sigma_\eta^2 V_0), \end{aligned}$$

where IG is the inverted gamma distribution and N is the normal distribution.

Given σ, the joint prior of β and σ_η is the normal-gamma prior used in regression analysis. Apart from σ, the priors are conjugate, and result in similar posteriors given the relevant likelihood. For σ, we use the inverted gamma consistent with a prior based on the time series of the stock returns. This is not more complicated than the diffuse prior case. The proper priors help model desired restrictions on the parameters. For example, one might expect β_1 to be centered on 1 rather than zero, and to be concentrated in the positive region.[14] One may also want to incorporate in the priors the no arbitrage conditions (the Merton (1973) bounds), by truncating the priors to eliminate parameter values that would violate such bounds. This can be done with simulation estimators by rejecting draws from the posterior that violate these bounds. Finally, when the sample is updated, the previous posterior distribution may be a natural basis for the formulation of the prior distribution.

Appendix A shows the posterior distributions for the extended model with proper priors. We will also allow the error η_t to be heteroskedastic. In the logarithm models, η_t is a relative pricing

[14]Also, an unbiased forecast in the log equation leads to a biased forecast in the level equation. This is due to the term -0.5 σ_η^2 in the mean of the lognormal distribution. One would then expect the intercept to be centered on 0.5 σ_η^2 for an unbiased model. The effect is negligible for commonly encountered parameter values.

error. Going from small to high option prices, one may expect the relative error to decrease. We will allow (up to three) different standard deviations σ_η depending on the moneyness ratio which proxies for the magnitude of the call values. [15]

We can not draw directly from the joint posterior $P(\beta, \sigma, \sigma_\eta \mid \underline{y}_t)$. So, we consider the three conditional distributions $P(\beta \mid \sigma, \sigma_\eta, \underline{y}_t)$, $P(\sigma_\eta \mid \beta, \sigma, \underline{y}_t)$, and $P(\sigma \mid \beta, \sigma_\eta, \underline{y}_t)$. By application of the Gibbs algorithm, a chain of draws from these three distribution will converge in distribution to draws from the desired marginal posterior distributions. $\mathrm{P}(\beta \mid \sigma_\eta, \sigma, \underline{y}_t)$ and $\mathrm{P}(\sigma_\eta \mid \beta, \sigma, \underline{y}_t)$ are shown to be a multivariate normal and an inverted gamma distribution respectively. Direct draws from both are possible. However, direct draws from $\mathrm{P}(\sigma \mid \beta, \sigma_\eta, \underline{y}_t)$ are not possible. Section 3.3 discusses the Metropolis algorithm needed to draw from $(\sigma \mid .)$.

3.3 Posterior Distribution of σ

This section and appendix B discuss the construction of the posterior distribution of σ in detail. The conditional posterior distribution of σ is

$$p(\sigma \mid \beta, \sigma_\eta, \underline{y}_t) \propto \frac{\exp\left\{-\frac{\nu_0 s_0^2}{2\sigma^2}\right\}}{\sigma^{\nu_0}} \times \exp\left\{\frac{\nu s^2(\sigma, \beta)}{2\sigma_\eta^2}\right\},$$

where $\nu s^2(\sigma, \beta) = (Y - X(\sigma))'(Y - X(\sigma))$. Call this distribution p. The main features of the algorithm are as follows. First, we select a (blanketing) distribution q with shape reasonably close to p, from which it is easy to make direct draws. Second, we do not need to compute the normalization constant of p or the cumulative distribution function of σ. Call p$\star$ the kernel of p. For every draw made from q, we know p$\star$/q and can compare it to the same quantity for the previous draw.

This is the basis for a probabilistic rule with three possible outcomes. First, the previous draw is actually repeated and the current draw is discarded. Second, the current draw is chosen. Third, the current draw is rejected and we make another candidate draw from q. The decision is made depending on the value of the ratio p$\star$/q at the candidate draw and at the previous draw.

[15]For brevity, the appendices are based upon the homoskedastic case.

Even if the shape of q is not close to that of p, the algorithm is valid, albeit inefficient with too many rejections or repeated draws. The closer the shapes of q and p are, the faster the algorithm generates informative draws on σ. In the limit, if draws from q were never rejected or repeated, q would in fact be equal to p.

A quantity c approximately equal to $p\star/q$ computed at various values of σ is used to tilt the algorithm towards rejections or repeats, or strike a balance between the two (see appendix B). So, for a given choice of q and c, a plot of the ratio $p\star/cq$ is a gauge of the effectiveness of the algorithm. The more $p\star$ looks like q, the flatter the ratio curve. We choose the blanket distribution q as a truncated normal with mean the mode of $p(\sigma)$.

Figure 1 documents the implementation of the estimator for a model with many parameters and a small sample. This serves to demonstrate the reliability of the method. We estimate model 3 allowing for 2 levels of σ_η: $\sigma_{\eta,1}$ for moneyness ratios below 1, and $\sigma_{\eta,2}$ for moneyness ratios above 1. The sample is made of the 140 quotes from Dec 1, 1989. The prior on σ is flat, based on the last 5 daily returns in November. If desired, one could incorporate more returns information into the prior.

The top left plot of figure 1 shows the ratio $p\star/cq$. The vertical dotted lines mark a +/- 2 standard deviations interval. In that interval, the ratio $p\star/cq$ remains very close to 1. This indicates that the shape of the chosen q is close enough to that of p to guarantee an effective algorithm. In this case, out of 5000 draws of σ, we had 32 rejections and 28 repeats. These numbers are typical of an efficient algorithm.

The benefit of the MCMC algorithm is that the random draws converge in distribution to the desired marginal posterior distribution. The question is then: when have we converged? Figure 1 shows the standard tools used to answer this question. The first tool is a time series plot of the draws, shown for $\sigma_{\eta,2}$ on the top right plot. We intentionally start the chain from unrealistic values for the parameters to check how quickly the draws settle down to a constant regime. Here it took less than 10 draws for the system to settle down. In the empirical section, we conservatively discard the first 500 draws. A further diagnostic is to compute sample quantities for different segments of the remaining sample of draws. When the sample quantities appear similar, the process has converged. The three boxplots shown in Figure 1 confirm that the series has converged. The final check available is a simple autocorrelation function of the draws past the first 50, bottom right plot of Figure 1. The autocorrelations die out quickly confirming that the sequence of remaining draws is stationary. This MCMC algorithm is implemented in our empirical analysis.

4 Empirical Application

This section provides the empirical analysis of the Black-Scholes with model error as applied to equity call options. The estimation is performed on the TOY call options data. The results presented are based on 3500 draws of the posterior distribution of the vector of parameters θ. This distribution is the basis for tests of the Black Scholes versus the extended models.

4.1 Parameters

First, we ask whether different models imply different values for common parameters. This question with respect to the volatility σ is crucial because: (1) an important empirical literature uses implied volatilities to analyze the informational efficiency of options markets, and (2) practitioners routinely back out implied volatilities from the basic Black-Scholes model. Let us first consider σ, the intercept, and the slope coefficients for the B-S model. Figure 2 shows the boxplots of the posterior distributions of these parameters for twelve models. The extremities of the boxplot are the 5^{th} and 95^{th} percentiles. The body of the boxplot shows the median, first and third quartiles of the distribution.

The top plots of figure 2 show that different models imply different volatility parameters. The difference is statistically significant. The logarithm models show median σ's going from 0.25 for the B-S to 0.27 for models 2 to 4. The levels models exhibit the opposite as changes go from 0.24 to 0.22. One may be surprised by the fact that different error specifications have opposed effects on a common parameter. This may come from the fact that the level models concentrate on high prices, i.e., in and at-the money options, while the logarithm models treat all options *equally*. Note that when the B-S is allowed to be in error, then σ does not necessarily estimate the return's standard deviation. It is only a free parameter which enables a functional form to fit the data better. This is especially true here since we use uninformative priors for σ.

Consider now the four bottom plots. As we go from model 0 to model 4, they give us an idea of the average bias around the B-S component in the extended models for a given sample. For the logarithm case, for example, the intercept becomes slightly negative while the coefficient multiplying the Black-Scholes goes down from 1 to below 0.9. This effect is offset by the fact that σ simultaneously goes up. For the level model, the changes in parameters are not as important. The values of the parameters for the logarithm and level models are not directly comparable.

Finally, it is clear from the plots in figure 2 that the parameters of model 4 are estimated with more uncertainty than the other models. This pattern is in fact true for all parameters of model 4, and most functions such as hedge ratios. The estimation for figure 2 was based on 456 quotes for the week of December 4 to 8, 1989. This figure shows that model 4 with 14 parameters (3 levels of σ_η) is difficult to estimate. Additional parameters can be costly and there is a marked difference in the width of the uncertainty when going from model 3 to model 4.

We now turn to σ_η. This parameter is of primary interest since it is the standard deviation of the model error. It is itself a diagnostic of the average error size - the logarithm model error is relative while the level models error is in dollars.

Do the errors of the more complex models have a lower standard deviation? Figure 3a shows the posterior distribution of σ_η for three models. The top left plot is in the case of homoskedastic errors. The other three plots represent σ_η for the same models where different model error standard deviations are allowed for out-of, at, and in-the money options. Two conclusions follow from the inspection of these plots. First, on average models 2 and 3 significantly reduce model error from about 10.5% down to 7%. Second, there is strong evidence of heteroskedasticity. The mean standard deviation of the model 2 error is 12% out-of, 4.5% at , and 2.5% in the money. In fact, the improvements obtained by models 2 and 3 are for out-of and at-the money. Models 2 and 3 actually exhibit higher standard deviations for in-the money options. However the trade off favors models 2 and 3: they bring error standard deviation out-of-the money down from 20% to 12%, at-the money down from 5% to 4.5%, while increasing in-the money from 1.7% to 2.5%.

Figure 3b reproduces these diagnostics for the level models. The top left plot shows σ_η for the homoskedastic models. The improvements due to the extensions (models 2 and 3) are not so impressive as for the logarithm models. The other three plots show that the pattern of heteroskedasticity is opposite from the logarithm case. Direct comparison between the values in figures 3a and 3b is not easy, as the first are in terms of relative errors and the second in terms of dollars. Models 2 and 3 improve the level specification for out-of (8 to 3 cents) and in-the (15 to 13 cents) money errors, but fail to do so for the at-the money errors (about 12 cents).

4.2 Hedge Ratios

We now document the ease of posterior distribution construction for deterministic functions of the parameters, via a study of hedge ratios. Asymptotic estimation for the parameters would require the use of delta methods to obtain an approximate value for the standard deviation of a hedge ratio. The estimate of the hedge ratio would then be assumed to be normally distributed. Instead, we easily obtain random draws of the posterior distribution for the hedge ratios by direct computation for each model.

Contingent claim models are used for two purposes, pricing and hedging. The previous section gave an analysis of pricing errors through inspection of the estimates of the model error's standard deviation. We nex want to investigate whether different models have different policy implications for hedging. Consider the instantaneous hedge ratio Δ. Different model functional forms imply different functional forms for Δ. They are calculated by computing the derivatives of the models specified in equation (4).

A random draw of the parameters yields a draw of the hedge ratio by direct calculation. Figure 4a shows the posterior distribution of Δ for B-S and the logarithm models 2 and 3. Figure 4b computes the hedge ratios for the corresponding level models. Consider figure 4a. The six plots show Δ for out-of, at, and in-the money, short and long maturity options. Models 2 and 3 have similar hedging implications, different from the B-S. The difference is statistically significant even though the larger number of parameters yield wider distributions for Δ. For example, 5 day, 4% out of the money calls have a median Δ of 15.4% per models 2 and 3, and only 13% per the Black-Scholes. The magnitudes of differences are relevant though not very large. They are reliably estimated. Figure 4b shows the same result for the levels model. It is interesting to compare the logarithm B-S in figure 4a with its level counterpart in figure 4b. Their Δ's are not identical. This shows that the distributional assumption of the error alone can affect important model charateristics such as hedge ratios.

Figure 4 and 4b show that the extended models have different policy implications with respect to the design of hedged portfolios relative to the basic Black-Scholes. Note that as in Merton (1973), nothing forces a ratio to be between 0 and 1. But if one believes that the model is a convex function of the stock, then the priors can be formulated to constrain the hedge ratio. In this case one can reject draws which imply ratios outside the range (0,1). We did not impose this restriction.

4.3 Biases

The extended models incorporate additional functions of moneyness and maturity with possibly a fair number of additional parameters. Rather than inspecting the values for each of these parameters, we now ask whether they produce, as a group, pricing implications markedly different from the Black-Scholes. For example, in the logarithm model 2, the call price is multiplied by $\exp\{\beta_2 z + \beta_3 z^2 + \beta_4(\tau - \bar{\tau})\}$. For various values of τ and z, we want to see how close to a flat surface this function is. Again a random draw of this function is obtained by direct computation from a draw of the parameters.

Figure 5 documents the posterior distribution of these functions for the logarithm model estimated from the 419 quotes of December 11 to 15, 1989. We have allowed for heteroskedastic pricing errors. The top left plot shows the mean and the 5^{th} and 95^{th} quantiles of the bias function for model 3 with $\tau = 60$ days. There is very strong evidence of moneyness biases at medium maturities, both statistical and economic. That is, the biases are precisely estimated as the 5% and 95% bands show, and the magnitudes of the biases are large. The top right plot shows the bias function for models 2 to 4 at $\tau = 5$ days. The three models give similar evidence of bias. The bias does not appear different from 60 days shown on the left. The bottom left plot shows the bias functions for model 3 at different maturities. There does not appear to be evidence that the bias function is different at different maturities. Finally we plot the bias function versus time to maturity for models 3 and 4 at three different levels of moneyness. First, models 3 and 4 produce similar biases. Second, this confirms that time to maturity affects biases in only a minor way.

Figure 6 show three-dimensional plots of the bias surfaces for the levels model 3 estimated for two succesive weeks of December 1989. The top plot is for the week of December 4 to 8, with 456 quotes. The bottom plot is for the week of December 11 to 15, with 419 quotes. Figure 6 shows that the mispricing implied by model 3 changes slowly with time. This is due to the time variation in the parameters. The biases, here in dollars, can be large especially for out-of the money, short maturity calls

4.4 Residual Analysis and In-Sample Specification Tests

Models based upon a set of stochastic assumptions can be tested by residual analysis. Standard residual analysis computes a statistic, e.g., autocorrelation, and a hypothesis test is performed based upon an asymptotic sampling distribution for this statistic. In contrast, Bayesian residual analysis uses the exact posterior distribution of the statistic which is computed as dicussed in section 3. Given a random draw of the parameters, we compute for each observation the residual $\log C_t - \log m_t(x_t, \theta)$. This is the basis for residual analysis, see Chaloner and Brant (1988). The most common diagnostics follow immediately. For example, for a given observation, we can compute mean and quantiles, or plot the histogram of the posterior distribution of the residual, conducting a standard outlier analysis. In our case, an outlier has the interpretation of a possible market error. Then, the diagnostic has a policy implication - it may be followed by a trade designed to take advantage of the perceived mispricing. This is dangerous if the model used is misspecified. A quote which appears as an outlier for model 0 may not be for model 2. In this case, unlike the user of the basic model, the user of model 2 would not undertake any trade.

Tests of the relationship between the error structure and right hand side variables are easily conducted with the residuals. The posterior distribution of the correlation of the residuals with an observable input to the model is available. For each draw of the residual vector, compute its correlation with the observable. This produces a draw of the posterior distribution of this correlation. This can be done for any (function of) observables which one suspects is not properly accounted for in the model. If the option price data covers a span of calendar times, then, violations from the Black-Scholes mode, e.g. stochastic volatility, may induce autocorrelation in the pricing errors of the basic model. [16] The posterior distribution of the autocorrelation function of the residuals can be constructed in the same manner as described above for the correlation with a right-hand-side variable. Also of interest are the posterior distributions of the average residuals or average squared residuals for different subsamples, for example, long and short maturity, or in- and out-of-the money. Before looking at our empirical application, we discuss the out-of-sample tests.

4.5 Dynamic Specification Test

The previous discussion pertained to an analysis of quotes or trades belonging to the sample used to estimate the model. We now turn to predictions made for options not part of the sample. In the absence of a misspecification, the predictive density of a yet unobserved value is the best representation of the uncertainty about that value. The predictive density can be compared to the realization when it occurs. This suggests a dynamic specification test and a model comparison procedure, leading to the selection of the model with the best predictive track record. Period by period, one can track the predictive performance of a model, across different types of options, or separated by category, e.g., moneyness or maturity.

Say that we have available at time t, for $\tau = 0, ..., L-1$, L previously formulated predictive densities $P(C_{t-\tau} \mid \underline{y}_{t-\tau-1})$ and L subsequently observed market realizations $C_{t-\tau}$. These are the basic elements for dynamic model comparisons. A predictive density can be formulated for each model. The mean of this predictive densities $\mathrm{E}\Big(C_{t-\tau} \mid \underline{y}_{t-\tau-1}\Big)$ can then be computed. Once $\mathrm{C}_{t-\tau}$

[16] In such as situation, one can introduce in the extended model an observable related to volatility such as trading volume or a time series volatility forecast.

is known, the error are recorded. For each model there are L realizations of this error. The different models can then be compared on the basis of (an average of) their errors and mean squared errors. Time series plots of these error functions can be constructed and updated dynamically for each model. These bias and squared errors can also be computed for subsets of the predictions, e.g., moneyness, maturity.

The above tests concentrate on the mean as a point estimate. Although the prediction mean determines the direction of a strategy, the predictive density also quantifies the uncertainty surrounding the prediction mean. This aspect of the prediction process is hardly ever implemented because standard methods do not provide reliable predictive densities. Models should be compared on their ability to forecast the uncertainty as well as the mean of future realizations. For example, we can compute the interquartile range of a predictive density. Ideally, this interquartile range should cover the realization 50% of the time.

The above tests can also be directed to alternative *economic* criteria. For example, if the model is used to compute a hedge ratio every period, the statistical specification of C_t may not be the best benchmark for model comparison. Instead, ex-ante predictive densities can be computed for the dollar error in delta hedging the call. If the probability of a loss larger than a given value is obtained, then gamma hedging or vega hedging can be considered, with predictive densities as well. If the probability of a loss is still too large, the trade can be avoided. These predictive densities, computed for a trader's entire book, can be used to help manage model error risk. Given the model error covariance structure, standard portfolio theory can be applied to minimize this component of risk. The loss function of interest to the hedger may be the realized squared error

$$SE_t = \left([C_t - C_{t-1}] - E\left(\frac{\partial c_{t-1}}{\partial S_{t-1}} \mid \underline{y}_{t-1} \right) \; [S_t - S_{t-1}] \right)^2 ,$$

of a hedge portfolio. The mean $E\left(\frac{\partial c_{t-1}}{\partial S_{t-1}} \mid \underline{y}_{t-1}\right)$ is that of the posterior density of the hedge ratio. The tracking of $\mathrm{SE}_{t-\tau}, \tau = 0, .. ,$ L-1 can serve to rank competing models. [17]

[17] Partial derivatives may not produce a variance minimizing criterion, specially if the investment horizon is discrete. The quantity δ_t which minimizes a portfolio's expected variance over a fixed investment horizon, t to t+1 may be more appropriate and can be computed by simulation.

4.6 In and Out of Sample Tests

We now implement some of the tests previously discussed. We concentrate on the logarithm heteroskedastic and homoskedastic models as well as the level heteroskedastic models. Table 1 contains the within-sample tests conducted over the first two weeks of December 1989. The parameters are reestimated each week, and then the errors are aggregated. This represents a sample of 871 quotes. The first panel, labeled residual analysis summarizes the residuals' biases and root mean squared errors (RMSE). These numbers are also computed for the out-of and in-the money subsamples, as well as the subsample of quotes for which the mean prediction falls outside the bid ask spread. There is an economic rationale for this last subsample. Predictions and fitted values that fall inside the bid ask spread may not be the basis for a trading strategy, and therefore not as relevant as those that fall outside.

A quick glance at the first panel of table 1 shows that average biases are far smaller than the RMSE's. The extended models appear to improve the bias of the basic B-S models with the exception of the level model for outside B-A spreads. The logarithm extended models have smaller RMSE than the B-S. Again, although there is evidence that the level extended models have lower RMSE than the B-S, it is not the case for quotes outside the B-A spread.

We now reformulate the diagnostic of the logarithm models in terms of *dollar* pricing errors. The residuals in the first panel are *relative* pricing errors and are not directly comparable to the residuals for the level models. The pricing errors for the logarithm models are shown in the second panel of table 1. Note the large pricing errors of model 4 which is not reliably estimated. The RMSE's reveal three facts. First, the incorporation of heteroskedasticity drastically improves pricing precision, even though it did not have an effect on the fit of the model. This is because what is only a second moment effect (variance of the residual) in the logarithms, becomes incorporated in the mean when we exponentiate to compute the price. Second, the extended models significantly improve upon the B-S. Third, the level models seem to have marginally better RMSE's, and less bias than the logarithm models. This does not mean that the level models should be preferred.

Given a fixed amount to invest, relative error may be the more relevant criterion.[18]

The third panel of table 1, labeled distribution analysis, documents the specification of the predictive density. The in-sample predictive density is obtained as discussed in section 2. The first column shows the percentage of quotes for which the mean of the predictive density falls outside the B-A spread. The extended models generate a sizeable reduction in this number. It is below 20% for models 2. The third column is the specification test which computes the percentage of quotes falling inside the interquartile range. The level models are remarkably well specified for this criterion with 50% hit rates. The heteroskedastic logarithms models are not as well specified.

We now turn to the second column entitled fit cover. The fit cover represents the cover rate when we use, not the predictive density, but the fit density. The fit density is obtained by drawing from the parameters with the model error set equal to zero. This is consistent with studies which have tested the B-S model only allowing for parameter uncertainty, see Lo (1986). [19] As shown in column 2, one would easily reject any of the models. Their 50% intervals cover the true value only 2% of the time for the B-S, and 13% of the time for the level model 2. In fact, this approach is flawed. It is not the B-S that does not generate enough variability, it is the fit density which is inappropriate for the test because it does not include the model error. When one uses the predictive density, even the B-S appears reasonably well specified for this test.

The final columns of the panel deal with the problem of the intrinsic bounds. No call model should generate negative call values, bound B_1=0, or values below B_2=S-PV(X). The logarithm models avoid the negativity problem by construction. We investigated the extent of the problem for the levels models. The column entitled B_1 shows the percentage of quotes for which the predictive density implies a probability of negative call values larger than 0.1%. There are no more than 3% of such quotes for model 2, and 0.3% for model 3. Now consider the second lower bound, B_2. Actually, 20% of the quote midpoints and 3% of the asks violate this second bound. We left these points in the estimation sample. We then computed the number of observations such that the first

[18]We could for example compute the relative error implied by the levels models.

[19]The procedure is not exactly the same though since our parameter uncertainty comes from the option prices, not from the time series of the underlying asset's return.

quartile of their predictive density violated B_2. For this test, we considered only the observations with a bid not violating this same bound. After the necessary correction for heteroskedasticity, the logarithm and level models were similar in this respect, see the last column. Less than one in a hundred predictions had this undesirable property.

The previous results, even the predictive densities, were in-sample tests computed for the observations used to estimate the parameters. Given these parameters, we now turn to the out-of-sample analysis. For each of the two weeks, we used the parameter draws to compute residuals, fit, and predictive densities for the quotes of the following week. This resulted in an out sample of 1043 quotes.

Table 2 summarizes the evidence in a format similar to table 1. Again, the biases are small and we do not discuss them. Instead consider the RMSE's. The first result is that the extended models do not improve the B-S's out of sample RMSE. This is the case for both residuals and pricing errors, levels or logarithms. Second, the magnitude of these out of sample errors is similar to their within sample counterparts in table 1. There is no significant deterioration of performance when going from within to out of sample estimation. Looking more closely at outside B-A spread RMSE's, one can conclude that: 1) the B-S model has similar performance in and out of sample, while the performance of models 2 and 3 has deteriorated to worse than B-S, and 2) the level models seem to have deteriorated more than the logarithm models.

This is in sharp contrast with the results of Dumas, Fleming, and Whaley (1995) who report a severe deterioration of their models in out of sample tests, while having a nearly perfect fit in sample. This is most likely due because we allow for model error in our estimation. This model error has properties which do not deteriorate out of sample. We do find, however, that the Black-Scholes model is a formidable competitor in out of sample specification tests.

We complete the out-of sample tests with the third panel of table 2. The first column yields results consistent with the in-sample tests. As indicated earlier, 20% of the in sample means of predictions were outside the B-A spread, now 30% of the out-of-sample predictions are outside the spread. We also break down the interquartile range (IQR) coverage ratios by the in-the money and out-of-the money categories. This shows how crucial the heteroskedasticity correction is. The homoskedastic models predictive IQR covers the quote 98% of the time for in the money quotes, a poor performance. The heteroskedasticity brings it down significantly.

5 Extensions

This section reports further empirical results and discusses extensions to the previous models

5.1 Updating

In most cases, one wants to reestimate the model regularly and keep predictions for a future time near the estimation time. This is due to the fact that model misspecifications are likely to result in time varying parameters. For example, this holds when volatility, hedging demand, or liquidity are time varying and only accounted for through the expansion and the intercept terms of the extended model.

The estimation is then based on updated (daily) cross-sections of option prices. The cross-sections are indexed by t. Technically, the calibration of priors p_t to the desired time varying mean and variance is easy. For an inverted gamma prior $p_t(\sigma)$, it is easy to choose ν_0 and s_0^2 to match a desired mean and variance. Similarly for σ_η, one can set ν_1 and s_1^2 to obtain the desired prior mean and variance. Finally consider β. We wish to formulate a prior $p_t(\beta) \sim N(\beta_0, \sigma_\eta^2 V_0)$. We could use the joint random draws of the time t-1 posterior of (σ_η, β). The sample mean of these draws yields β_0. The sample covariance matrix of the draws of $\frac{\beta}{\sigma_\eta}$ yields V_0.

Do we want to use just the time t-1 posterior to build the time t prior? This may be desired if the daily parameter estimates followed a random walk. Given the probable reasons for the time variation of the parameters, this is unlikely. For example, the most probable source of misspecification - volatility - is unlikely to be non-stationary even though it is strongly autocorrelated. The information on previous, t-2, t-3, .., periods may also be important. A candidate for the formulation of time t priors follows from the time series analysis of the previous posteriors and a one step ahead forecast. There are two alternatives to this approach. First, to keep several panels of options in the likelihood and to model the time series variation of the parameters. Second, to extend σ by some function of proxies for time varying volatility. The MCMC estimator can accommodate either of these extensions.

We now document the performance of the models through a simplified updating scheme. Using 8749 quotes from January 01 to March 30 1990, we reestimate the parameters every second trading day. Given little evidence supporting improvement the larger extended models, we restrict the comparison to B-S and model 2, allowing for heteroskedasticity.

Figure 7 documents the time series variation of some of the parameters. The top left plot shows the volatility σ. The top right plot shows the model error standard deviation σ_η for the homoskedastic models. The parameters of model 2 show more time variation than those of the B-S. The top right plot shows that the model 2 error's standard deviation is consistently lower than that of the B-S. The bottom plots show the two levels of σ_η for the heteroskedastic models. Except for one day near the end of the period, the out of the money quotes ($\sigma_{\eta,1}$) have large errors than the in the money quotes.

We next compute the pricing errors for B-S and model 2. They are summarized as follows:

In Sample Performance: Bi-Daily Reestimation

Criterion	B-S	Model 2
RMSE	0.31	0.18
MAE	0.18	0.12
RMSE outside B-A	0.49	0.31
Number of Pred. outside B-A	3200	2260

MAE is the mean absolute error. In sample, model 2 exhibits significantly smaller errors by all of the above criteria. Further, there were 627 observations for which model 2's mean prediction was outside the B-A spread and the B-S prediction inside, but the reverse happened 1568 times. Both models had prediction means simultaneously outside the B-A spread 1632 times, and in the BA spread 4922 times. Figure 8 shows where model 2 improves over the B-S. We plot, for each model, the pricing error versus the moneyness. Model 2, the top plot, suffers less from the overpricing of the out-of-the money quotes than does B-S. The gap in the plots around the zero line occurs because we only plotted the predictions outside the B-A spread.

5.2 Error Specification

We now propose a more general model of the error consistent with the presence of intermittent mispricing. The intuition underlying our formulation is that most quotes contain no market error and their only error is η_i. In rare occurrences, however, an additional error ϵ_i with standard deviation σ_ϵ occurs. We have already indicated how to conduct a residual analysis which may identify outlying observations. It can be superior, however, to incorporate this possibility of rare errors into the model's error distribution. The resulting diagnostics are then easier to interpret, e.g., an estimate for the probability of a given observation having the extra error can be obtained. The estimation and prediction procedures are also more reliable since they incorporate the existence of these mispricing errors. The cost of this extension, the added burden on the estimation, has to be weighted against the likelihood that such errors are present.

We model the *market* error as

$$\epsilon_i \begin{cases} = & 0 \text{ with prob. } 1-\pi, \\ \sim & N(0, \sigma_\epsilon) \text{ with prob. } \pi. \end{cases}$$

This error formulation implies that the mispricing error is zero most of the time. It is similar to formulations used to model regime changes in Economics, see Hamilton (1987). [20] The estimation of this model can be conducted in two ways. First, it may be based upon a given value for π, the unconditional probability of mispricing. Second, we can formulate a prior distribution for π and let the data update it. We outline the second approach which subsumes the first.

First, we specify a prior for π, which can be centered close to zero. We choose a Beta distribution which parameters can generate a great variety of shapes on [0,1], from uniform to centered as near

[20] π can be made to depend on observables. This would be more in the spirit of a model specification test.

zero as desired. The model can be thought of involving a state variable s_i equal to 1 (market error) or 0 (no market error). For a sample of size n, the 2^n possibilities for the state vector S make both traditional Bayesian and maximum likelihood analysis very complicated. Also, given the size of the parameter space, the asymptotic approximation underlying the maximum likelihood method becomes questionable for conventional sample sizes.

As before, the resolution of this problem is facilitated by the incorporation of the Gibbs sampling algorithm. [21] The parameters are $\sigma, \sigma_\eta, \sigma_\epsilon, \beta, \pi$. We can use the Gibbs algorithm if we first augment the parameter space by the state vector S, see Tanner and Wong (1987). Given (S, π, σ_ϵ), the system is similar to that already analyzed. The errors do not have equal variance, but they can be standardized given σ_ϵ and S. Next, we can find the posterior distribution of σ_ϵ given (β, σ_η, S). For the priors formulated in equation (10), appendix C shows how to make draws from the following N + 3 conditional posterior distributions:

$$p(\sigma_\eta, \sigma, \beta \mid \underline{y}_t, \sigma_\epsilon, S),$$

$$p\left(\sigma_\epsilon \mid \underline{y}_t, \beta, \sigma_\eta, S\right),$$

$$p(s_i = 1 \mid \underline{y}_t, S_{-i}, .), \quad \text{for } i = 1, \ldots, N$$

$$p(\pi \mid \underline{y}_t, S),$$

where $S_{-i} = S - s_i$. The Gibbs algorithm guarantees that a sequence of draws from these conditional posterior distributions converges to the desired marginal posteriors. Consider the first of these and recall that we can not draw directly from σ_η, but must use a a Metropolis algorithm. Direct draws can be made from the other N + 2 distributions.This estimator is a Markov Chain estimator combining the Gibbs and the Metropolis algorithms. It can be shown to converge (see Jacquier, Polson, and Rossi (1994) for an example). For each observation, we obtain the posterior probability that is contains a market error. We also obtain the overall posterior probability of market error, $P(\pi \mid \underline{y}_t)$, the posterior distribution of σ_ϵ, and the usual parameters and predictions discussed in the previous sections.

[21] See Geman and Geman (1984), Gelfand and Smith (1990), and Mc Culloch and Tsay (1993) for a related analysis.

5.3 Unobservable Inputs

This section discusses the inclusion of other unobservable inputs into the estimation procedure. An estimator when the stock price or the interest rate are not observable can be formulated in a fashion similar to the previous analysis. This approach models the uncertainty introduced by the measurement error directly, rather than lumping it into the additive error term. For example, the input for the stock's error in the Black-Scholes formulation would be

$$S_i = S_i^* + \nu_i, \;\; \nu_i \sim N(0, \sigma_n u),$$

where S_i^* is the observation in the database. Priors can be imposed on σ_ν to reflect bid-ask spread limits. Given the small size of ν compared to the magnitude of S, the normality assumption is a reasonable start. Given simulation evidence (Renault (1995)) concerning non-synchroneity errors, the above formulation may be a worthwhile extension.

Alternatively, one may want to allow for an observation error for the risk free rate. The Black-Scholes input would then be $r_{f,i} = r^{\star}_{f,i} + \nu_i$, where $r^{\star}_{f,i}$ is the most relevant observed risk free rate, and ν_i an observation error. In this case, the risk free rate is an unobserved state variable. Again, the estimator can be extended to nest an optimal signal extraction for the state variable. It is the hierarchical structure of MCMC estimators that allows the extension to unobservable state variables. In practice, this results in additional conditional distributions from which to draw, as seen in section 5.2.

6 Conclusion

This paper develops a new method which incorporates model error into the estimation of contingent claim models. This method allows one to estimate deterministic functions of the parameters, to generate the residual of an in-sample observation and assess its abnormality, and to determine the predictive density of an out-of-sample estimation. The MCMC estimators which we implement are very flexible. Additional state variables can be incorporated to study more complex error structures, or allow for non observable inputs such as the underlying price or the risk free rate.

We apply this method to about 9600 quotes on stock calls options, and document the behavior of several competing models nesting the B-S. The competing models are justified as an approximation to an unknown model or a complex model too costly to implement. We formulate the error in both relative terms (logarithm models) and dollar terms (level models). Our analysis shows within sample evidence of Black Scholes mispricing. By some criteria, the extended models dominate the B-S within sample. Indeed they reduce root mean squared errors of pricing and residuals. The improvement, however, deteriorates for the larger models with 12 or more parameters. The extended models also have different hedging and pricing implications than does the B-S model. Allowing model error to be heteroskedastic is shown to stongly improve the performance of most models.

We also show that the failure to include model error in specification tests results in a severe bias towards rejection. For example, the interquartile range of a predictive distribution which should cover the true value 50% of the time, would be wrongly believed to cover the true value 2% of the time, leading to a rejection of the model.

The Black-Scholes model appears robust in out of sample performance as many of the advantages demonstrated by the extended models disappear. For most models, the out of sample performance does not display a drastic deterioration compared to the in sample performance. This insight is obtained because our estimation technique formally incorporates model error. Our results are in strong constrast with those of recent studies which do not allow for model error. The extended models imply different hedge ratios than the basic model. Therefore, they have different implications if these difference are important out of sample. The analysis of these differences awaits further research.

APPENDIX

A Homoskedastic Extended Model: Posterior Distributions

Consider the model

$$\begin{aligned} \log C_i &= \beta_1 \log BS\left(\sigma, x_{1i}\right) + \underline{\beta}_2' x_{2i} + \eta_i, \quad \eta_i \sim N\left(0, \sigma_\eta\right) \\ R_t &= \mu + \xi_t, \qquad\qquad\qquad\qquad\qquad \xi_t \sim N\left(0, \sigma\right) \end{aligned}$$

The likelihood function is:

$$\ell(\sigma_\eta, \sigma, \beta_1, \beta_2' \mid \underline{y}_t) \propto \frac{1}{\sigma_\eta^N} \times \exp\left\{ -\frac{\sum_1^N \left[\log C_i - \beta_1 \log b(\sigma, x_{1i}) - \beta_2' x_{2i}\right]^2}{2\sigma_\eta^2} \right\}$$

To simplify the notation, let $x_i' = (\log m(\sigma, x_{1i}),\, x_{2i}')$, $X' = (x_1, ..., x_N)$, $\beta' = (\beta_1, \beta_2')$, and $Y' = (\log C_1, ..., \log C_N)$. We formulate the following joint prior distribution for the parameters

$$\begin{aligned} p\left(\sigma, \sigma_\eta, \beta\right) &\propto p(\sigma) p(\sigma_\eta) p(\beta \mid \sigma_\eta) \\ &= \mathrm{IG}(\sigma : \nu_0, s_0^2)\ \mathrm{IG}(\sigma_\eta : \nu_1, s_1^2)\ \mathrm{N}(\beta : \beta_0, \sigma_\eta^2 V_0) \end{aligned}$$

To reflect the fact that p(σ) is based on the returns data, let $\nu_0 s_0^2 = \sum_1^{T_r} (R_t - \bar{R})^2$ and $\nu_0 = T_r + 1$. Apply Bayes theorem. The joint density of the parameters is

$$p(\sigma_\eta, \sigma, \beta \mid \underline{y}_t) \propto \frac{\exp\left\{-\frac{\nu_0 s_0^2}{2\sigma^2}\right\}}{\sigma^{\nu_0}} \times \frac{1}{\sigma_\eta^{N+\nu_1}} \times \exp\left\{ -\frac{(Y - X\beta)'\,(Y - X\beta) + \nu_1 s_1^2}{2\sigma_\eta^2} \right\}$$

Let k be the dimension of β. Consider the quantities

$$\hat{\beta} = (X'X)^{-1} X'Y, \quad V = \left[X'X + V_0^{-1}\right]^{-1}, \quad \bar{\bar{\beta}} = V\left[X'X\hat{\beta} + V_0^{-1}\beta_0\right],$$

and

$$\nu_\eta = N - k + \nu_1 - 1, \quad \nu_\eta s_\eta^2 = \left(Y - X\bar{\bar{\beta}}\right)' \left(Y - \bar{\bar{\beta}}\right) + \left(\beta_0 - \bar{\bar{\beta}}\right)' V_0^{-1} \left(\beta_0 - \bar{\bar{\beta}}\right) + \nu_1 s_1^2.$$

The joint density can be rewritten as

$$p\left(\sigma_\eta, \sigma, \beta \mid \underline{y}_t\right) \propto \frac{1}{\sigma^{\nu_0}} \times \exp\left\{\frac{-\nu_0 s_0^2}{2\sigma^2}\right\} \times \frac{1}{\sigma_\eta^{\nu_\eta + 1}} \times \exp\left\{ \frac{-\nu_\eta s_\eta^2 - \left(\beta - \bar{\bar{\beta}}\right)' V^{-1} \left(\beta - \bar{\bar{\beta}}\right)}{2\sigma_\eta^2} \right\}.$$

It is analogous to that resulting from a standard regression model with the twist that X, $\hat{\beta}$, $\bar{\bar{\beta}}$, V, and $\nu_\eta s_\eta^2$, are functions of σ. We can now break down the joint density in the conditionals of interest. First,

$$p(\beta \mid \sigma, \sigma_\eta, \underline{y}_t) \sim N\left(\bar{\bar{\beta}}, \sigma_\eta^2 V\right). \tag{6}$$

The joint density of σ and σ_η is then

$$p\left(\sigma_\eta, \sigma \mid \underline{y}_t\right) \propto \frac{1}{\sigma^{\nu_0}} \times \exp\left\{\frac{-\nu_0 s_0^2}{2\sigma^2}\right\} \times \frac{1}{\sigma_\eta^{\nu_\eta+1}} \times \exp\left\{\frac{-\nu_\eta s_\eta^2}{2\sigma_\eta^2}\right\} \times |V|^{1/2}.$$

The conditional posterior density of σ_η is

$$p\left(\sigma_\eta \mid \sigma, \underline{y}_t\right) \sim \mathrm{IG}(\nu_\eta = N - k + \nu_1, \nu_\eta s_\eta^2(\sigma)) \tag{7}$$

The posterior density of σ is

$$p(\sigma \mid \underline{y}_t) \propto \frac{1}{\sigma^{\nu_0}} \exp\left\{\frac{-\nu_0 s_0^2}{2\sigma^2}\right\} \times \left[\nu_\eta s_\eta^2(\sigma)\right]^{-\nu_\eta/2} \times |V|^{1/2}. \tag{8}$$

The distribution in equation (8) is the marginal posterior distribution of σ. A draw from σ_η, σ, and β can be made by: (1) a metropolis draw from expression (8), (2) a draw from (7), and (3) a draw from (6). The Gibbs step is not needed. It is however needed for extensions of the model, e.g., heteroskedastic model error.

B σ: The Metropolis Step

This appendix discusses the σ draws. In the extended model with prior distribution, the posterior distribution of σ conditional on the other parameters is shown to be

$$p(\sigma \mid \beta, \sigma_\eta, \underline{y}_t) \propto \frac{\exp\left\{-\frac{\nu_0 s_0^2}{2\sigma^2}\right\}}{\sigma^{\nu_0}} \times \exp\left\{\frac{\nu s^2(\sigma, \beta)}{2\sigma_\eta^2}\right\},$$

where $\nu s^2(\sigma, \beta) = (Y - X(\sigma))'(Y - X(\sigma))$. Note that it is different from expression (9). For computational convenience, we introduce the sample statistic $\nu_\eta s_\eta^2\star$, the mode of the kernel, and rewrite the posterior density of σ as

$$p(\sigma \mid \beta, \sigma_\eta, \underline{y}_t) \equiv \frac{K}{\nu_\eta s_\eta^2\star} \times \mathrm{IG}(\nu_0, \nu_0 s_0^2) \times \exp\left\{\frac{\nu s^2(\sigma, \beta)}{2\sigma_\eta^2}\right\} \tag{9}$$

Recall that we draw in sequence $(\beta \mid .)$, $(\sigma \mid .)$, and $(\sigma_\eta \mid .)$, building a chain of such draws. There is no analytical expression for K, but it could be computed numerically by importance sampling from the first kernel in equation (9). This is unrealistic as: (1) we would have a new K to compute everytime we make a draw of σ because β and σ_η have changed, and (2) even then, direct draws from (9) by conventional methods such as inverse CDF are unrealistic. Instead, we use an Metropolis algorithm that does not require the computation of K.

The Metropolis algorithm (see Metropolis et al. (1953) and Tierney (1991)) nests a simpler algorithm, the accept/reject, (see Devroye (1987)), which requires the knowledge of K. We explain the accept/reject algorithm first. We cannot draw directly from the density p(σ). There is a *blanketing* density q(σ) from which we can draw, and which meets the condition that there exists a finite number c such that $cq(\sigma) > p(\sigma)$, for all σ. Draw from q a number σ and accept the draw with probability p(σ)/$cq(\sigma)$. The intuition of why this produces a sample of draws with distribution p(σ) is simple: We draw from q and for each draw we know by how much cq dominates p. p/cq is not the same for every value of σ because p and q do not have the same shape. The smaller p/cq, the more q dominates p, the more likely we are to draw too often in this area, the less likely the draw is to be accepted. If the parameter space is unbounded,a finite c such that $cq(\sigma) > p(\sigma), \forall\sigma$, exists only if the tail of q drops at a slower rate than the tail of p. For density (9), this can be accomplished if q is an inverted gamma with parameter $\nu \leq T_r - 1$. Given that c exists, an ideal density is such that p / q is relatively constant over σ. Otherwise c needs to be very large, and we will waste time rejecting many draws. Experimentation shows that the inverted gamma may have a shape very different from (9), particularly if the option kernel is more informative than the returns kernel. This is because q must have low degrees of freedom ($\nu \leq T_r - 1$) for c to exist. q is not allowed to tighten when the information in the options data increases. An extreme case of this occurs if we only use option data. Also, the calculation of c is non trivial. One must first calculate K rather precisely, and then solve for the minimum of p/q over σ. Therefore the accept-reject algorithm alone does not help us.

This is where the Metropolis algorithm intervenes. For any candidate density q, we can always find a c such that cq > p, for most values of σ. For some values of σ, cq < p, i.e., the density q does not dominate p everywhere. In these regions, we do not draw often enough from q, and therefore, underestimate the mass under the density p. The Metropolis algorithm is a rule of how to *repeat* draws, i.e., build mass for values of σ, where q does not draw often enough. This algorithm does not require dominance everywhere, and gives us more choices for the density q and the number c. For a given density q, too large a c leads to frequent rejections, and too low a c produces many repeats, but the algorithm is still valid. A c which trades off these two costs can be computed very quickly. Furthermore we do not need to compute K in (9) anymore. This is because the Metropolis is a Markov Chain algorithm with transition kernel a function of the ratio p(y) / p(x), where x and y are the previous and the current candidate draws. K disappears from the ratio. Consider an independence chain with transition kernel f(z) $\propto$ min{ p(z), cq(z)}. The chain repeats the previous point x with probability 1-α, where $\alpha(x,y) = min\left\{\frac{w(y)}{w(x)}, 1\right\}$, where w(z) $\equiv$ p(z)/f(z). If cq > p, w(z)=1, and if cq < p, w(z) >1. The decision to stay or move is based upon $\frac{w(y)}{w(x)}$ which compares the (lack of) dominance at the previous and the candidate points.

We implement the Metropolis Accept Reject algorithm as follows. A truncated normal distribution was found to have a shape close to p. We choose it as blanketing density q. The truncation is effected by discarding negative draws. We have not encountered such draws even in the smallest samples where the mean is still more than 6 standard deviations away from 0. A possible alternative to the normal blanket would be the lognormal distribution. We set the blanket mean equal to the mode of p($\sigma \mid \underline{y}_t$). The mode is found quickly in about 10 evaluations of the kernel. The variance of q is then set to best match the shape of q to that of p. For this, the discrepancy function p$\star$/q, where p$\star$ is the kernel of p, is computed and minimized at 3 points, the mode and 1 point on each side of the mode, at which p is half the height of p at the mode. They are found in a few evaluations of the kernel. The minimization requires and additional 10 evaluations. This brings q as close as possible to p in the bulk of the distribution where about 70 % of the draws will be made. Possible values for c are the ratios $p\star/q$ at these three points. We choose c so as to slightly favor rejections over repeats. The top left plots of figure 1 show that the ratio p⋆/cq is close to 1 almost everywhere. The intuition of the ratio p⋆/cq is as follows. If a candidate draw is at the mode, ratio = 1, and the previous draw is at the upper dotted line, ratio = 1.1, then there is a 1/1.1 chance that the previous draw will be repeated rather than the candidate draw chosen. Also, a draw at 0.27, ratio = 0.93, has a 7% chance of being rejected. The potential efficiency of the algorithm is verified when we keep track of the actual rejections and repeats in the simulation. Even in a very short sample such as the case of figure 1, 140 quotes and 10 parameters, we got no more than 32 rejections and 28 repeats over 5000 draws.

C Analysis of Market Error

Consider

$$\begin{aligned}
\log C_i &= \beta_1 \log m(\sigma, x_{1i}) + \beta_2' x_{2i} + a_i, \text{ where } a_i = \eta_i + s_i \epsilon_i \\
&= \beta' x_i + a_i \\
R_t &= \mu + \xi_t \\
\eta_i &\sim N(0, \sigma_\eta), \quad \epsilon_i \sim N(0, \sigma_\epsilon), \quad \xi_t \sim N(0, \sigma) \\
s_i &= \begin{cases} 0 & \text{with prob. } 1 - \pi \\ 1 & \text{with prob. } \pi \end{cases}
\end{aligned}$$

The variance of a_i is $\sigma_i^2 = \sigma_\eta^2 + s_i \sigma_\epsilon^2 \equiv \sigma_\eta^2(1 + s_i\omega)$. Introduce the state vector S$=\{s_1, ..., s_N\}$, a sequence of independent Bernouilli trials. Consider the prior distributions

$$\begin{aligned}
\pi &\sim B(a, b) \\
(\beta \mid \omega, \sigma_\eta) &\sim \mathrm{N}\left(\beta_0, \sigma_\eta^2(1 + \frac{a}{a+b}\omega) V_0\right) \\
\sigma_\eta &\sim \mathrm{IG}\left(\nu_1, s_1^2\right) \\
\sigma &\sim \mathrm{IG}\left(\nu_0, s_0^2\right).
\end{aligned} \tag{10}$$

where IG and B are the Inverted Gamma and the Beta distributions. These priors can be made arbitrarily diffuse by setting ν_0 and ν_1 to 0, and the diagonal elements of V_0 to large values. Note that σ_ϵ is modelled through the specification of ω. The goal is to obtain the posterior distributions of $\beta, \sigma, \sigma_\eta, \pi, \omega$, and S either joint or marginal. The first conditional posterior is that of $(\beta, \sigma, \sigma_\eta \mid \underline{y}_t, \omega, S)$:

1:

$$p(\sigma_\eta, \sigma, \beta \mid \underline{y}_t, \omega, S) \propto \frac{\exp\left\{\frac{-\nu s^2}{2\sigma^2}\right\}}{\sigma^{\nu_0 + T_r}} \times \frac{\exp\left\{\frac{-\nu_1 s_1^2}{2\sigma_\eta^2}\right\}}{\sigma_\eta^{\nu_1+1+k+N_0}\left(\sigma_\eta\sqrt{1+\omega}\right)^{N-N_0}}$$

$$\times \exp\left\{-\frac{(\beta-\beta_0)' V_0^{-1}(\beta-\beta_0)}{2\sigma_\eta^2\left(1+\frac{a}{a+b}\omega\right)}\right\} \times \exp\left\{-\frac{(Y^* - X^*\beta)'(Y^* - X^*\beta)}{2\sigma_\eta^2}\right\}$$

where N_0 is the number of observations for which s_i is zero. $Y^* = (\log C_1^*, ..., \log C_N^*)'$, where $\log C_i^* = \log C_i/(1+\omega s_i)^{.5}$. The same transformation is applied to the vector X, i.e. each element is divided by $\sqrt{1+s_i\omega}$. After this transformation, a draw of this posterior is made as shown in section 3. Now consider ω introduced above. Given S, the likelihood function of ω depends only on the $N_1 = N - N_0$ observations for which $\mathrm{s}_i = 1$. Consider for $\bar{\omega} = 1+\omega$, a truncated inverted gamma prior distribution $\mathrm{IG}(\nu_2, s_2^2)\, I_{\bar{\omega}>1}$. The posterior distribution of $\bar{\omega}$ conditional on the other parameters is

2:

$$\begin{aligned} p\left(\bar{\omega} \mid \underline{y}_t, \beta, \sigma_\eta, S\right) &\propto \frac{1}{\bar{\omega}^{1+\nu_2+N_1}} \times \exp\left\{ -\frac{\sum_{i\in N_1} (Y_i - \beta' x_i)^2}{2\sigma_\eta^2 \bar{\omega}^2} + \nu_2 s_2^2 \right\} \mathrm{I}_{\bar{\omega}>1} \\ &\sim IG\left(\nu_2 + N_1, \nu_\omega^2 s_\omega^2 = \nu_2 s_2^2 + \sum_{i\in N_1} \left(\frac{Y_i - \beta' x_i}{\sigma_\eta}\right)^2\right) \mathrm{I}_{\bar{\omega}>1} \end{aligned}$$

where $\mathrm{I}_{\bar{\omega}>1}$ is the indicator function for $\bar{\omega} > 1$. A draw of ω is obtained directly from a draw of $\bar{\omega}$ since $\bar{\omega} = 1+\omega$. We now need the conditionals $\mathrm{p}\left(s_i \mid \underline{y}_t, S_{-i}, .\right)$ where "." stands for all the other parameters, and S_{-i} refers to the state vector without s_i. Following McCulloch and Tsay (1993), they are written as

3:

$$\begin{aligned} p(s_i = 1 \mid y, S_{-i}, .) &= \frac{\pi p(y_t \mid s_i = 1, .)}{\pi p(y_t \mid s_i = 1, .) + (1-\pi) p(y_t \mid s_i = 0, .)} \\ &= \frac{1}{1 + \frac{1-\pi}{\pi} \times \frac{p(y_t \mid s_i = 0, .)}{p(y_t \mid s_i = 1, .)}} \end{aligned}$$

For the set up considered here the denominator term is simply:

$$\frac{y_t \mid s_i = 0, .}{y_t \mid s_i = 1, .} = \sqrt{(1+\omega)} \exp - \frac{(\log C_i - \beta' x_i)^2}{2\sigma_\eta^2} \times \frac{\omega}{1+\omega}$$

We now need the last conditional posterior of π . It depends exclusively on. With N_1 the number of s_i's equal to 1, we have

4:

$$p(\pi \mid S, .) \sim B(a + N_1, b + N - N_1)$$

References

Bates, D, (1995): "Testing Options Pricing Models", Rodney L. White Working Paper, 14-95, The Wharton School. To appear in the Handbook of Statistics, vol. 14; G.S. Maddala and C.R. Rao editors, Amsterdam: Elsevier.

Black, F., and Scholes, M., (1973): "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy*, May, 637-659.

Bossaerts, P., and Hillion, P. (1994): "Method of Moments Estimation of Contingent Claims Asset Pricing Models," *Journal of Mathematical Finance.*

Bossaerts, P., and Hillion, P. (1994): "Local Parametric Analysis of Hedging in Discrete Time", California Institute of Technology Working Paper.

Casella, G., and George, E., (1992): "Explaining the Gibbs Sampler", *The American Statistician*, 46, 167-174.

Chaloner, and Brant, (1988): "A Bayesian Approach to Outlier Detection and Residual Analysis", *Biometrika.*

Clément, E., and Gouriéroux, C.,, and Montfort, A., (1993): "Prediction of Contingent Price Measures", *Working Paper I.N.S.E.E.*, no 9329.

Devroye, L. (1987): *A Course in Density Estimation*, Boston, Birkhauser

Dumas, B., Fleming, J.,, and Whaley, R., (1995): "Implied Volatility Functions: Empirical Tests", H.E.C. and Duke Working Paper.

Engle, R., Kane, A.,, and Noh, J., (1993): "Index-Option Pricing with Stochastic Volatility and the Value of Accurate Variance Forecasts", N.B.E.R. Working Paper, # 4519.

Gelfand, A., and Smith, A., (1990): "Sampling Based Approaches to Calculating Marginal Densities", *Journal of the American Statistical Association*, 85, 398-409.

Geman, S., and Geman, D., (1984): "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.

Gultekin, N., Rogalski, R. , and Tinic, S., (1982): "Option Pricing Model Estimates: Some Empirical Results", *Financial Management*, Spring, 58-69.

Hamilton, J. D., (1989): "A new approach to the Economic Analysis of Time Series and the Business Cycle", *Econometrica*, 57, 357-384.

Harrison, J., and Pliska, S., (1981): "Martingales and Stochastic Integrals in the Theory of Continuous Trading", *Stochastic Processes and Their Applications*, 11, 215-260.

Heston, S. (1993): "A Closed Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options". *Review of Financial Studies*, 6, 2, 327-343.

Hull, J.,, and White, A. (1987): "The Pricing of Options on Assets with Stochastic Volatility", *Journal of Finance*, 3, 281-300

Hutchinson, J., Lo, A, and Poggio, T. (1994): "A Non Parameteric Approach to Pricing and Hedging Derivative Securities via Learning Networks". *Journal of Finance*, 49, 3.

Jacquier, E., Polson, N., and Rossi, P., (1994): "Bayesian Analysis of Stochastic Volatility Models", *Journal of Business and Economic Statistics*, 12, 4, 371-418.

Jacquier, E., Polson, N., and Rossi, P., (1995): "Efficient Option Price Prediction under Stochastic Volatility", Manuscript, The Wharton School, University of Pennsylvania.

Jarrow, R.A., and Rudd, A., (1982): "Approximate Option Valuation for Arbitrary Stochastic Processes", *Journal of Financial Economics*, 10, 347-369.

Lo, A., (1986): "Tests of Contingent Claims Asset Pricing Models", *Journal of Financial Economics*, 17, 143-173.

Macbeth, J., and Merville, L., (1979): "An Empirical Examination of the Black-Scholes Call Option Pricing Model", *Journal of Finance*, 35, 285-301.

McCulloch R., and Tsay, R., (1993): "Bayesian Inference and Prediction for Mean and Variance Shifts in Autoregressive Time Series", *Journal of the American Statistical Association*,88,968-978.

Merton, R. C. (1973): "Theory of Rational Option Pricing", *Bell Journal of Economics and Management Science*, 4, 141-183.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953), "Equations of State Calculations by Fast Computing Machines," *J. Chemical Physics*, 21, 1087-1091.

Platen, E., and Schweizer, M., (1995): "On Smile and Skewness", *Working Paper, Stochastic Analysis Group, Australian National University*

Renault, E. (1995): "Econometric Models and Option Pricing Errors", Invited Address at the 1995 World Econometric Congress in Tokyo. I.D.E.I. Toulouse working paper.

Renault, E., and Touzi, N. (1992): "Option Hedging and Implicit Volatility", CREST / GREMAQ Working Paper.

Rubinstein, M., (1985): "Nonparametric Tests of Alternative Option Pricing Models using all Reported Trades and Quotes on the 30 Most Active CBOE Option Classes from August 23, 1976 through August 31, 1978", *Journal of Finance*, 40, 455-480.

Rubinstein, M., (1994): "Implied Binomial Trees", *Journal of Finance*, 49, 3, 771-818.

Schotman, P., (1994): "A Bayesian Approach to the Empirical Valuation of Bond Options", forthcoming *Journal of Econometrics.*

Tanner, T.,, and Wong, W., (1987): "The Calculation of Posterior Distributions by Data Augmentation", *Journal of the American Statistical Association*, 82, 528-549.

Taylor, S.J.,, and Xu, X., (1993): "The Magnitude of Volatility Smiles: Theory and Empirical Evidence for Exchange Rates", *Working Paper, Lancaster University.*

Tierney, L. (1991): "Markov Chains for Exploring Posterior Distributions", Technical Report 560, University of Minnesota School of Statistics.

Whaley, R., (1982): "Valuation of American Call Options on Dividend-Paying Stocks: Empirical Tests", *Journal of Financial Economics,* 10, 29-58.

Figure 1: MCMC Estimation, Sigma and Sigeta_2, TOYS'R US, Dec 1, 89: 140 quotes, 10 Parameters

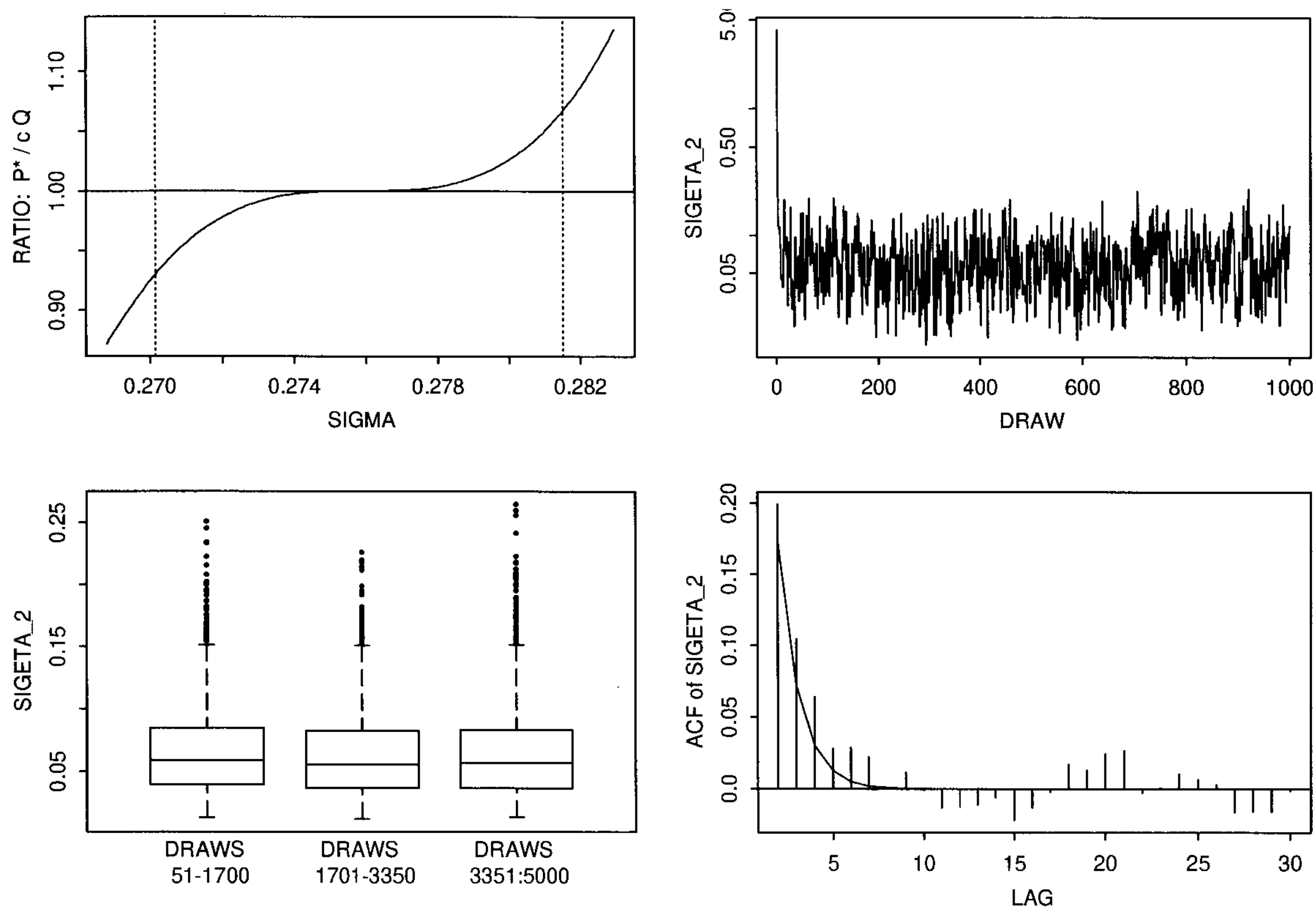


Figure 2: Sigma, Intercept and Slope, Logarithm and Level Models
TOYS'R US, 456 quotes Dec. 04-08, 1989

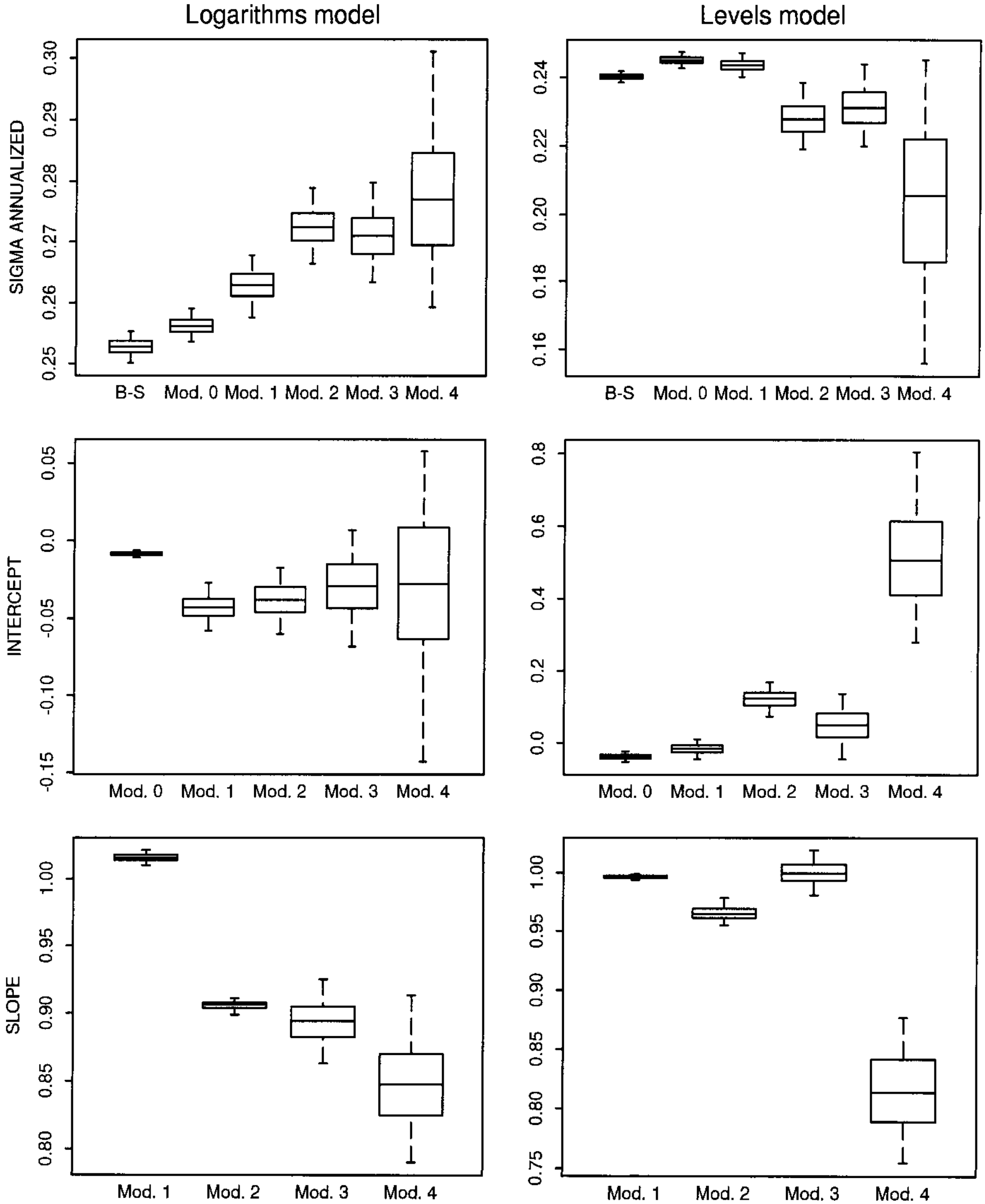


Figure 3a: Model Error and Heteroskedasticity, Logarithms Model
TOYS'R US, 456 quotes Dec. 04-08, 1989

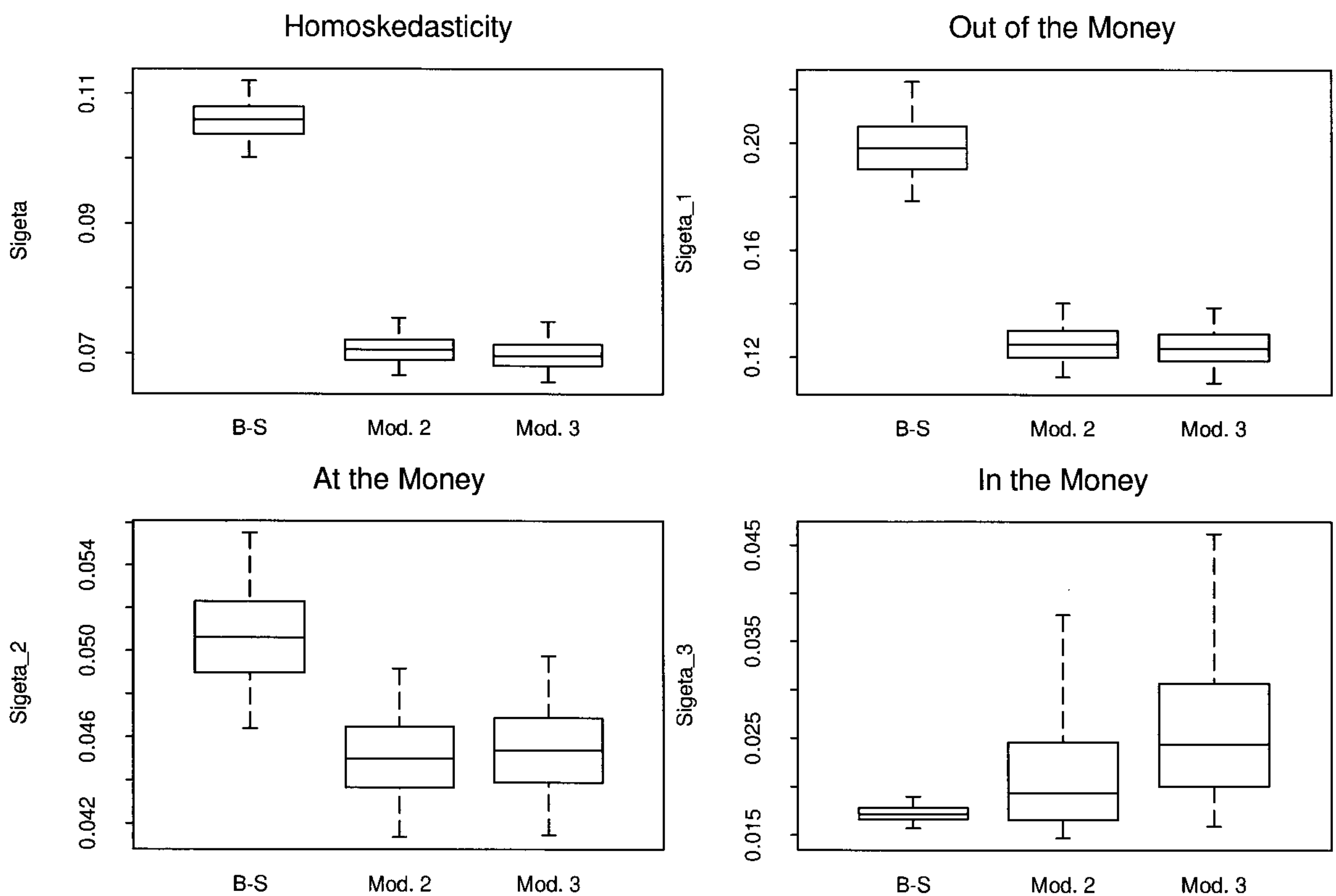


Figure 3b: Model Error and Heteroskedasticity, Levels Model
TOYS'R US, 456 quotes Dec. 04-08, 1989

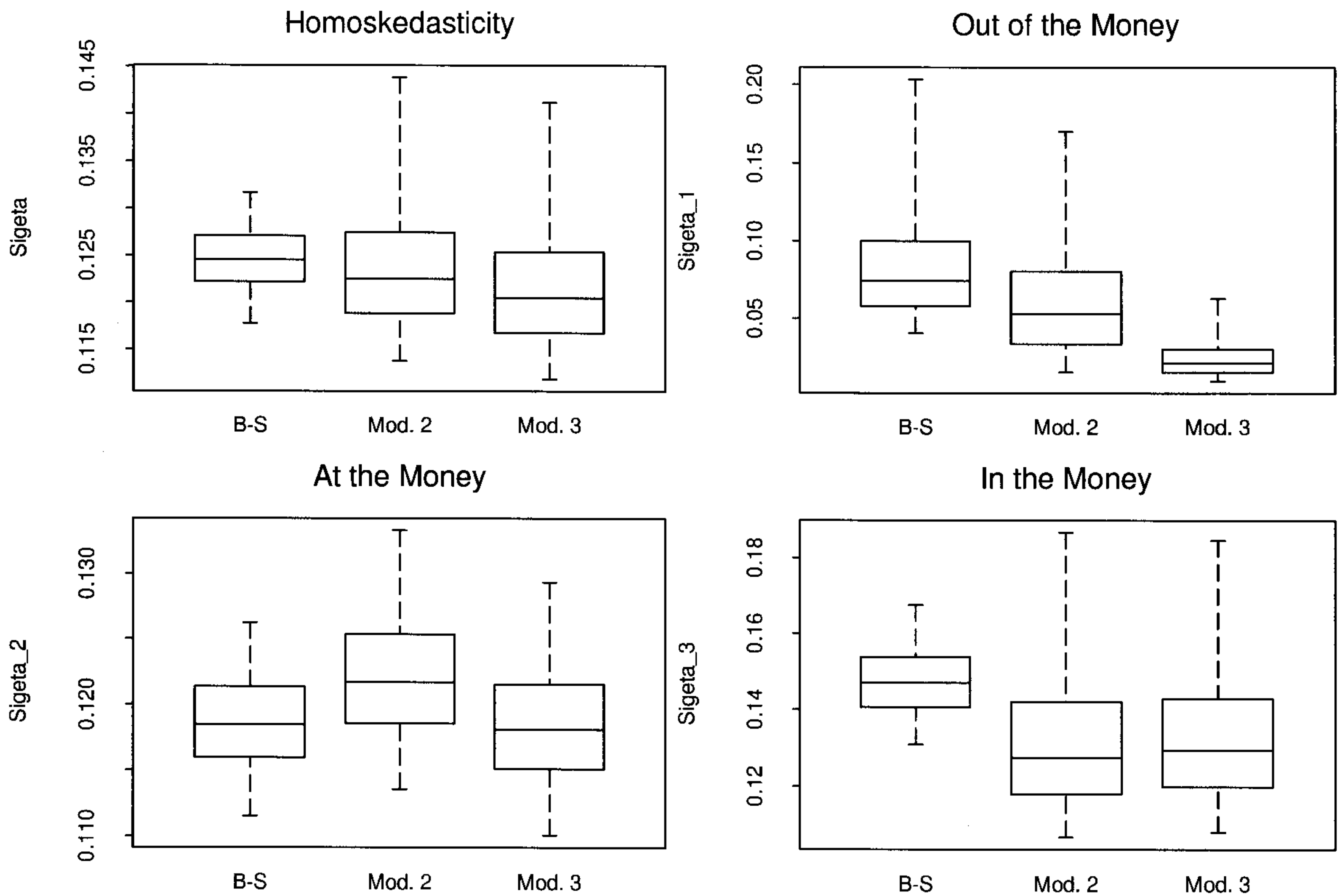


Figure 4a: Hedge Ratios for 3 Logarithms Models, TOYS'R US, Dec 04-08 1989: 452 quotes

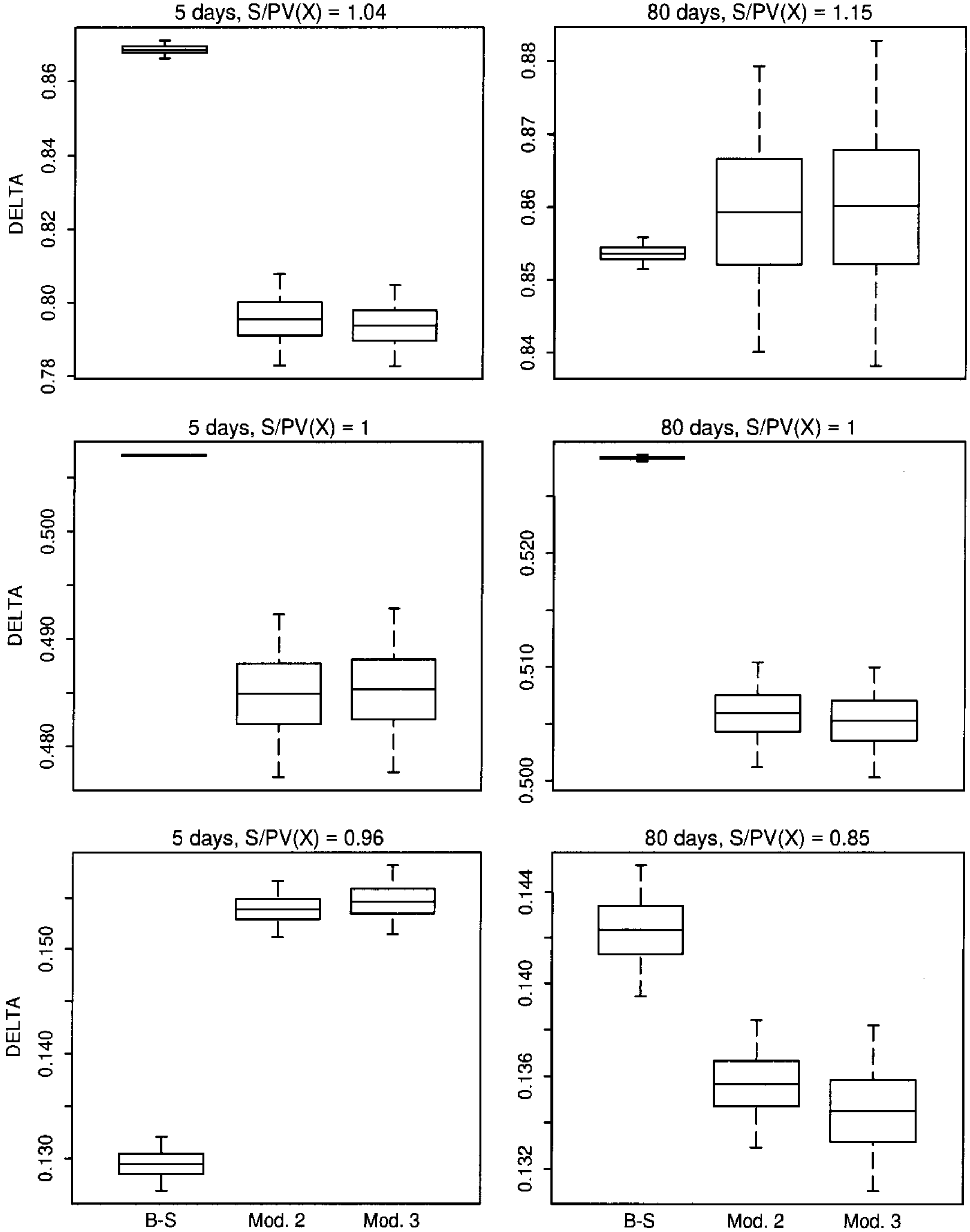


Figure 4b: Hedge Ratios for 3 Levels Models, TOYS'R US, Dec 04-08 1989: 452 quotes

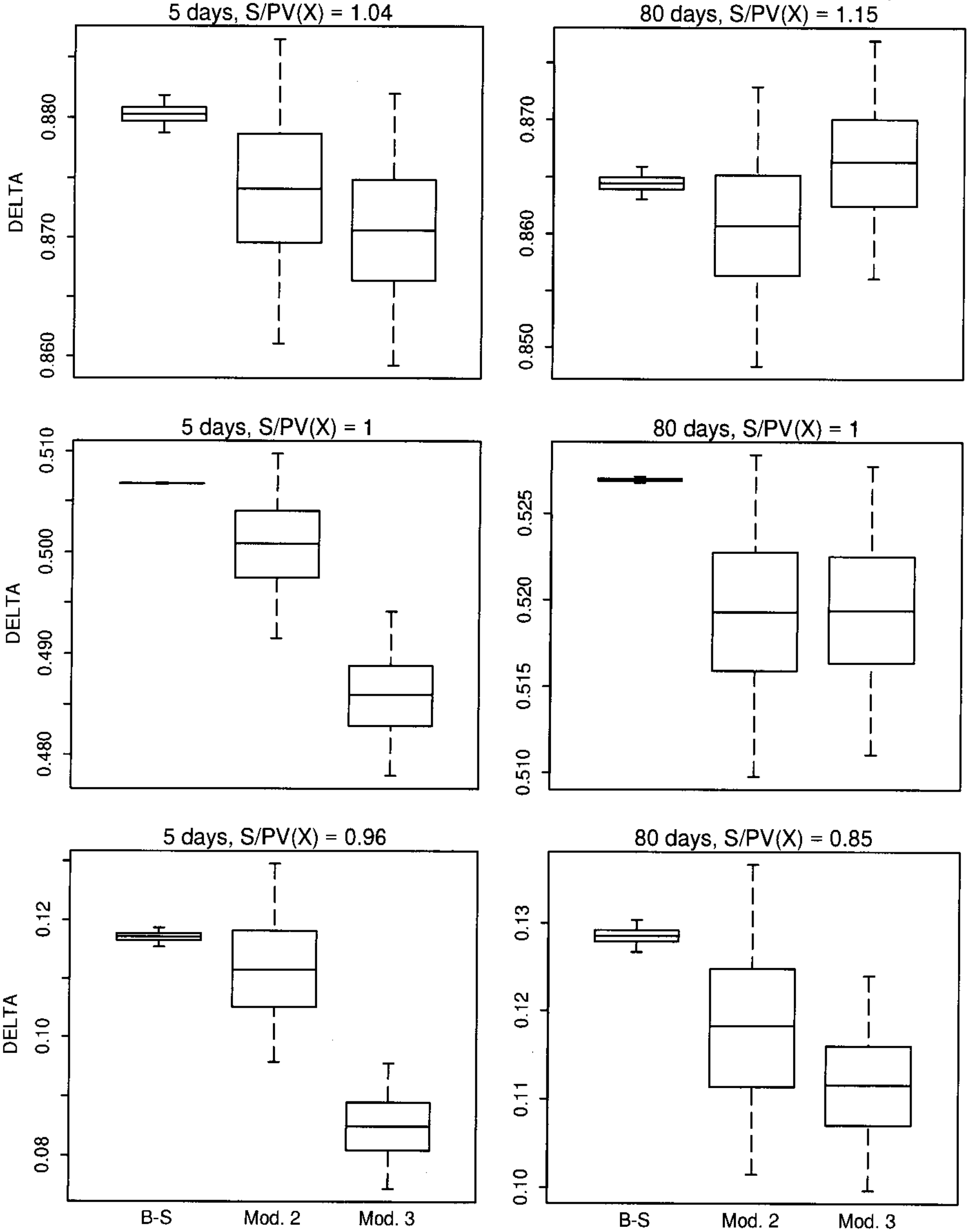


Figure 5: Money and Maturity Biases, Logarithms Models, TOYS'R US, Dec 11-15 89, 419 Quotes

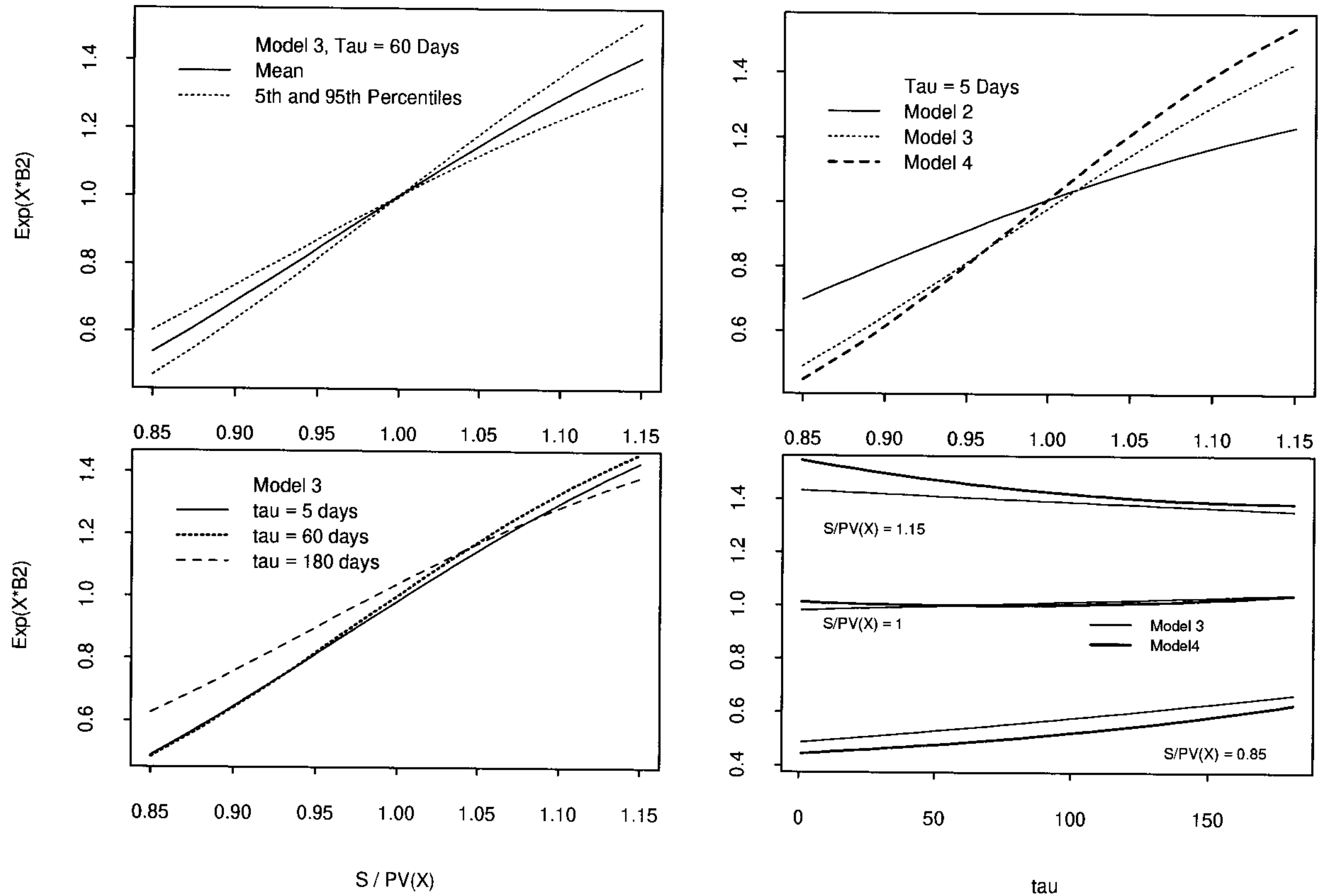
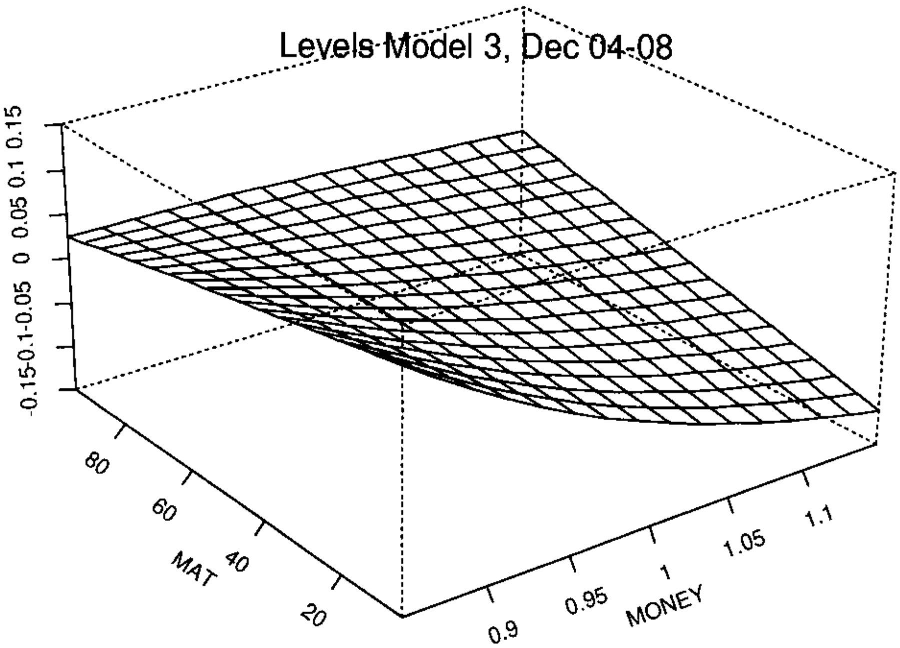


Figure 6: Bias Surfaces, TOYS'R US, Dec 89, 1927 Quotes



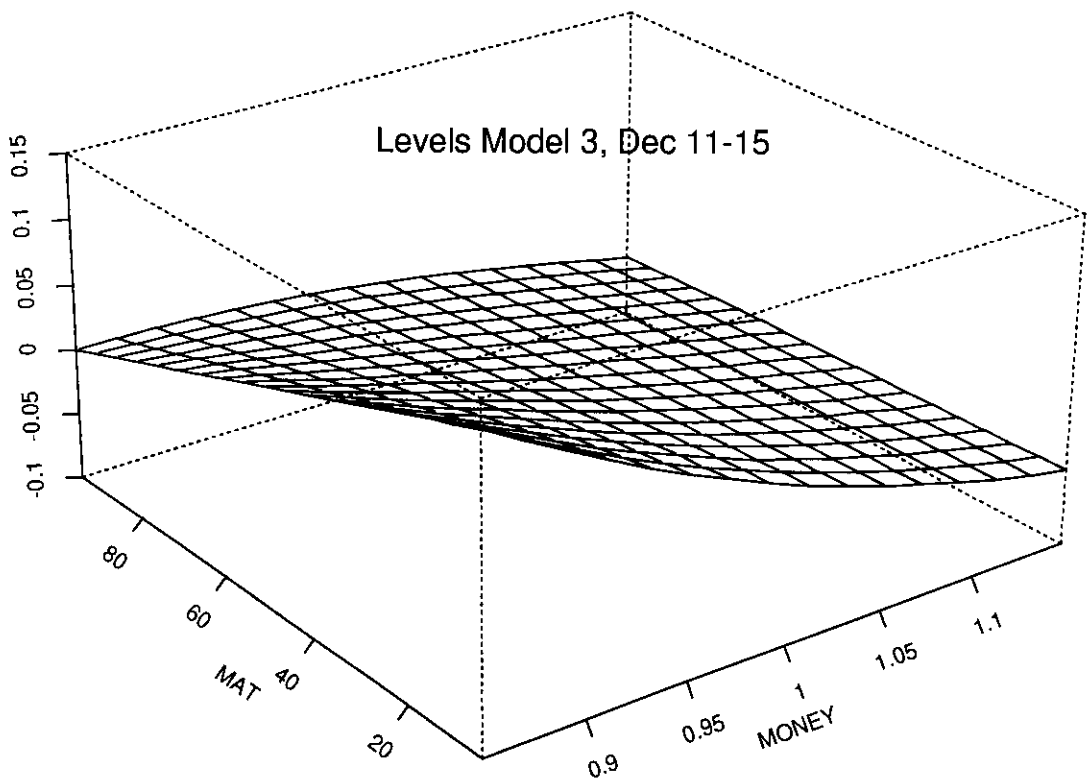


Figure 7: Time Series Plots of Parameters, Logarithms Models, TOYS'R US, January-March 1990

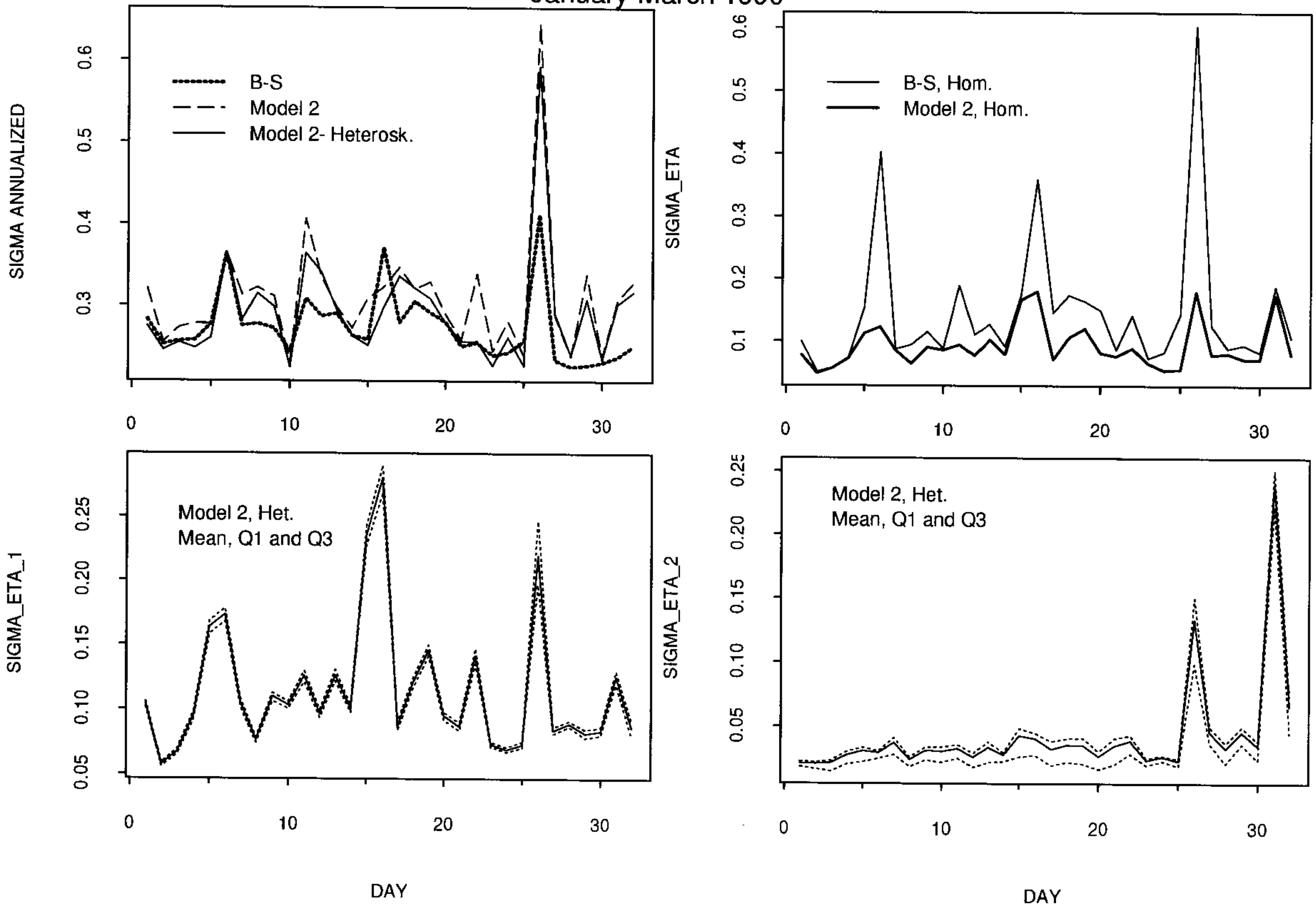
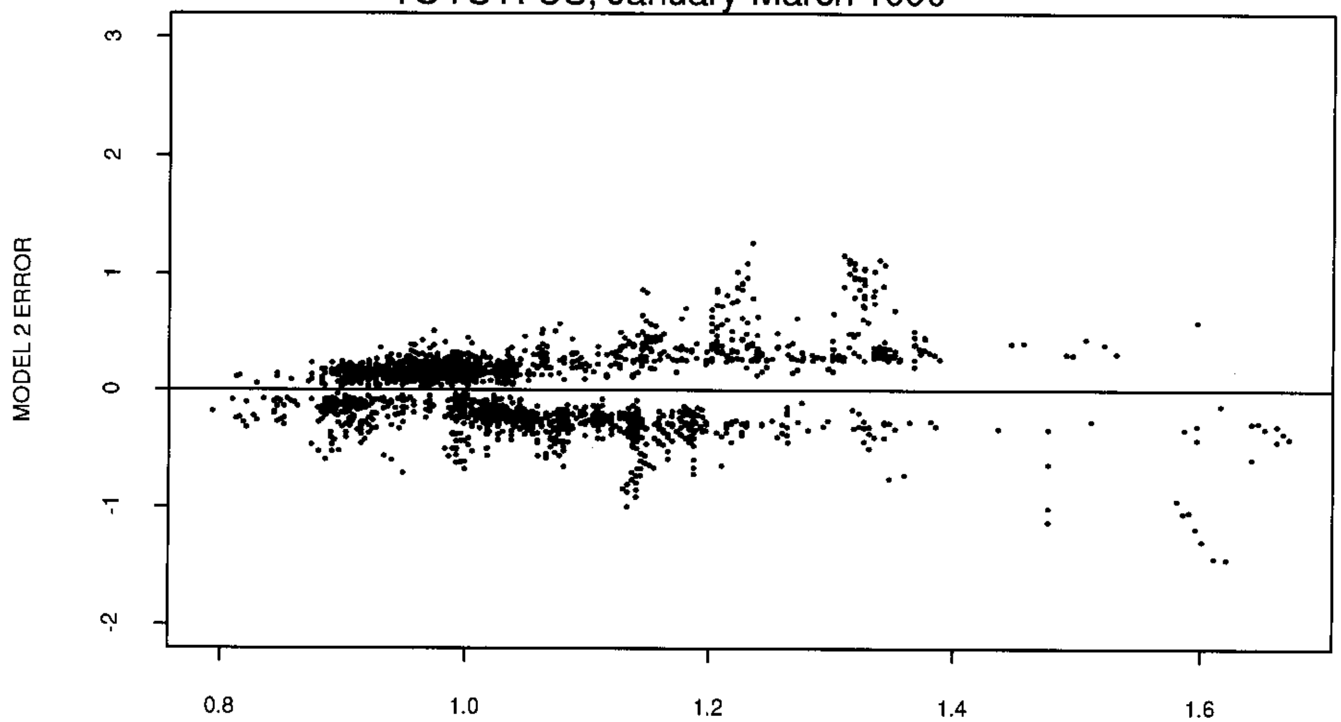


Figure 8: Pricing Error vs Moneyness, Logarithms Models, TOYS'R US, January-March 1990



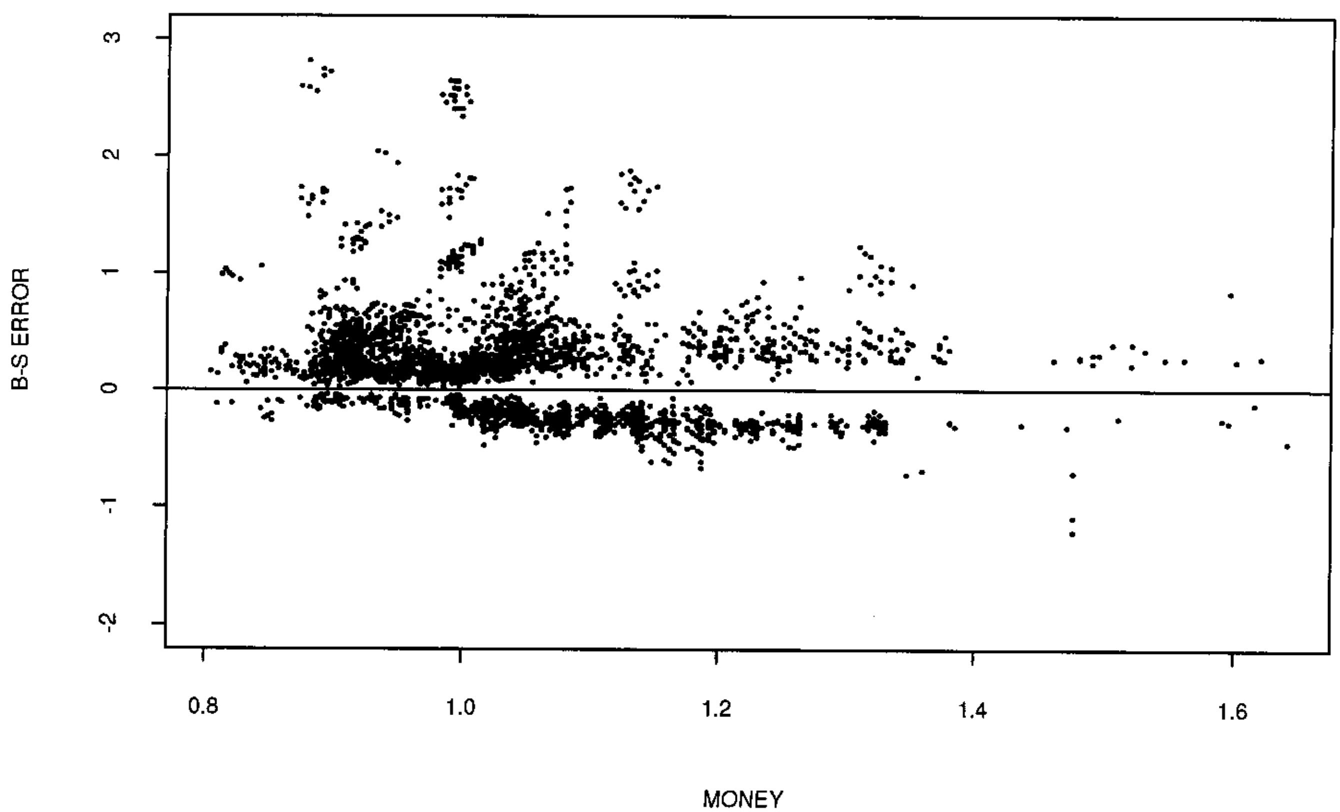


Table 1: In-Sample Performance Analysis TOYS'R US, Dec 4 to Dec 15, 1989[1]

Residual Analysis

Model	BIAS all	BIAS oom	BIAS out BA	RMSE all	RMSE oom	RMSE im	RMSE out BA
Log-Hom							
B-S	-0.018	-0.04	-0.05	0.10	0.18	0.02	0.15
2	0	-0.01	0.001	0.07	0.12	0.02	0.12
3	-0.	-0.01	0.001	0.066	0.11	0.02	0.12
4	0.002	-0.004	0.006	0.065	0.11	0.02	0.07
Log-Het		im					
B-S	-0.012	-0.006	-0.028	0.10	0.18	0.017	0.16
2	-0.005	-0.001	-0.017	0.07	0.12	0.015	0.13
3	-0.004	-0.002	-0.013	0.07	0.12	0.016	0.12
4	0	0.007	0.002	0.07	0.11	0.020	0.10
Lev-Het		im					
B-S	-0.013	-0.047	0.005	0.12	0.092	0.144	0.19
2	-0.002	0.007	0.012	0.11	0.087	0.126	0.19
3	0	0.009	0.017	0.11	0.081	0.122	0.19
4	0.005	0.031	0.014	0.11	0.078	0.134	0.21

Pricing Analysis

Model	BIAS all	BIAS oom	BIAS im	BIAS out BA	RMSE all	RMSE oom	RMSE im	RMSE out BA
Log-Hom								
B-S	-0.07	-0.11	-0.06	-0.16	0.16	0.17	0.17	0.24
2	0.001	-0.004	-0.01	0.003	0.12	0.09	0.14	0.20
3	0.	-0.01	-0.025	0.006	0.12	0.084	0.15	0.20
4*	-8.	-0.01	-26	-29	49	0.092	85	60
Log-Het								
B-S	-0.07	-0.10	-0.06	-0.13	0.144	0.16	0.145	0.22
2	-0.01	-0.03	-0.007	-0.009	0.11	0.098	0.128	0.19
3	-0.01	-0.03	-0.016	0.001	0.11	0.09	0.13	0.18
4	-0.12	-0.02	-0.36	-0.34	0.39	0.085	0.68	0.49

Table1 - continued

Distribution Analysis

Model	% Pred. out BA	Fit Cover	Pred Cover[2]	B1[3]	Cm<B2	A<B2	Q1<B2|B>B2
Lev-Het							
B-S	24	2	50	4	20	3	0.4
2	18	13	52	3	20	3	0.5
3	18	15	51	0.3	20	3	0.6
4	15	31	64	0.7	20	3	3
Log-Het							
B-S	29	2	58	na	-	-	0.4
2	18	21	67	na	-	-	1.1
3	19	24	68	na	-	-	2
4	25	36	78	na	-	-	10
Log-Hom							
B-S	32	1	71	na	-	-	6
2	19	24	74	na	-	-	6
3	19	28	75	na	-	-	6
4	63	41	85	na	-	-	7

[1]The models have been estimated over the week of Dec. 4-8, and reestimated for the Dec. 11-15 week. The errors have then been aggregated. Symbols used: all: all quotes used, oom: out of the money quotes, im: in the money quotes, out BA: quotes where the mean prediction is outside the Bid-Ask spread, B: Bid, A: Ask, B1,B2: intrinsic lower bounds on call price.

[2]Percentage of the observations for which the interquartile range of fit or prediction covers the true value.

[3]B1:Percentage of observations such that Prob(Pred<0) >0.001. B2 is the other intrinsic bound, S-PV(X). The next columns show the percentage of market prices, ask prices, and first quartile of predictive density violating B2.

Table 2: Out-of-Sample Performance Analysis TOYS'R US, Dec 4 to Dec 15, 89[1]

Residual Analysis

Model	BIAS all	BIAS oom	BIAS out BA	RMSE all	RMSE oom	RMSE im	RMSE out BA	RMSE 1 day ahead all	RMSE 1 day ahead oom	RMSE 1 day ahead im
Log-Hom										
B-S	-0.006		-0.009	0.10	0.18	0.02	0.15			
2	0.007		0.017	0.11	0.23	0.02	0.18			
3	0.005		0.021	0.12	0.24	0.02	0.18			
Log-Het										
B-S	-0.001	0.038	0.005	0.1	0.20	0.02	0.16	0.12	0.23	0.02
2	0.003	0.023	0.003	0.1	0.21	0.02	0.17	0.10	0.19	0.02
3	0.003	0.016	0.009	0.11	0.22	0.02	0.18	0.10	0.19	0.02
Lev-Het										
B-S	0.02	0.08	0.10	0.15	0.14	0.14	0.23			
2	0.03	0.05	0.09	0.15	0.13	0.14	0.22			
3	0.03	0.05	0.09	0.15	0.14	0.14	0.23			

Pricing Analysis

Model	BIAS all	BIAS oom	BIAS im	BIAS out BA	RMSE all	RMSE oom	RMSE im	RMSE out BA
Log-Hom								
B-S	-0.06	-0.02	-0.08	-0.07	0.15	0.11	0.16	0.22
2	-0.00	0.05	-0.05	0.01	0.16	0.14	0.17	0.23
3	-0.01	0.05	-0.08	0.02	0.16	0.13	0.17	0.23
Log-Het								
B-S	-0.041	-0.013	-0.057	-0.05	0.15	0.12	0.15	0.22
2	-0.005	0.040	-0.035	0.004	0.14	0.13	0.15	0.21
3	-0.005	0.036	-0.050	0.015	0.14	0.12	0.15	0.22

Table2 - Continued

Distribution Analysis

Model	% Pred. out BA	Fit	IQR cover[2] Pred: all	im	oom	B1[3]	Cm<B2	A<B2	Q1<B2 | B>B2
Lev-Het									
B-S	29	-	41	46	39	2.2	8	0.9	0.4
2	30	-	39	51	35	1.2	8	0.9	0.2
3	31	-	39	53	35	0.6	8	0.9	0.4
Log-Het									
B-S	31	1.6	54	64	39	na	8	0.9	0.7
2	31	15	52	75	31	na	8	0.9	0.6
3	29	19	52	78	33	na	8	0.9	0.6
Log-Hom									
B-S	31	-	69	98	35	na	8	0.9	5
2	36	-	59	97	24	na	8	0.9	3
3	34	-	58	97	24	na	8	0.9	2

[1]Models estimated as in table 1. The out of sample statistics are computed over the week following the estimation. all: all quotes used, oom: out of the money quotes, im: in the money quotes, out BA: quotes where the mean prediction is outside the Bid-Ask spread. B: Bid, B1,B2: intrinsic lower bounds on call price.

[2]Percentage of the observations for which the predictive interquartile range covers the true value.

[3]B1:Percentage of observations such that Prob(Pred<0) >0.001. B2 is the other intrinsic bound, S-PV(X). The next columns show the percentage of market prices, ask prices, and first quartile of predictive density violating B2.