BAYESIAN ANALYSIS OF CONTINGENT CLAIM MODEL ERROR

Eric Jacquier[*]
Boston College and CIRANO
Robert Jarrow
Cornell University and Kamakura Corporation

This paper formally incorporates parameter uncertainty and model error into the implementation of contingent claim models. We make hypotheses for the distribution of errors to allow the use of likelihood based estimators consistent with parameter uncertainty and model error. We then write a Bayesian estimator which does not rely on large sample properties but allows exact inference on the relevant functions of the parameters (option value, hedge ratios) and forecasts. This is crucial because the common practice of frequently updating the model parameters leads to small samples. Even for simple error structures and the Black-Scholes model, the Bayesian estimator does not have an analytical solution. Markov Chain Monte Carlo estimators help solve this problem. We show how they extend to some generalizations of the error structure. We apply these estimators to the Black-Scholes. Given recent work using non-parametric function to price options, we nest the B-S in a polynomial expansion of its inputs. Despite improved in-sample fit, the expansions do not yield any out-of-sample improvement over the B-S. Also, the out-of-sample errors, though larger than in-sample, are of the same magnitude. This contrasts with the performance of the popular implied tree methods which produce outstanding in-sample but disastrous out-of-sample fit as Dumas, Fleming and Whaley (1998) show. This means that the estimation method is as crucial as the model itself.

JEL Classification: C1, C11, C15, G1

1 Introduction

Since the original Black and Scholes (1973) and Merton (1973) papers, the theory of option pricing has advanced considerably. The Black-Scholes (hereafter B-S) has been extended to allow for stochastic volatility and jumps. The newer equivalent martingale technique allows to solve complex models more simply. At the same time it appears that the finance literature has not spent as much energy on the effect of the econometric method used to estimate the models. [2] Whaley (1982) used several options to compute the B-S implied volatility by minimizing a sum of squared pricing errors. The non linear least squares method is still the most often used in empirical work. The method of moments is sometimes used, e.g., Bossaerts and Hillion (1994a), Acharya and Madan (1995). Giannetti and Jacquier (1998) report potential problems with the asymptotic approximation. Rubinstein (1994)'s paper started a strain of non parametric empirical work with aim to retrieve the risk neutral pricing density implied by option prices. This literature is reluctant to assume a model error which is inconsistent with the no-arbitrage deterministic models within which it works. The results in Dumas, Fleming and Whaley (1998), hereafter DFW, show that this can lead to catastrophic out-of-sample performance even if, or maybe because the in-sample fit is quasi perfect.

Absolute consistency with the no arbitrage framework is not consistent with the data. Deterministic pricing models ignore market frictions and institutional features too hard to model. The over-fitted implied density is affected by model error. There is no tool to asses the effect of model error on the estimate. The results of DFW are consistent with an over-fitting scenario: The B-S models is too restrictive to suffer from over-fitting. So it performs much better out-of-sample than the more general non parametric models. Maybe DFW did not test the models as much as the estimation method.

We develop an estimator to allow its user to assess the effect of model error on the inference. It allows specification tests. The cost of this is that we must make explicit hypotheses on the pricing error distribution. We can then write the likelihood function of the parameters. [3] Like the existing methods, non linear least squares and methods of moments, the maximum likelihood estimator only has an asymptotic justification. We develop a Bayesian estimator which is also justified in small sample. This is useful for the small samples typical due to the common practice of updating model parameters daily.

The Bayesian estimator does not have an analytical solution even for the B-S and simple pricing error structures. We solve this problem by constructing a hierarchical Markov Chain Monte Carlo estimator (hereafter MCMC). This simulation based estimator provides the second improvement over standard methods. It delivers exact inference for any non linear function of the parameters instead of relying on the usual combination of delta method and normality assumption. The simulation based estimator produces draws of the desired posterior or predictive densities, e.g., option price , hedge ratio, correlation structure of the errors, all non linear in the parameters. [4] Before taking a position, an agent wants to assess if a quote is *far* from the price predicted by a model. For

---

[2]See Campbell, Lo and MacKinlay (1997) for a review of the option pricing literature, Bates (1995) for an empirical review, and Renault (1995) for a discussion of the econometric literature.

[3]We are not aware of likelihood based option pricing estimation. But some papers do add an explicit error to deterministic pricing models. Justifications often include measurement errors. For example, Brown and Dybvig (1986) add an error to the CIR term structure model.

[4]For a basic B-S model, the *parameters* is the underlying standard deviation.

this type of tail based diagnostic, it is crucial to use an exact rather than an approximate density. This is because predictive and posterior densities are often non normal.

The Bayesian estimator allows for prior information on the parameters, which standard methods do not. This is useful for several reasons. First, the underlying time series can be used as prior information on the parameters. Second, imperfect models often result in time varying parameters and need to be implemented in a setup allowing for periodic re-estimation. In the resulting small samples, priors may be used to improve precision. The result of a previous sample can serve as prior information for the next sample. Finally, priors can help resolve potential multicollinearities when nesting competing models. We also extend the MCMC estimator to heteroskedastic pricing errors and intermittent mispricing where the error variance is sometimes larger than usual. The latter extension is more than a way to model fat tail errors. It parameterizes a quantity of economic interest to the user, the probability of a quote being an outlier. This is in line with common views where models are thought to work most of the time, and some quotes may occasionally be out of equilibrium and include a *market* error.

A key here is that we put a distribution on the errors. Let us see what this allows us to do which is not done in the empirical finance literature. Usually, once a point estimate for the parameters is obtained by non linear least squares, model prices are computed by substitution into the price formula. These price estimates are then compared to market prices. Mispricings, e.g., smiles, are characterized. Competing models are compared on the basis of these price estimates and smiles. What is not produced is a specification test based upon confidence intervals around the price reflecting parameter uncertainty and model error. Quotes, whether in or out of sample, should be within this interval with the expected frequency. Without further distributional assumptions, the distribution of the model error in a method of moment or least squares estimation is not clear. Standard methods can, but this is not done in the literature, use the asymptotic distribution of the parameter estimator to get an approximate density for the model price reflecting parameter uncertainty alone. Coverage tests could follow. First, this (necessary) approximation can lead to flawed inference. Second, it does not allow the incorporation of the pricing error to form a predictive rather than a fit density . We produce both fit and predictive densities for option prices and document their behavior. The extent to which they produce different results has not been documented.

Of course, the modelling of pricing errors arises only if the data include the option prices. We state the obvious to contrast the estimation of model parameters from option prices and from time series of the underlying. For models with risk premia, the time series may not give information on all the parameters of the model. Also, even for simpler models, the time series and the option prices may yield different inference. Lo (1986) computes coverage intervals and performs B-S specification tests. He uses the asymptotic distribution of the time series variance estimator to generate a sampling variability in option prices. The estimation has not used the information in option prices to infer the uncertainty on the parameter or the model error. For a model with constant variance, this is a problem since the length of the time series affects the precision of the estimator of variance. In the limit, a long sample implies very high precision and a short one implies to low precision, with the corresponding effect on the B-S confidence intervals. This test is hard to interpret.[5] Our method can incorporate the underlying historical information if desired, but the core of the information is the option data.

---

[5]Jacquier and Polson (1995) propose to construct efficient option price predictors that reflect the uncertainty in the forecast of stochastic volatility. Stochastic volatility forecasts are not degenerate even in large sample.

One does not expect any model to have no errors. Rather, an acceptable model would admit errors which: (1) are as small as possible, (2) are unrelated to observable model inputs or variables, and (3) have distributional properties preserved out-of-sample. (1) and (2) say that the model uses the information set as best as possible. (3) says that the inference can be trusted out-of-sample. Our estimator can produce small sample diagnostics of these criteria. We document its behavior for the B-S model. We well know that the B-S does not fare well at least by criteria (1) and (2) above. DFW's results seem to imply that it fares better than some more complex models by criterion (3). There are known reasons for the B-S sytematic mispricings. The process of the underlying asset may have a stochastic volatility, or jumps. The inability to transact continuously affects option prices differently depending on their moneyness. Time varying hedging demand can also cause systematic mispricings. Practitioners' common practice of gamma and vega hedging with the Black-Scholes reveals their awareness of model error.

It also shows their reluctance to use the more complex models available. Despite known problems, the B-S is still by far the most used option pricing model. One reason is that practitioners see more complex models as costly or risky. They may not have an intuition on their behavior. The estimation of the parameters can be complex, lead to unfamiliar hedge ratios or forecasts. Jarrow and Rudd (1982) argue that in such cases, auxiliary models such as a polynomial expansions of known inputs may be useful extensions to a basic model. Their argument could be relevant beyond the B-S case. At any point, there is a better understood basic model - the status-quo, and more complex models entertained which are harder to learn and implement. To improve upon the current basic model while the more complex model is not yet understood, the expansions need to capture the patterns omitted by the basic model. Recently, Hutchinson, Lo and Poggio (1993) showed that non parametric forms can successfully recover existing B-S patterns. [6] We add an extension of the B-S in this spirit and incorporate it to our estimator. We essentially will test if the expansions, similar to the larger models of DFW, capture the systematic patterns omitted by the B-S.

In the empirical analysis, we document the use of the estimator. We show the non-normality of the posterior densities. We then show that tests of the B-S model (and its expansions) which only account for parameter uncertainty (fit density) do not give a reliable view of the pricing uncertainty to be expected out-of-sample. The use of the predictive density markedly improves the quality of the forecasts. The fit density is of course tighter than the predictive density. It leads to massive rejection of all the models used. Its use by a practitioner would lead to an overestimation of the precision of a model price. We show that the non parametric extended models have different in-sample implications than the simple Black-Scholes. They also improve the model specification. However we also show that these improvements do not survive out-of-sample.

Section 2 introduces the *basic* model b(.) and its *extended* counterpart. It also introduces two candidate error structures, multiplicative (logarithm) and additive (level), both possibly heteroskedastic. Section 3 discusses the estimators and details the implementation of the method in the case of both *basic* and *extended* Black-Scholes models. [7] Section 4 shows the results of the application of the method to stock options data. Section 5 concludes and discusses areas of future research.

---

[6] For a survey of the recent non parametric literature, see Ghysels et al. (1997).

[7] Technical issues and proofs are discussed in an appendix available from the authors upon request. jacquier@jacquier.bc.edu

2 Models and Errors

2.1 Basic model, additive and multiplicative forms

Consider N observations of a contingent claim's market price, C_i, for $i \in \{1,2,...,N\}$. We think of C_i as a limited liability derivative, like a call or put option. Formally, we can assume that there exists an unobservable equilibrium or arbitrage free price c_i for each observation. Then the observed price C_i should be equal to the equilibrium price c_i. There is a basic model $b(x_{1i}, \theta)$ for the equilibrium price c_i. The model depends on vectors of observables $x_{1,i}$, and parameters θ. We assume that the parameters are constant over the sample span. The model is an approximation, even though it was theoretically derived as being exact. There is an unobservable pricing error, η_i. A quote C_i may also sometimes depart from equilibrium. The error then has a second component ϵ_i, which can be thought of as a market error. ϵ_i and η_i are not identified without further assumptions. In the empirical analysis contained in this paper, we merge these two errors into one common pricing error η_i. In section 5, we propose an error structure to better identify outlying quotes which may originate from intermittent mispricing. Formally,

$$\log C_i \equiv \log b_i(x_{1i}, \theta) + \eta_i, \tag{1}$$

This implies a multiplicative error structure on the level, which guarantees the positivity of the call price for any error distribution.

The introduction of a non-zero error η_i is justified. First, simplifying assumptions on the structure of trading or the underlying stochastic process made to derive tractable models. They result in errors, possibly biased and non i.i.d. For example, Renault and Touzi (1994) 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 recording of underlying and option prices can measurement can cause skewed Black-Scholes implied volatility smiles. Bakshi, Cao, and Chen (1998) show that adding jumps to a basic stochastic volatility process further improves pricing performance. Bossaerts and Hillion (1994b) show that the assumption of continuous trading also leads to smiles while Platen and Schweizer (1995)'s hedging model causes 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, in typical models, the rational agents are unaware of market or model error and know the parameters of the model. Such models could be biased in the "larger system" consisting of expression (1). [8]

The error in (1)is multiplicative. This could be preferred to an additive structure for two reasons. First, it insures the positivity of C_i, independently from the distribution of η_i. This is the first Merton lower bound, hereafter B_1. [9] Second, it models relative rather than absolute errors. This insures 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 measure of model risk. The difference between multiplicative and additive errors is an unexplored empirical question.

---

[8]Clément, Gouriéroux, and Montfort (1993) address model uncertainty by randomizing the equivalent martingale measure. They show that this may induce non i.i.d. pricing errors related to the model inputs. Our goal here is more to implement and test known models.

[9]This does not address the second Merton lower bound, B_2: A call option must exceed the stock price less the present value of the strike price, S- PV(X).

So we also implement the simple additive (level) form:

$$C_i = b_i(x_{1i}, \theta) + \eta_i, \tag{2}$$

For each formulation, we will document how acute the problem of Merton bound violations are.

2.2 Non parametric extended models

For many *basic* models, say $b_i(x_{1i}, \theta)$, the pricing error η_i is not i.i.d. We introduce an *extended* model m_i such that:

$$\log C_i = \beta_1 \log b_i(x_{1i}, \theta) + \beta_2' x_{2i} + \eta_i \equiv \log m_i + \eta_i. \tag{3}$$

Model m_i differs from the basic model b_i by the addition of a coefficient β_1 and linear combination $\beta_2' x_{2i}$. The variables x_{2i} may include functions of the observables x_{1i}, of other relevant variables and an intercept term. They can also include one or several competing models to b_i. Then the *extended* model equation allows an estimation procedure for nesting competing models. [10] The extended model is intended to capture the biases in the basic model. Jarrow and Rudd (1982) argue that it may be justified as a costless and intuitive non parametric approximation of a more general model when the more general model is either unknown or too costly to implement. [11] This is also in the spirit of the non parametric literature such as Hutchinson, Lo, and Poggio (1994) who show that simple non parametric models can recover Black-Scholes prices. Here we will check if they can recover patterns in addition to the Black-Scholes in actual prices. Under the simple additive error form, the extended model is given by

$$C_i = \beta_1 b_i(x_{1i}, \theta) + \beta_2' x_{2i} + \eta_i \equiv m_i + \eta_i. \tag{4}$$

The basic model $b(\sigma, x_{1i})$ is the Black-Scholes. So x_{1i} includes the stock price S_i, the time to maturity τ_i, the relevant interest rate r_{f,τ_i}, and the exercise price X_i. We assume that $\eta_i \sim N(0, \sigma_\eta)$ for both logarithm and level models. The parameters θ include 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 z is the logarithm of $S/Xe^{-r_{f,\tau}}$, the ratio of the stock to the present value of the exercise price. Renault and Touzi (1994) show that under a stochastic volatility framework, the B-S implied volatility is parabolic in z, and decreasing in τ. The second variable is τ, the maturity in days. In the analysis, we will refer to the following models 0 to 4.

Insert Table 1 here

The logic 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

---

[10]Competing models are often 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 (3) may provide an inexpensive control of model error in hedging a trader's portfolio (book).

do not know the functional form of the better parametric model, it is unclear how far the expansion of the relevant variables needs to go. Model 4 will test the effectiveness of large expansions. One could also consider other variables, such as liquidity proxies, e.g., the bid-ask spread of the option or the stock.

2.3 Heteroskedasticity

The error in the level model is in dollars. Both quote and model value are small for far out of the money options. Deep in the money, they are both large. So one may expect a possible heteroskedasticity in the errors, with σ_η an increasing function of moneyness. A different argument can be made. The typical plot of the B-S value vs the stock value for a given maturity and exercise price shows that the distance between the Merton bound and the model value is smallest for far from the money options and largest at the money. This suggests that pricing errors may be more variable at the money than away from the money. Both arguments imply that some form of heteroskedasticity should be allowed. This could alleviate the potential Merton bound problem with the level model.

The error in the logarithm model is in relative terms. For a far out of the money call, the quote could be 12 cents and the model might be *centered* at 6 cents. The dollar pricing error is small but the relative error is large. Conversely, a large dollar error for a deep in the money option may imply a small relative error. So one may expect σ_η to be a decreasing function of moneyness in the logarithm model. A decreasing σ_η may also help alleviate the potential for predictive densities to violate the Merton bound, $B_2 : C > S - PV(X)$, still a potential problem in the logarithm model.

We extend our estimator to allow for heteroskedasticity and implement both homoskedastic and heteroskedastic errors. We model heteroskedasticity by simply allowing up to three different standard deviations $\sigma_{\eta j}, j = 1, 2, 3$ depending on the moneyness ratio which proxies for the magnitude of the call value. [12] A heteroskedastic model then has up to 2 more parameters than its homoskedastic counterpart. For example, model 2 - table 1, with three levels of pricing error variance, has nine, not seven parameters.

3 Estimation

3.1 Monte Carlo Estimation and Prediction

We now briefly outline the methodology for Monte Carlo estimation. The time-t information set consists of observations of the option prices $C_{s,i}$ collected for times s up to t, the vectors of relevant observable inputs $x_{\tau,i}$, and the history of the underlying asset price S_τ. We index quotes by t and i because for a time s, a cross-section of (quasi) simultaneous quotes can be collected. Let $\underline{y}_t$ be the vector of histories at t, the *data*. θ is the vector of all the relevant parameters including σ_η the standard deviation of the error η_i, and the coefficients β from expression (3) or (4). A prior density $p(\theta)$ is selected.

---

[12] We discuss the choice of the boundaries in section 4. An alternative is to model σ_η as a smooth function of moneyness.

Next, the contingent claim model and the distributional assumption on η_i yield the likelihood function $\ell(\theta \mid \underline{y}_t)$. Finally, by Bayes theorem, the posterior density of the parameters is $p(\theta \mid \underline{y}_t) \propto \ell(\theta \mid \underline{y}_t) p(\theta)$. [13] The specifics of the posterior and predictive distributions vary with the prior and likelihood functions and are discussed later. Even for the Black-Scholes, the posterior density of θ does not have an analytical solution. However a Monte Carlo estimator, i.e., a sample of random draws of the posterior density is feasible. Moments and quantiles are readily obtained from this sample with any desired precision. The main advantage of the Monte Carlo approach is that a draw of any deterministic function of θ is obtained by direct computation from each draw of θ. Unlike the delta method, this requires no approximation. We can for example generate samples of draws of the exact posterior distribution of the model value, b_i, m_i, or hedge ratios. The Monte Carlo approach removes the need to perform numerical integration as in standard Bayesian analysis.

An important deterministic function of the parameters is the residual for each observation. Each draw of θ implies a draw of the posterior distribution of η_i for each i by computation of: $\log C_i - \log m(x_i, \theta)$. This is the basis for residual analysis. [14] These residuals can be used for within sample tests of model specification, discussed later. Predictive densities, needed for predictive tests, are different. The predictive density of a quote C_f depends on the error $(\eta_f \mid \underline{y}_t)$.

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

Draws of this density are made and used as follows. Make one draw from $(\eta_f \mid \theta, \underline{y}_t)$ for each draw of $\theta \mid \underline{y}_t$. This yields a sample of joint draws of $(\eta_f, \theta \mid \underline{y}_t)$. The resulting draws of η_f are draws of $(\eta_f \mid \underline{y}_t)$ since they integrates out θ. Then for each draw, compute c_f as in (3) or 4). This yields a sample of draws of $(C_f \mid \underline{y}_t, x_f)$, the predictive density of the price. Quantiles of this density provide a model error based uncertainty around the point prediction, i.e. a probabilistic method for determining to where a market quote C_f lies in its relevant predictive density. We discuss predictive specification tests in sections 4.7 and 4.8

3.2 Markov Chain Algorithms

We now present the intuition behind the Markov Chain Monte Carlo (MCMC) algorithm needed here. The next section and the appendix contain more details.[15] Consider $p(\beta, \sigma_\eta \mid \sigma \mid .)$. Given σ, the model is a standard linear regression and direct draws of β, σ_η are available. The non linearity in σ makes it impossible to draw directly from the marginal posterior $p(\sigma \mid .)$.

This is partly because the integration constant of p does not have an analytical expression. The Metropolis algorithm solves that problem because it does not require the integration constant. [16] See Metropolis et al. (1953). The intuition of the Metropolis algorithm is as follows. First, select a

---

[13]Whenever possible, we will follow closely Zellner's notation. See Zellner (1971).

[14]See Chaloner and Brant (1988). The difference with standard residual analysis is that one obtains the exact distribution of the residual for each observation.

[15]Detailed derivations of all the posteriors and the likelihood function are available in a technical appendix upon request and on the first author's WEB page.

[16]This is essential for the 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 σ would require an optimization, each step of the optimization requiring a numerical integration.

blanketing density q from which one can draw directly. q should have a shape similar to p. Given a draw from q, the Metropolis algorithm is a probabilistic rule of acceptance of the draw which compensates for regions where q draws too often or not often enough. The resulting sample of draws of q converges in distribution to a sample of draws from p.

For the heteroskedastic extension where σ_η is a vector of parameters, we cannot write $p(\sigma \mid .)$, we can only write $p(\sigma \mid \sigma_\eta)$. The MCMC algorithm is adjusted by the use of a Gibbs cycle, see Geman and Geman (1984). It solves the following problem: Consider $(\sigma, \sigma_\eta$. We can not draw from their joint density. 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 density σ, σ_η. This algorithm extends to any number of conditionals. It is invariant to the initial values. This combination of these two algorithms constitutes the MCMC estimator which we use. See Jacquier, Polson and Rossi (1994), and Tierney (1994) for a recent discussion of the MCMC algorithms and convergence conditions. The draws converge in distribution to draws of the joint posterior, under very mild conditions. The key is that the conditions are often expressed in terms of the various transition kernels used by the algorithms. Even though the conditions are intuitive and mild, absolute proofs of convergence for all samples, on the basis of the likelihood and the priors only, is not easy . Some razor's edge cases could cause problems. For example, assume a proper prior and an unbounded parameter space. For a small sample the likelihood may be unbounded. Then, a small dataset with some outliers could have a very peaked and multimodal likelihood. If the (unimodal) prior is not strong enough, the algorithm could still get *stuck* around a local mode. However if the prior imposes compactness on the parameter space, the algorithm would converge even under multimodality. The issue of multimodality is addressed by Gelman and Rubin (1992ab). Polson and Roberts (1994) have convergence results for some families of distributions. The general consensus in the literature is that MCMC algorithms are very robust but care must be taken in the diagnostics. Kass et al. (1997) gives a good survey of the literature. We now specialize the MCMC estimator to the B-S model.

3.3 Application to the Black-Scholes Model

3.3.1 Priors and Posteriors

In the Black-Scholes economy, the underlying asset price S_t follows a lognormal distribution, i.e., $R_t = \log(S_t/S_{t-1}) \sim N(\mu, \sigma)$. In our application, the underlying asset is a stock. We assume that the stock price and the risk free rate $r_{f,i}$ are observed without error. We assume a zero correlation between the stock return and the model error. This is inocuous in our application because the stock returns collected predate the panels of option price data. If stock returns with a calendar overlap with the option price data were incorporated in the likelihood, the assumption of zero correlation might not hold under some market structure where trading in the underlying is not exogenous to trading in the option. [17] Here we view the stock return data as a source of prior distribution on σ. So the likelihood does not include the stock return data. [18]

---

[17]For this situation, allowing non zero correlation could provide a test of exogeneity of the underlying market. We do not explore this route here.

[18]Combining a historical kernel in the likelihood with a diffuse prior on σ gives the same result.

We use normal gamma priors for the parameters:

$$p(\sigma, \sigma_\eta, \beta) = p(\sigma)p(\sigma_\eta)p(\beta \mid \sigma_\eta) \sim \text{IG}(\sigma : \nu_0, s_0^2) \; \text{IG}(\sigma_\eta : \nu_1, s_1^2) \; \text{N}(\beta : \beta_0, \sigma_\eta^2 V_0), \qquad (6)$$

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. So they result in similar posteriors, see Zellner (1971). For σ, we use the inverted gamma prior because it is consistent with a posterior resulting from the time series of the stock returns. The priors can be made diffuse by increasing the variance V_0 and using small values of ν_0 and ν_1. They would still be proper.

Proper priors also allow the computation of odds ratios of desired. They also help model desired restrictions on the parameters. For example, one may want to center β_1 on 1 rather than zero, and concentrate it in the positive region.[19] One can also incorporate in the priors the Merton bounds, by truncating the priors to eliminate parameter values violating the bounds. With a Monte Carlo estimator, this is done by simply rejecting the posterior draws that violate the bounds. However, one can not guarantee that the predictive density will not violate the bound. We will diagnose this potential problem. Finally, when the sample is updated, the previous posterior distribution may be used as a basis for the next the prior. This is not as simple as the linear regression. Because of the non linearity in σ, one cannot find conjugate priors for all the parameters.

The joint posterior follows from Bayes theorem, see the appendix. We can not draw directly from it. Instead we can draw from $p(\sigma \mid \underline{y}_t)$ by a Metropolis step, then directly from $p(\sigma_\eta \mid \sigma, \underline{y}_t)$, and from $p(\beta \mid \sigma, \sigma_\eta, \underline{y}_t)$. In this case no Gibbs cycle is needed. A Gibbs step can be introduced by drawing from $p(\sigma \mid \beta, \sigma_\eta, \underline{y}_t), p(\sigma_\eta \mid \beta, \sigma, \underline{y}_t)$, and $p(\beta \mid \sigma, \sigma_\eta, \underline{y}_t)$ respectively. These conditional densities obtain readily from inspection of the joint posterior density. With heteroskedastic errors, we will need to use a Gibbs cycle. This is because we can not write the kernel of $p(\sigma \mid \underline{y}_t)$ when σ_η is a **vector** of error standard deviations and not a scalar. See Appendix A3 for the heteroskedastic model.

3.3.2   Posterior Distribution of σ

We now discuss the Metropolis step for σ. Appendix A2 contains the implementation details. Consider the conditional posterior distribution of σ

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

where $\nu s^2(\sigma, \beta)$ is a function of σ, β and x_i. This is the distribution p used when a Gibbs cycle is implemented. A draw of σ is made as follows. First, select a (blanketing) distribution q with shape reasonably close to p, from which one can draw directly. We do not need to know the normalization constant of p or the cumulative distribution function of σ. Call p⋆ the kernel of p above. Second make a draw from q. For this draw, we know p⋆/q The algorithm is a probabilistic rule with three

---

[19]Also, a zero intercept in the log model leads to a biased forecast for the Call price. This is due to the term 0.5 σ_η^2 in the mean of the lognormal distribution. One may then want to center the intercept on -0.5 σ_η^2 for an unbiased log model. The effect is small for typical parameter values.

possible outcomes. First, the previous draw is 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 value of the ratio p⋆/q at the candidate and the previous draws is used to compute the probability of each outcome.

The closer the shapes of q and p are, the faster the algorithm generates informative draws on σ. A quantity c chosen as p⋆/q computed for various values of σ is used to balance the number of rejections or repeats (see appendix A2). For a choice of q and c, a plot of the ratio p⋆/cq helps assess the effectiveness of the algorithm. The more p⋆ looks like q, the flatter the ratio curve. For q, we choose a truncated normal with mean the mode of p(σ). A number of diagnostics are available to decide how many *burn in* draws to eliminate, and how fast the algorithm generates information. We discuss these tools in section 4.2.

4   Empirical Application

4.1   Data

The options data come from the Berkeley database. Quotes for call options on the stock TOYS'R US, are obtained from December 89 to March 90. We use TOY for two reasons. First, it does not pay dividends so a European model is appropriate. Second, TOY is an actively traded stock on the NYSE. We define the market price as the average of the Bid and the Ask. Quotes are always given in simultaneous pairs by the market makers. We filter out quotes with zero bids, and quotes which relative spread is larger than the price. We use quotes rather than trades because it results in more precise estimation Trades in this market happen more often at the prevailing ask or bid than inside the spread. This causes additional error.[20]

There are between 80 and 300 daily quotes on TOY calls. We collect all quotes whatever their maturity and moneyness because we want a global model diagnostic, and we analyze extended models including bias functions. These models need rich panels to be identifiable. Table 2 summarizes the data used.

Insert Table 2 here

We estimate the models over several sub-samples of these four months, to illustrate our results under various setups. We use a one day sample to document small sample performance and convergence of the algorithm. One week samples are compromises between the need for a short calendar span and the need to accumulate information for the non-parametric extensions. When the samples cover longer calendar periods, the potential autocorrelation of the errors may affect the estimation. An estimator which does not take this into account may misrepresent the variability of the parameters. Month long and bi-daily re-estimated samples help examine the robustness of the results and the time variation of the parameters.

---

[20]We verified this by examining the trades and quotes. See George and Longstaff (1993) for the same conclusion on the SP100 market. We previously estimated the models on trades with resulting performance dramatically worse than presented here.

4.2 Convergence of the Algorithm

Here we estimate a level model 3 with 2 levels of σ_η: $\sigma_{\eta,1}$ and $\sigma_{\eta,2}$ for moneyness ratios below 1 and above 1. It is a 10 parameter model. We use the 140 quotes from Dec 1, 1989. The prior on σ is flat. We choose one (arbitrary) day and a relatively large model to document the applicability of the algorithm in a small samples. One needs to monitor the draws. Figure 1 documents some of the diagnostics used in the implementation of a MCMC algorithm.

When can we assume to have converged? A first tool is the time series plot of the draws, top plot in Figure 1 shown for $\sigma_{\eta,2}$ on the top right plot. We intentionally start the chain from several unrealistic values to check how quickly the draws settle down to a constant regime for **all** the parameters. Here it took less than 10 draws for the system to settle down. The result was similar for the other parameters and different starting values. If the starting value have a lasting effect on the initial draws, e.g., multimodality, Gelman and Rubin (1992ab) argue that diagnostics based on a single run can be misleading. 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 are stable , the process has converged. The boxplots in Figure 1 confirm that the series has converged for this run.

The next question is: How fast does information accumulate? The most intuitive check is the autocorrelation function of the draws past the burn-in period. The bottom plot of Figure 1 shows that the autocorrelations die out quickly. So the sequence of remaining draws is stationary. That together with the fact that we checked multiple starting values confirms that the algorithm converges. The autocorrelations are low. This means that information accumulates nearly as fast as for a simple iid sequence. One can assess the precision of a sample quantile estimate given the sample size and the degree of non *i.i.d.-ness* of the algorithm. Raftery and Lewis (1992) propose such a diagnostic based solely on the sequence of draws. Our results will be based on 3500 parameter draws.

4.3 The Parameters: β_0, β_1, σ

For the same model 3 and data as above, Figure 2 shows the posterior distributions of σ and $\sigma_{\eta,2}$. The left plot indicates that the distribution of σ does not depart too much from normality. However the right plot shows that $\sigma_{\eta,2}$ has a very non normal distribution. So, for a small sample, asymptotic inference based upon a normal approximation can be flawed. This itself is a reason to favor a small sample estimator.

We now ask whether different models imply different values for their common parameters. This question with respect to the volatility σ is crucial because: (1) a visible empirical literature uses implied volatilities to analyze the informational efficiency of options markets, and (2) practitioners routinely back out implied volatility from the basic Black-Scholes model. We now use the first complete week in December, i.e. 452 quotes generated between December 4 and 8. A week is a compromise calendar span. It allows us to collect more information on the parameters of the non parametric extensions. More than the sample size itself, the variation in the stock level generates a richer pattern of moneyness. For a pure cross-section, where the stock price does not move, models such as 3 or 4 may be difficult to identify. The cost of a longer window is that over five days, time

11

variation in volatility may cause the models to be more misspecified and the errors to be larger and predictable. Panel B in Table 2 shows the cross-correlations of the expansion variables over the week. Those involving the powers of moneyness are high. Over a day, they are even higher. Over a month, they are a little bit lower.

We estimate 12 models, the basic B-S and models 0 to 4 with heteroskedastic errors, for both log and level structures. From now on, we use diffuse priors for the parameters to make our results more comparable with the literature which only uses option prices to estimate the models. The models in figure 3 allow for three levels of heteroskedasticity. Figure 3 shows boxplots for the posterior distributions of σ, the intercept, and the slope coefficients of the models. The whiskers of the boxplots are the 5th and 95th percentiles. The body of the boxplot shows the median, first and third quartiles.

The top plots of figure 3 show that different models imply different volatility parameters. For the log models, median σ's go up from 0.25 for the B-S to 0.27 for models 2 to 4. For the level models, they go down from 0.24 to 0.22. This may be because level models concentrate on high prices, i.e., in and at-the money options, while log models emphasize out-of-the money options. [21] Since σ changes with the model, it has not reason to be an unbiased estimated of the return's standard deviation. It is a catch-all parameter which enables a functional form to fit the data better, even more so since we use uninformative priors for σ.

σ has a very concentrated distribution for the B-S model. This is because the model is so restrictive that there is no leeway for the parameter. As parameters are added, the distribution of σ spreads. Bossaerts and Hillion (1994a) note the same phenomenon. This happens, to some extent to all the parameters. The fact that the posterior densities of σ, the intercept and the B-S slope in models 1 and above are cross-correlated (not shown) also contributes to this.

Consider now the four bottom plots. For models 0 to model 3, 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. The values of the parameters for the logarithm and level models are not comparable because of the different functional forms.

Finally, consider the parameters of model 4. They are estimated with far less precision than the other models. In fact, the distribution for model 4 parameters have large outliers which do not appear on the plots. Figure 4 shows that model 4 with 14 parameters (3 levels of σ_η) is difficult to estimate even with 452 quotes collected over a week where the stock price is allowed to vary. This casts doubt on the adequacy of such a model in a true cross-sectional implementation. Additional parameters are costly and there is a marked difference in the width of the uncertainty between models 3 and 4. A lot of the problems from model 4 come from the high cross-correlations shown in Panel B of Table 2.

---

[21] As a heuristic check, we re-estimated the level models on only the out-of-the money options. The posterior means for σ were indeed higher.

4.4 The Parameters: σ_η

Let us turn to σ_η, the standard deviation of the model error. Do the more complex non-parametric models have lower standard deviation of errors? [22] Is there a need for heteroskedastic errors? To implement a 3 level heteroskedasticity, we need two cutoff values. Sample independent cut-off values sound appealing. Say that moneyness below 95% and 105% are groups 1 and 2, the rest is group 3. This poses a problem for small panels where one may not have enough data to identify each group. We select the cutoff values so that the three groups are identified. Here, we had 141, 192, and 119 quotes in each group. The cutoffs were 4% out- and 18% in the money. This reflects the fact that there are more quotes in- than out-of the money, see Table 2, Panel A.

Figure 4a shows the posterior distribution of σ_η for B-S, and models 2 and 3 in the log form. The top left plots show σ_η for the homoskedastic models. The other three plots represent $\sigma_{\eta,(1,2,3)}$ for out-, at-, and in-the money quotes. The top left plot shows that on average models 2 and 3 reduce model error from about 10.5% down to 7%. The other plots show the very strong heteroskedasticity. The mean standard deviation of the model 2 error is 12% out-, 4.5% at-, and 2.5% in-the money. The improvement due to models 2 and 3 take place out- and at-the money. Models 2 and 3 do worse than B-S in-the money,. The trade off still favors them: They bring down the error standard deviation, out-of-the money from 20% to 12%, at-the money from 5% to 4.5%, while increasing in-the money from 1.7% to 2.5%. Finally, Model 3 does not improve over model 2. Model 4, not in the picture actually did worse. This is another warning that adding terms to an expansion is not always rewarding when the terms need to be estimated.

Figure 4b shows these diagnostics for the level models. Direct comparison between the values in figures 4a and 4b is not easy. The first are relative and the second are dollar errors. We tackle this issue later in section 4.7. The top left plot shows σ_η for the homoskedastic models. The improvements due to models 2 and 3 are not as impressive. The medians are better, but the spread of the distributions increases. The other three plots confirm that the pattern of heteroskedasticity is opposite from the logarithm case. 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). The inspection of the posterior densities on Figure 4b reveals that many are strongly skewed. Even with 452 quotes collected over a week, the use of asymptotic normality could lead to flawed inference.

We now estimate the heteroskedastic models 2 and 3 for the entire December 1989, a sample of 1923 quotes. Figure 4c shows the posterior distribution of $\sigma_{\eta,(1,2,3)}$. For comparability with the weekly data, we used the same cutoff points for the heteroskedasticity. First, we compare model performance for weekly vs monthly samples. Take logarithm model 2. The median σ_η were 12%, 4.5%, and 2% for out-,in-,at-the money calls for the one week sample (figure 4a). The left plots of Figure 4c show that they are 28%, 7.4%, and 2% for the 1 month sample. Again, model 3 does not improve on model 2. This confirms that the use of a long sample decreases the fit not only for the B-S model but also for the extended models. Also, the non parametric expansions of models 2 and 3 do not alleviate the problem for the log model. The right plots show however that the level errors for the one month sample are not substantially higher than for the one week sample of figure 4b. The level specification seems more robust in long samples than the log.

---

[22]Unlike for the R^2 of a regression, this is not automatically the case.

This shows that increasing the sample size via an increase in the calendar span may not give a better chance to the non parametric expansions. There may be reasons for this. Model 3 for example says that the smile is a function of the time to maturity. Assume a Hull-White stochastic volatility world. Say that today's volatility is well below its mean. Model 3 picks it with a higher and flatter smile for longer maturities since volatility is expected to rise. If tomorrow's volatility is above its mean, the smile will be lower and flatter for longer maturities. Mix these two days into a sample and the pattern may be lost. [23]

4.5 Hedge Ratios

Contingent claim models are used for two purposes, pricing and hedging. The previous section gave an analysis of pricing errors via the estimates of the model error's standard deviation. We investigate whether the models have different hedging implications. Consider for example, the instantaneous hedge ratio Δ. Asymptotic estimation for the parameters would require the use of delta methods to get an approximate value for the standard deviation of a hedge ratio. The estimator of Δ would then be assumed to be normally distributed. Instead, we obtain draws of the posterior distribution of Δ by computation of the derivatives of the models specified in equations (3) or (4). Different models imply different functional forms for Δ.

Figure 5 shows the posterior distribution of Δ for B-S and logarithm models 2 and 3. The six plots show Δ for out-, 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 identified with precision, mostly because the B-S Δ is incredibly concentrated. This is because σ itself is concentrated for the B-S. 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. These differences are of the order of 2% for the short maturity calls and 1% for the long maturity, not economically relevant. The results for the level models, not shown, are similar. Again the differences are not economically meaningful. The differences between the log and level models are small too. For the level model, the differences are not reliably estimated either. The distributions have large overlaps from model to model.

4.6 The Expansion Function βx

The extended models incorporate additional functions of moneyness and maturity with possibly a fair number of additional parameters. We do not inspect each parameter. Instead we ask whether they result in different pricing implications from B-S. For example, in level model 2, the B-S part of the call price is incremented by $\beta_0 + \beta_2 z + \beta_3 z^2 + \beta_4 \tau$, a value which varies with τ and z. We want to see how flat this function is. For any value of z and τ, the parameter draws yields a sample of draws of the expansion function. It is not straightforward to interpret this as a bias function in the traditional sense. Since the parameters σ and β_1 are different for the various models, the B-S part of the models will be different too, though not across money and maturity. This is because we estimate these functions together with the B-S parameter values. [24]

---

[23]In this situation, the best hope for an ad-hoc model to track the dynamics is via the introduction of an observable related to volatility such as trading volume or a time series volatility forecast.

[24]For a strict bias analysis, one could set $\beta_1 = 1$ and tighten the prior on σ.

Figure 6 documents the posterior distribution of these functions for level models 2 and 3, estimated from the 452 quotes of December 4 to 8, 1989. We have allowed for three levels of σ_η. The vertical axis is in dollars. The horizontal axis shows the moneyness; stock divided by present value of exercise price.

On the top left plot, the solid lines are the 5th, 50th and 95th quantiles of the expansions. The dashed lines are their posterior mean, and +/- 1.64 standard deviation. The expansions are computed for model 3 and $\tau = 100$ days. For this slightly above average maturity the mean function shows no moneyness effect. Notice that the distribution appears normal since the quantiles are close to their normal values. For a 100 day option, the 90% confidence interval (hereafter CI) covers about 30 cents, 15% in the money. If properly specified this is quite precise since it is smaller than a typical Bid-Ask spread of 50 cents. However the CI for out of money calls is 20 cents. This is larger than typical spreads of below 18.75 cents. Recall that these intervals only reflect parameter uncertainty and not the B-S part and the model error.

The top right plot shows the 10%, 50%, 90% quantiles for models 2 and 3 and a 5 day call. Model 3 is a negative function of moneyness, while model 2 is a positive function. Allowing the smile to vary with the maturity (model 3) drastically modifies the pricing relative to the money for short term options. For low and high moneyness, the 90% CI of the two models do not overlap. For this short maturities, the CI's are quite narrow always below 20 cents, well below 10 cents at the money. This is smaller than the relevant bid-ask spreads. Now consider the bottom right plot. It shows these two models CI's for a 180 day option. Model 2 shifted up owing to the presence of τ in its expansion. Model 3's prediction however is now positively related to maturity. The CI's for both model are now much larger. This is specially noticeable for model 3 which generate CI's of 40 cents. The bottom left plot shows the posterior mean of the expansion for model 3 for different maturities. The reversal of the function from short to long maturities is striking. The 5 day, 15% out-of-the money calls are worth 20 cents more than in-the money calls, *relative to the B-S part of the model*. But 180 day 15% out-of-the money calls are 15 cents less than in-the money calls, relative to the B-S part of the model.

These functions change over time. The patterns of Figure 6 are however typical for a sample of this size. To conclude, model 3 has very different pricing implications than model 2 even though it has the same hedging implications. This brings up an interesting distinction between likelihood and non likelihood based methods. An econometrician accustomed to the GMM or Least Squares approach might say: "This is because you minimized pricing error. The estimation found the parameters optimizing the pricing of the various models, this will not highlight their potentially different hedging behavior". In fact a branch of the classical literature suggests that parameters should be estimated by optimizing the hedging (pricing) behavior of the model if the model is to be used for hedging (pricing). Given the non linearity in these functions, the point estimates will be different in small sample, but not in large sample however. There is therefore an inconsistency between the choice of objective function, i.e., only matters in small sample, and the theoretical validity of the estimation method, i.e., only asymptotic. Bayesian analysis does not optimize any criterion to produce a point estimate. It gives the posterior distribution reflecting the prior and likelihood information. The decision theoretic part is kept separate unlike for some non likelihood based classical method. [25]

The uncertainty around the expansion functions, due to parameter uncertainty only, varies

---

[25]We thank Peter Bossaerts for highlighting this point.

a lot with the moneyness and maturity. It typically is the largest for long maturity calls. However, it is most often well within the bid-ask spread. If this uncertainty is representative of the model performance, we should expect model prices to fall inside the Bid-Ask spread most of the time. We discuss this further.

4.7   In-Sample Specification Tests

Residual analysis helps test the stochastic assumptions of models. Bayesian residual analysis uses the exact posterior distribution of the residual as dicussed in section 3.1. We use the fitted density to analyze the performance of the logarithm heteroskedastic and homoskedastic models and the level heteroskedastic models. We conducted in-sample tests for the the first two weeks of December 89. We reestimated the models for the week of December, 11 to 15. This added 419 quotes to the 452 of the week of the 4th to the 8th. The results are in table 3.

Insert Table 3 here

Panel A shows the biases and root mean squared errors (RMSE) of the models on the basis of the posterior mean of the fit density of the quotes. We also look at the out-of and in-the money sub-samples. We also show the bias and RMSE, using only the quotes where the posterior mean is outside the spread. This brings an economic rationale on the model comparison. If two models yield posterior means inside the bid ask spread, it is of little interest which one is closest to the mid point.

The left half of panel A shows that average biases are far smaller than the RMSE's. Biases are never more than a few cents for the level models. The biases for the log models are also in the order of 1 or 2%. This is smaller than the typical percentage spread. Consider now the right half of Panel A. The logarithm models 2 have smaller RMSE than the B-S. For quotes with posterior means outside the spread, model 2 reduces the RMSE from 16% to 13%. There is no improvement beyond model 2. The in-sample improvement brought by model 2 are very minor, nil outside the spread. Recall that if the errors exhibit cross-sectional or time series correlation, the posterior mean of the parameters may mis-represent their variability. We can check how close the posterior mean of σ_η is to the RMSE's in Panel A. For the B-S, the posterior mean is below the RMSE. For the extended models, the posterior means are closer to the RMSE's. Specific forms of autocorrelation in the errors can be diagnosed by computing the posterior distribution of the autocorrelation function.

Do log models perform better than level models? The residuals of the log models are *relative* pricing errors. They are not comparable to the residuals for the level models. For the log models, we compute the posterior mean of the model value and hence, the dollar pricing errors. Due to the exponentiation, the variance of the relative error may induce biases in the pricing error. The results are in Panel B. Note the large pricing errors of model 4 which was not reliably estimated. This confirms the high cost of an increase of the expansion, even if the sample spans a week. We do not discuss model 4 any further. The log B-S models biases are larger than the level B-S biases, but they are within common bid ask spread values. The log extended formulations have no relevant bias.

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. What is only a second moment effect in the log model, becomes incorporated in the mean in the exponentiation.

Second, the extended models significantly improve upon the B-S. Third, the level models seem to have marginally better RMSE's 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.

These diagnostics are based upon the posterior mean of the residuals. Panel C, distribution analysis, documents the specification of the predictive and fit density. The in-sample predictive density is obtained as discussed in section 3.1. 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 greatly reduce this number, from 25 to 30% for the B-S, down to below 20% for models 2. We then compute the percentage of quotes falling inside the interquartile range of the fit density (column 2) and the predictive density (column3). Column 2 shows that the fit covers grossly underestimate the variability of the model. For example, the IQ range of the fit density for the B-S covers the quotes only 2% of the time. Those of the extended models from 15 to 24%. Column 3 shows the predictive covers. The Homoskedastic models all have very misleading predictive densities. Allowing for heteroskedasticity greatly improves the predictive densities. The level models perform much better than the log models on this account. As a whole the predictive densities appear to be the appropriate benchmarks for a specification test. The fit densities which do not account for model error, underestimate the variability of the models.

The final columns of the panel deal with the problem of the intrinsic bounds. Call model should not generate negative call values, bound $B_1=0$, or values below $B_2=S-PV(X)$. The level models may have a problem with B_1. 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 bound B_2. Actually, 30% of the asks violate this bound. [26] We computed the number of quotes for which the first quartile of thee predictive density violated B_2. For this, we only used the quotes which bid was above B_2. After allowing for heteroskedasticity, the logarithm and level models give similar results. The last column shows that less than 1% of the predictions have this undesirable property.

We conducted similar test on different samples. We also estimated the models on the entire month of December, 1923 quotes. The typical RMSE's were 12 and 10% for log models B-S and 2, 14 and 13 % for level models B-S and 2. The log models had the same pricing RMSE's as the level models. First, this shows that when the sample has a longer calendar span, the fit may be degraded. Second, the extended models do not improve the fit as much. We conjectured earlier in the paper that this could happen if the patterns that they fit vary with time. A longer sample can mix the patterns and reduce the efficacy of the expansion.

So we also investigated samples covering short calendar spans. We conducted a bi-daily re-estimation from January 2 to March 31, 1990, a total of 8749 quotes. Given the little evidence supporting improvement by the larger extended models, we compared only the B-S and model 2. The RMSE's were 0.31 (B-S) and 0.18 (Model 2). Outside the Bid-Ask spread, they were 0.49 and 0.31. 37% (B-S) and 26% (Model 2) of the posterior mean prices were outside the spread. These errors are actually higher than for the week long samples. Also, model 2 exhibits significantly smaller errors. Further, there were 627 observations for which model 2's mean prediction was outside the

---

[26]The violations are small and for short term options. If the bound S-X is used, only 2 quotes violate the bound. For short term options, even a small transaction cost on the Tbill would make an arbitrage impossible.

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 7 documents the time series variation of some of the parameters. The top left plot shows the volatility σ. Parameters exhibit large variations. The model will be well specified out-of sample if the posterior densities of the parameters are in line with the magnitude of these variations. The two right plots of Figure 7 show the mean, first and third quartiles of σ and σ_η. If the IQ range of these densities covered the next period posterior mean, this would indicate that the parameter's uncertainty is reasonably specified with respect to the uncertainty of it future values. The inspection of these plots indicates that the posterior distributions seem to underestimate the time series variability of the parameters. The bottom left plot shows that the posterior means of σ_η for the two models are not very different from one another, even though model 2 often produces smaller values.

4.8   Out-of-sample Specification Tests

The previous results, even the *predictive* densities, were in-sample. We now turn to the out-of-sample analysis. For each of the two weeks, we used the parameter draws to compute diagnostics for the quotes of the following week. This resulted in an out-sample of $419 + 624 = 1043$ quotes. Table 4 summarizes the out-of-sample evidence in a format similar to table 3.

Insert Table 4 here

Again, the biases are small and we do not discuss them. Consider the RMSE's. First, the extended models do not produce better RMSE's than the B-S. This is true for residuals and pricing errors, levels and logarithms. Second, the log B-S out-of sample RMSE's are about the same as in-sample. The better in-sample RMSE's of log models 2 and 3 has vanished. They are now at par with the B-S. The level B-S RMSE's have however deteriorated, mostly due to a deterioration for out-of the money quotes. For the pricing errors in Panel B, the out-of sample RMSE of the log models put them at par with the level models.

This contrast with the results of DFW who report a severe deterioration of their extended models in out of sample tests, while having a nearly perfect fit in sample. First we did not allow a perfect fit in sample. Second, the performance degradation was not as strong. The RMSE's increased from 7% to 10% (11 to 14 cents) for log-model 2, 11 to 15 cents for level model 2. DFW only sampled quotes every Wednesday. So their out sample was seven days apart from their estimation sample. We check if this is an issue by looking at the RMSE's for the first day out of sample. The results, right side of panel A, are that the 1 day ahead RMSE's are worse than the RMSE's based on 1 to 5 days ahead. This confirms that the deterioration in performance occurs right away.

Panel B in table 4 completes the out-of sample tests. The first column confirms a noticeable, but not catastrophic degradation of performance. 20% of the in sample means of predictions were outside the B-A spread. 30% of the out-of-sample predictions are outside the spread. The B-S does as well as the other models. Next, the IQ range coverage ratios, broken down by moneyness, show how crucial the heteroskedasticity is. The homoskedastic B-S and model 2 predictive IQR covers the in-the money quotes 98% and 97% of the time, a gross overestimation. They cover the

18

out-of the money quotes 35% and 25% of the time. These numbers are markedly improved by the heteroskedasticity. Still, even heteroskedastic, the log models predictive densities do a poor job of reflecting the variability of the quotes. The level models appear better specified. The last column shows that the predictive densities do not violate bounds anymore out-of sample than in sample.

We also used the estimates from the whole month of December to make out-of sample forecasts for the first week of January. As above, the extended models did not do better than the B-S. However in that case the B-S also performed worse out-of sample than in sample.

5   Conclusion

We incorporate model error into the estimation of contingent claim models. We design a Bayesian estimator which to produce the posterior distribution of deterministic functions of the parameters, or the residual of an in-sample observation and assess its abnormality, and to determine the predictive density of an out-of-sample estimation. We document the potential non-normality of some posterior distributions. This shows the added value of the Bayesian approach.

We apply this method to quotes on stock calls options, and document the behavior of several static non parametric expansions nesting the B-S. These types of non parametric models have received attention in the recent literature. They could be justified as an approximation to an unknown model or a complex model too costly to implement. We formulate the error in both relative (logarithm) and dollar terms (level models), and allow it to be heteroskedastic. Our analysis shows that, within sample the small expansions improve on the Black Scholes mispricing. They reduce root mean squared errors of pricing and residuals. Whatever the calendar span, the larger expansions are unable to improve on the B-S even in sample. In-sample, the extended models have similar hedging implications but different pricing implications than does the B-S model. The log models do not exhibit better performance than the level models. One must allow for heteroskedastic model error specially for the log models.

Specification tests on the basis of the fit and predictive densities of the call price, show that the failure to include model error in prediction intervals leads to an underestimation of the variability of the quotes. 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.

Out-of sample, the non parametric expansions most always fail to improve on the B-S. Specification tests show that even the heteroskedastic models do not model the variability of out-of and in-the money quotes properly. These results cast a serious doubt on the usefulness of static non parametric expansions. Their better fit does not survive out of sample. However, out of sample performances are not drastically worse than in sample performances. These results are in constrast with those of recent studies which exhibit quasi perfect fit in sample and disastrous fit out-of sample. The difference lies in the estimation technique which allows us to make more realistic, though not satisfactory, out-of sample inference on the basis of our in-sample estimation

This paper is a first pass at the much neglected likelihood based estimation of contingent claim models. The MCMC estimators which we implement are flexible. They can be extended to estimate more sophisticated error structures or functional forms. We discuss some of these extensions

below.

In the appendix we show how to extend the model to allow for a more general error structure consistent with the presence of intermittent mispricing. The intuition underlying this formulation is that sometimes, an additional error ϵ_i, possibly a market error, is added to η_i. This extension leads to outlier diagnostics easy to interpret. The probability that a given observation has the extra error can be estimated. The estimation and prediction procedures may be more robust as they allow for these mispricing errors, a form of fat-tailness. The cost of this extension, the added burden on the estimation, has to be weighted against the likelihood that such errors are indeed present.

In some cases one must relax the assumption that the underlying or the risk free rate are observed exactly. For example, the martingale restriction built in the B-S model can be relaxed with either the stock price or the risk free rate being free parameters, e.g., Longstaff (1993) and Giannetti and Jacquier (1998). Renault (1995) argues that non synchroneous measurement may have an impact on the pricing formula. The stock price in the B-S could be modelled as $S = S^* + \mu + \nu$, where S^* is the measurement and ν_i is random with appropriate priors possibly reflecting bid-ask spreads. S becomes an unobserved state variable. The hierarchical structure of MCMC estimators allow extensions to unobservable state variables. A worthwhile research direction can be to investigate whether errors inside the B-S (or other) box obviate the need for errors outside.

A class of models assumes the homogeneity of the option price, see Garcia and Renault (1997). The input variable is the ratio of the stock over the exercise price. Such non parametric models could be implemented. The extensions could take the place of the standard normal cumulants d_1 and d_2. Alternatively σ could be expanded inside the B-S function, e.g., $\sigma = \sigma_0 + \sigma_1 z + \sigma_2 z^2..$, as in DFW. Placing errors and extensions inside the model allows to randomize it without ever violating the no-arbitrage restrictions.

Model mispecification often results in time varying-parameters. One can argue that static non parametric extensions can not perform well out-of sample because they do not capture the dynamics of this time variation. So a worthwhile direction is to allow the non parametric forms to capture the missing dynamics via a well chosen prox, e.g., stochastic volatility. Within this framework, the model presented here begs to be extended to a dynamic estimation setup. Even better will be an extension of the estimator to parametric models that nest the B-S.

References

Acharya, S., and Madan, D., (1995): "Asset Pricing in an Incomplete Market with a Locally Risky Discount Rate", Working Paper, Finance Department, University of Maryland.

Bakshi, G., and Cao, C., and Chen, Z., (1998): "Empirical Performance of Alternative Option Pricing Models", *Forthcoming Journal of Finance.*

Bates, D, (1995): "Testing Options Pricing Models", 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. (1994a): "Method of Moments Estimation of Contingent Claims Asset Pricing Models," *Journal of Mathematical Finance.*

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

Brown, S., and Dybvig, P. (1986): "The Empirical Implication of the CIR Theory of the Term Structure of Interest Rates", *Journal of Finance*, 41, 617-630.

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.

Dumas, B., Fleming, J.,, and Whaley, R., (1998): "Implied Volatility Functions: Empirical Tests", *Journal of Finance*, 53, 2059-2106.

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.

Garcia, R., and Renault, E., (1998): "Risk Aversion, Intertemporal Substitution and Option Pricing", CIRANO Working Paper, www.cirano.umontreal.ca, 1998.

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.

Gelman, A., and Rubin, D.B., (1992a): "A Single Series of the Gibbs Sampler provides a False Sense of Security", *Bayesian Statistics 4*,(J.M. Bernardo et al., eds.). Oxford University Press, pp. 625-632.

Gelman, A., and Rubin, D.B., (1992b): "Inference from Iterative Simulation with Multiple Sequences (with discussion)", *Statistical Science*, 7, 457-511.

George, T., and Longstaff, F., (1993): "The Cross-Sectional Distribution of Bid-Ask Spreads and Trading Activity in the S&P 100 Index Options Market", *Journal of Financial and Quantitative Analysis*, 28, 381-397.

Ghysels, E., and Patilea, V., and Renault, E.,and Torres, O.,(1997): "Non Parametric Methods and Option Pricing", Working Paper CIRANO, Montreal: www.cirano.umontreal.ca

Giannetti, A. , and Jacquier, E., (1998): "The Cross-Section of Option Prices: Test of the Equivalent Martingale Measure" Working Paper Boston College: www2.bc.edu/ jacquier.

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., and Polson, N. , (1995): "Efficient Option Price Prediction under Stochastic Volatility", Manuscript, Finance Department, Boston College.

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

Kass, R.E., and Carlin, B.P., and Gelman, A. and Neal, R.M. (1998): "MCMC in practice: A roundtable discussion", *The American Statistician.*

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

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*

Polson, N., and Roberts, G.O., (1994): "On Geometric Convergence of the Gibbs Sampler", *Journal of the Royal Statistical Society*

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. (1994): "Option Hedging and Implicit Volatility", Mathematical Finance.

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

Schotman, P., (1996): "A Bayesian Approach to the Empirical Valuation of Bond Options", *Journal of Econometrics*, 75, 183-215.

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

Tierney, L. (1994): "Markov Chains for Exploring Posterior Distributions", *Annals of Statistics*, 22, 1701-1762

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

Zellner, A. (1971): "An Introduction to Bayesian Inference in Econometrics", John Wiley & Sons, New York.

Table 1: List of Models

Model #	Model Feature	Number of Parameters[a]
B-S	none	2: σ, σ_η
0	Add intercept	3: $\sigma, \sigma_\eta, \beta_0$
1	Add slope coefficient	4: $\sigma, \sigma_\eta, \beta_0, \beta_1$
2	Add τ, z, z^2	7
3	Add $\tau z, \tau z^2$	9
4	Add τ^2, z^3, z^4	12

[a]This is for models with homoskedastic pricing errors. Models with heteroskedastic errors, see below, have more than one σ_η parameter.

Table 2: TOYS'R US Quotes Data, Dec. 89 to March 90[1]

Panel A: Summary Statistics

	BA$	BA%	Moneyness%	Maturity
Median	0.375	8	9	46
Q1,Q3	0.25, 0.5	6, 13	1.3,21	22, 94
10%, 90%	0.125, 0.5	4, 22	-6, 31	8, 164

Panel B: Cross-correlation of input variables over Dec 4-8, 1989

	z	z^2	τ	$z\tau$	$z^2\tau$	τ^2	z^3
z^2	0.89						
τ	-0.19	-0.12					
$z\tau$	-0.14	-0.20	0.54				
$z^2\tau$	-0.25	-0.24	0.51	0.88			
τ^2	-0.10	-0.10	0.59	0.07	0.05		
z^3	0.87	0.98	-0.12	-0.17	-0.23	-0.11	
z^4	0.82	0.97	-0.10	-0.15	-0.19	-0.10	0.994

[a]BA is the Bid-Ask Spread. The moneyness is in percentage of the present value of the exercise price. The Maturity is in days.

Table 3: In-Sample Performance Analysis: Dec 4 to Dec 15, 1989[a]

Panel A: Residual Analysis

	BIAS		RMSE			
Model	all	out BA	all	oom	im	out BA
Log-Hom						
B-S	-0.018	-0.050	0.10	0.18	0.02	0.15
2	0.000	0.001	0.07	0.12	0.02	0.12
3	-0.000	0.001	0.066	0.11	0.02	0.12
4	0.002	0.006	0.065	0.11	0.02	0.07
Log-Het						
B-S	-0.012	-0.028	0.10	0.18	0.017	0.16
2	-0.005	-0.017	0.07	0.12	0.015	0.13
3	-0.004	-0.013	0.07	0.12	0.016	0.12
4	0.000	0.002	0.07	0.11	0.020	0.10
Lev-Het						
B-S	-0.013	0.005	0.12	0.092	0.144	0.19
2	-0.002	0.012	0.11	0.087	0.126	0.19
3	0.000	0.017	0.11	0.081	0.122	0.19
4	0.005	0.014	0.11	0.078	0.134	0.21

Panel B: Pricing Analysis

	BIAS				RMSE			
Model	all	oom	im	out BA	all	oom	im	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

Table 3 - continued
 Panel C: Distribution Analysis

Model	% Pred. out BA	Fit Cover	Pred Cover[b]	B1[c]	A<B2[c]	(Q1<B2|B>B2)
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
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
Lev-Het						
B-S	24	2	50	4	30	0.4
2	18	13	52	3	30	0.5
3	18	15	51	0.3	30	0.6
4	15	31	64	0.7	30	3

[a]The models have been estimated over the week of Dec. 4-8, (452 quotes) and re-estimated for the Dec. 11-15 week (419 quotes). That is 871 quotes used to compute the above statistics. The symbols used are, 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.

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

[c]B1:Percentage of observations such that Prob(Pred<0) >0.001. B2 is the other intrinsic bound, S-PV(X). The next column show the number of ask quotes violating the bound. Only 2 quotes violated the bound S-X. The last column is the percentage of quotes for which the first quartile of the predictive density violated the bound, counted over only the quotes which Bid was above the bound.

Table 4: Out-of-Sample Performance Analysis, Dec 11 to Dec 23, 89[a]

Panel A: Residual and Pricing Analysis

Model	Residual RMSE							Pricing RMSE			
	Up to 5 days ahead				1st day ahead						
	all	oom	im	out BA	all	oom	im	all	oom	im	out BA
Log Homoskedastic:											
B-S	0.10	0.18	0.02	0.15				0.15	0.11	0.16	0.22
2	0.11	0.23	0.02	0.18				0.16	0.14	0.17	0.23
3	0.12	0.24	0.02	0.18				0.16	0.13	0.17	0.23
Log Heteroskedastic:											
B-S	0.1	0.20	0.02	0.16	0.12	0.23	0.02	0.15	0.12	0.15	0.22
2	0.1	0.21	0.02	0.17	0.10	0.19	0.02	0.14	0.13	0.15	0.21
3	0.11	0.22	0.02	0.18	0.10	0.19	0.02	0.14	0.12	0.15	0.22
Level Heteroskedastic:											
B-S	0.15	0.14	0.14	0.23							
2	0.15	0.13	0.14	0.22							
3	0.15	0.14	0.14	0.23							

Panel B: Distribution Analysis

Model	% Pred. out BA	IQR cover[b] Pred: all, im, oom			B1[c]	A<B2[c]	(Q1<B2 | B>B2)
Log Homoskedastic:							
B-S	31	69	98	35	na	0.9	5
2	36	59	97	24	na	0.9	3
3	34	58	97	24	na	0.9	2
Log Heteroskedastic:							
B-S	31	54	64	39	na	0.9	0.7
2	31	52	75	31	na	0.9	0.6
3	29	52	78	33	na	0.9	0.6
Level Heteroskedastic:							
B-S	29	41	46	39	2.2	0.9	0.4
2	30	39	51	35	1.2	0.9	0.2
3	31	39	53	35	0.6	0.9	0.4

[a]The statistics are computed over the week after the estimation week. 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

[b]Percentage of the observations for which the predictive interquartile range covers the market price.

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

Figure 1: MCMC estimation diagnostics, Dec 1, 89: 140 quotes,
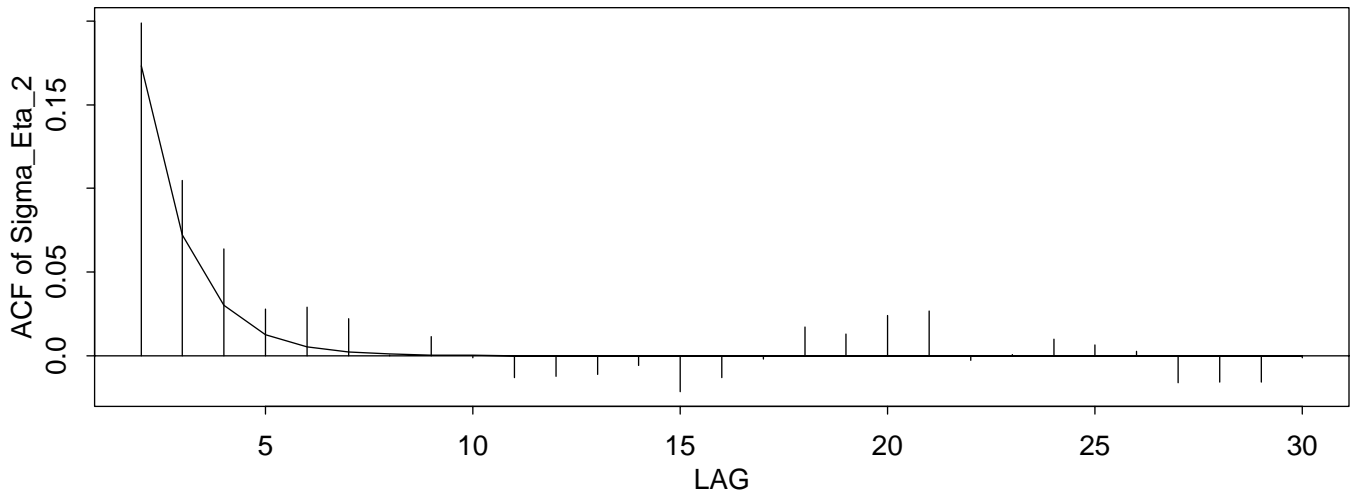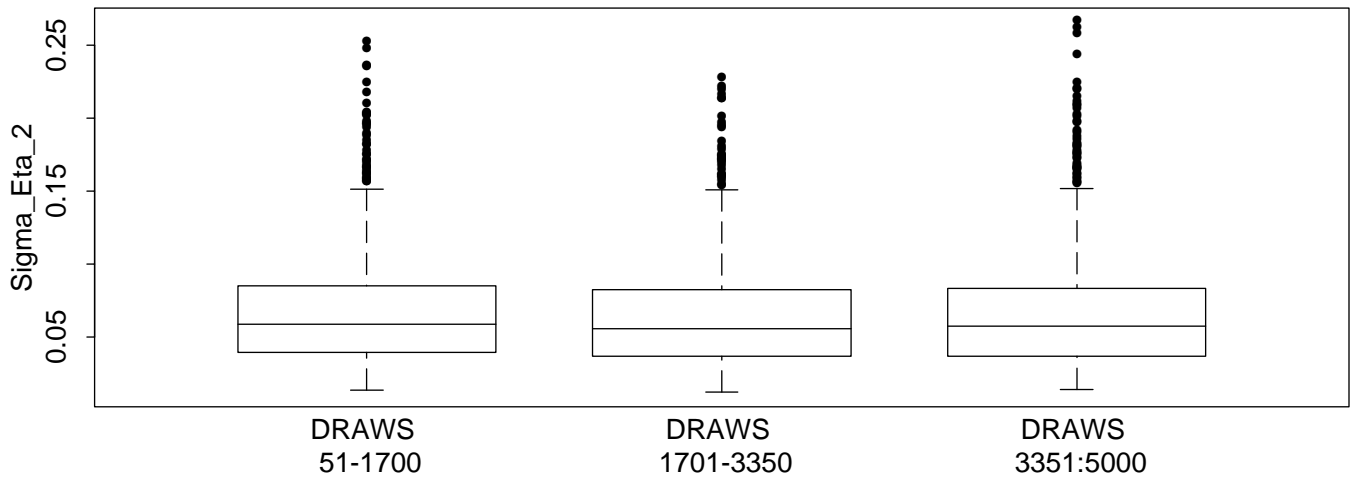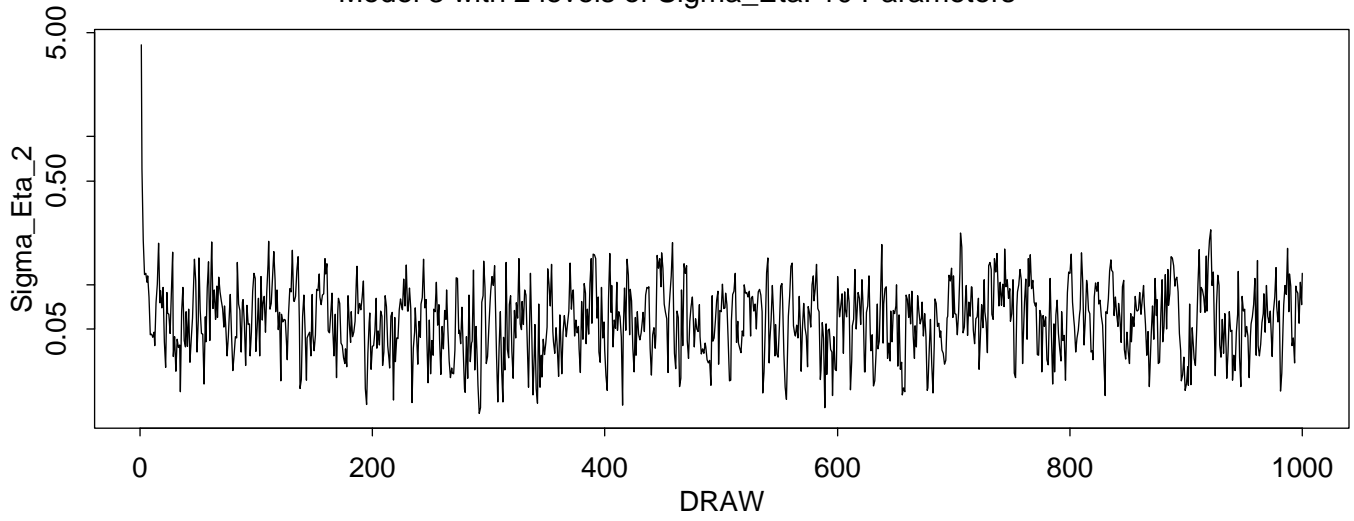Model 3 with 2 levels of Sigma_Eta: 10 Parameters

Figure 2: Non normal posterior density, Dec 1, 89: 140 quotes,
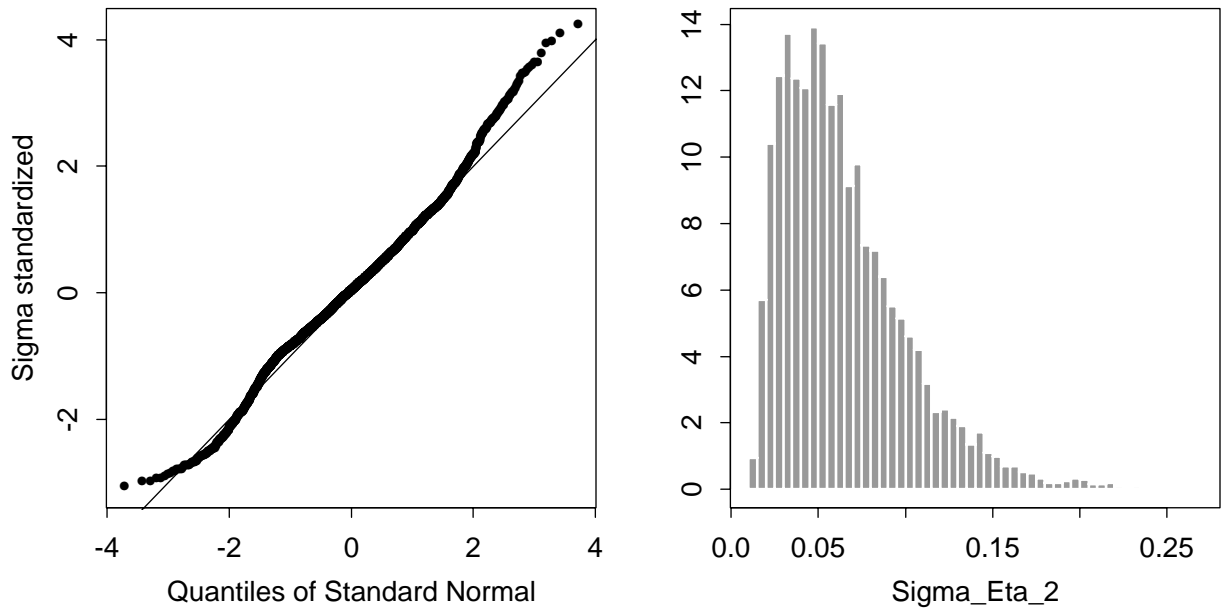Model 3 with 2 levels of Sigma_Eta: 10 Parameters

Figure 3: Sigma, Intercept and Slope, Logarithm and Level Models
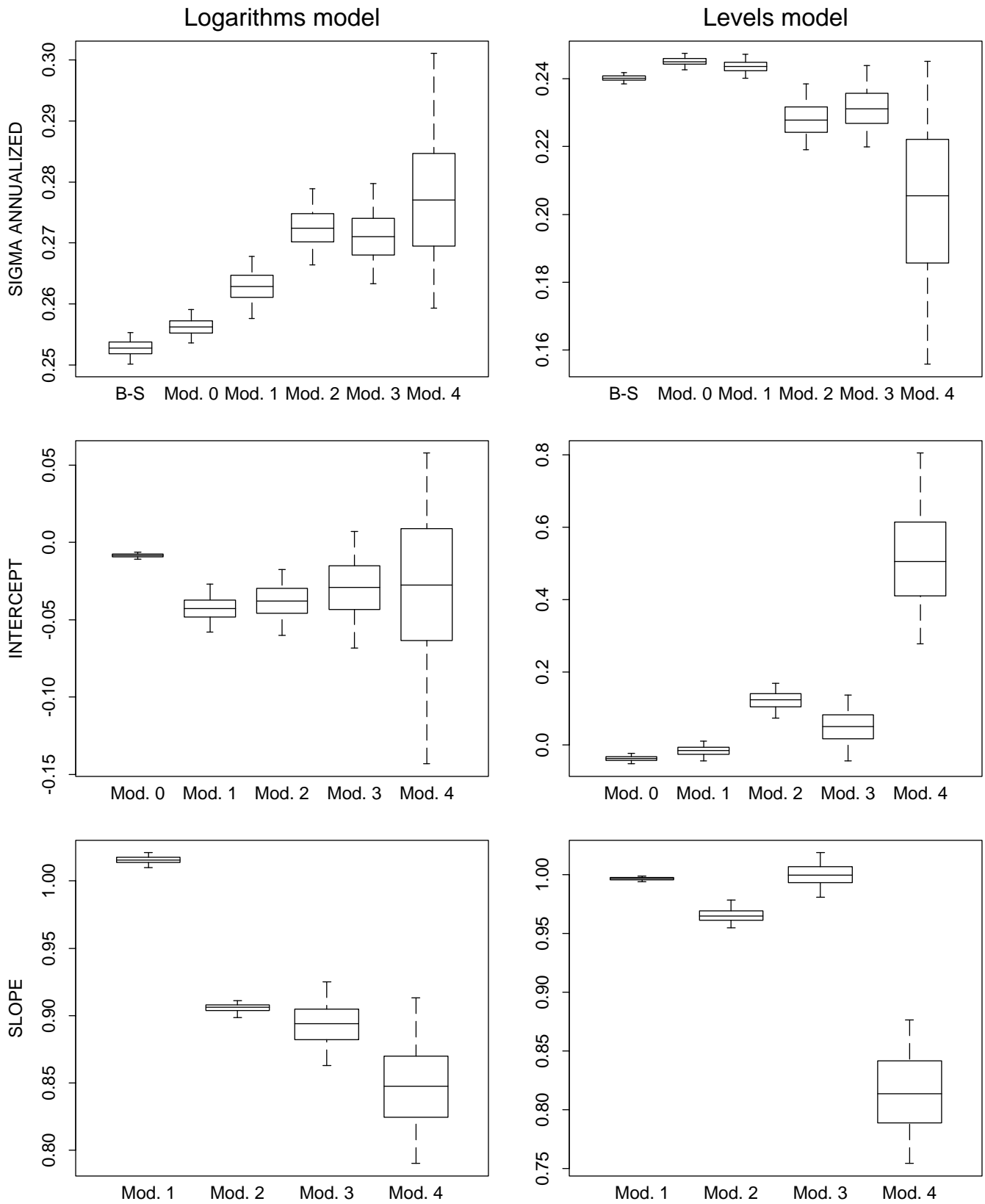TOYS'R US, 452 quotes Dec. 04-08, 1989

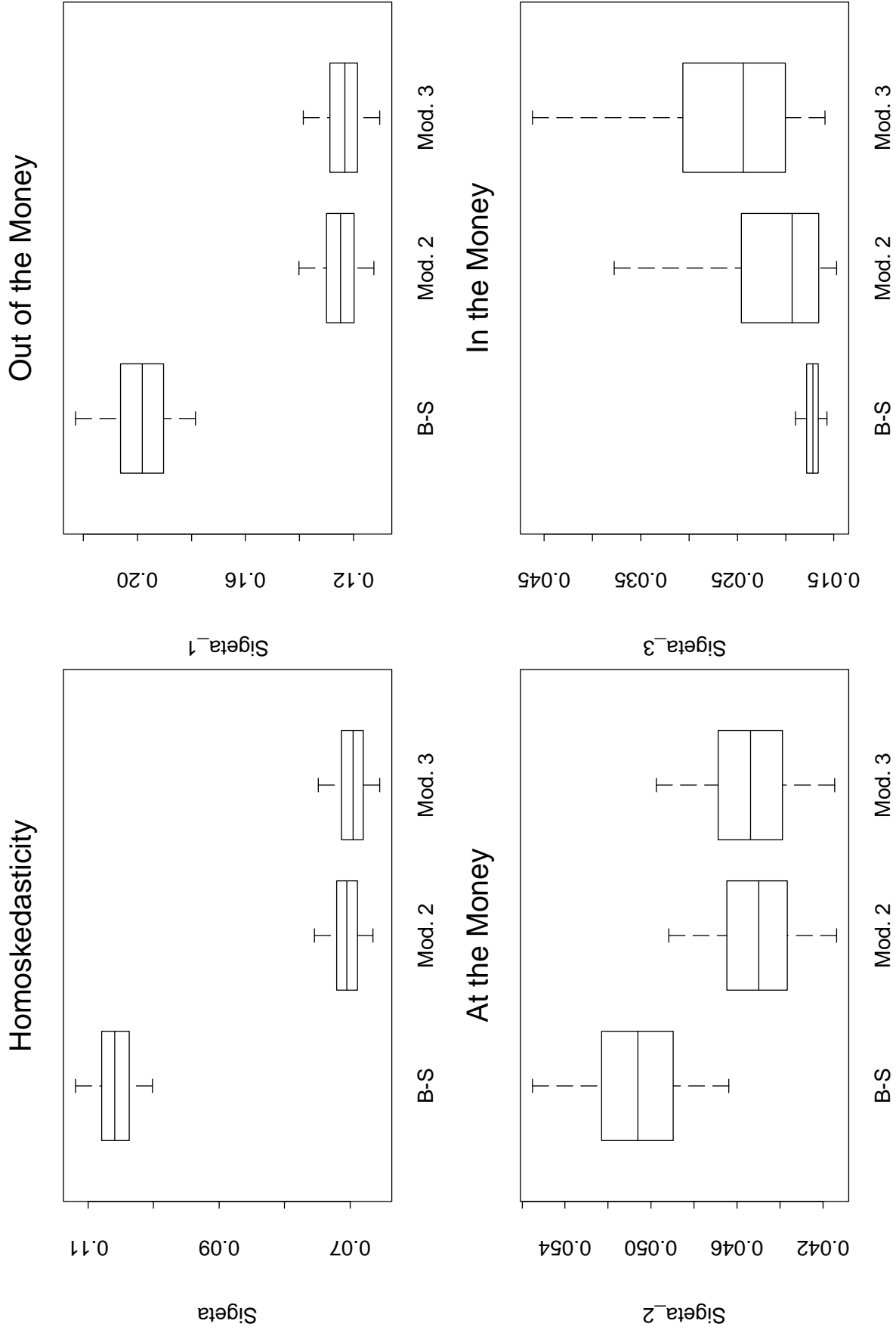Figure 4a: Logarithm Model Error and Heteroskedasticity, 452 quotes Dec. 04-08, 1989

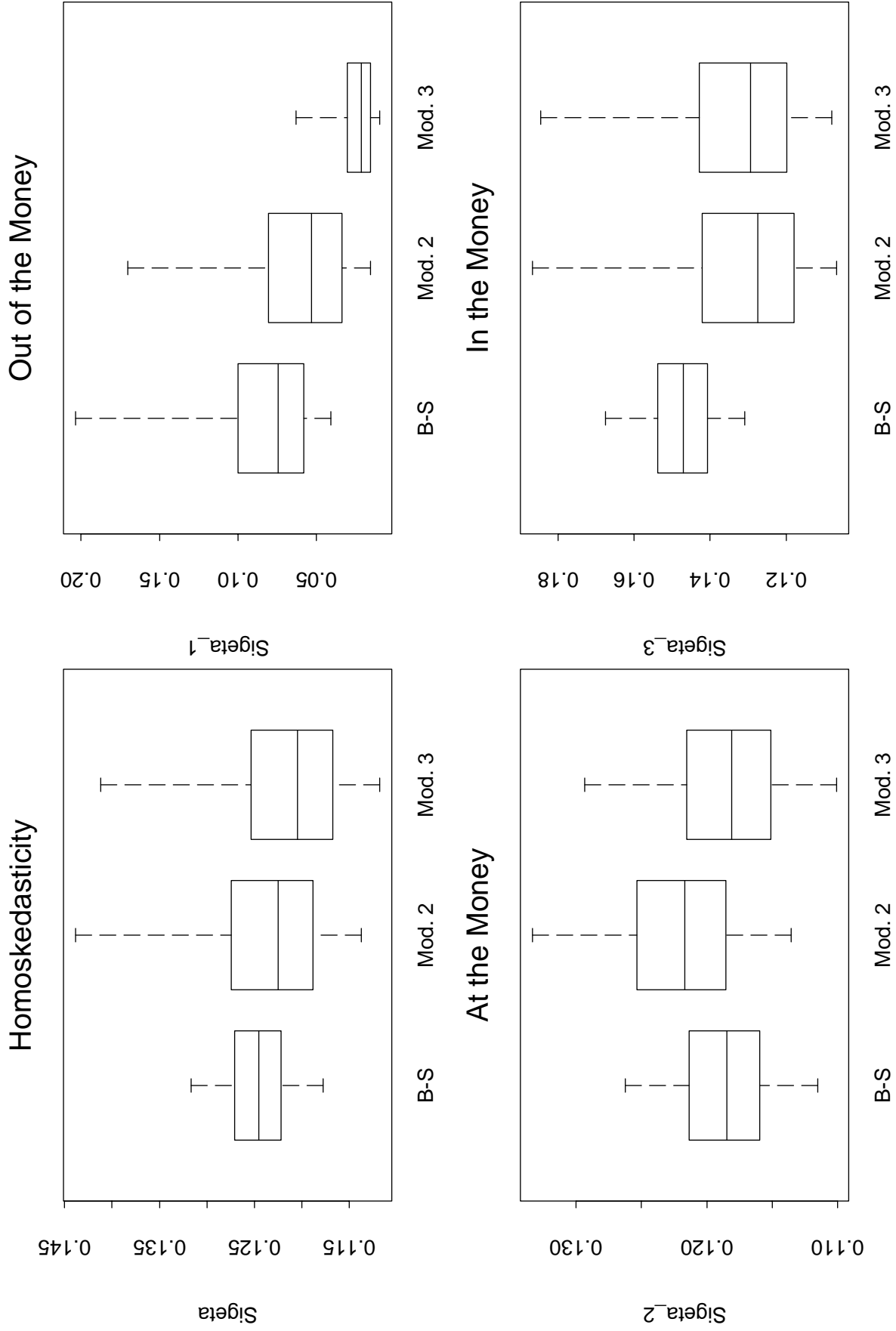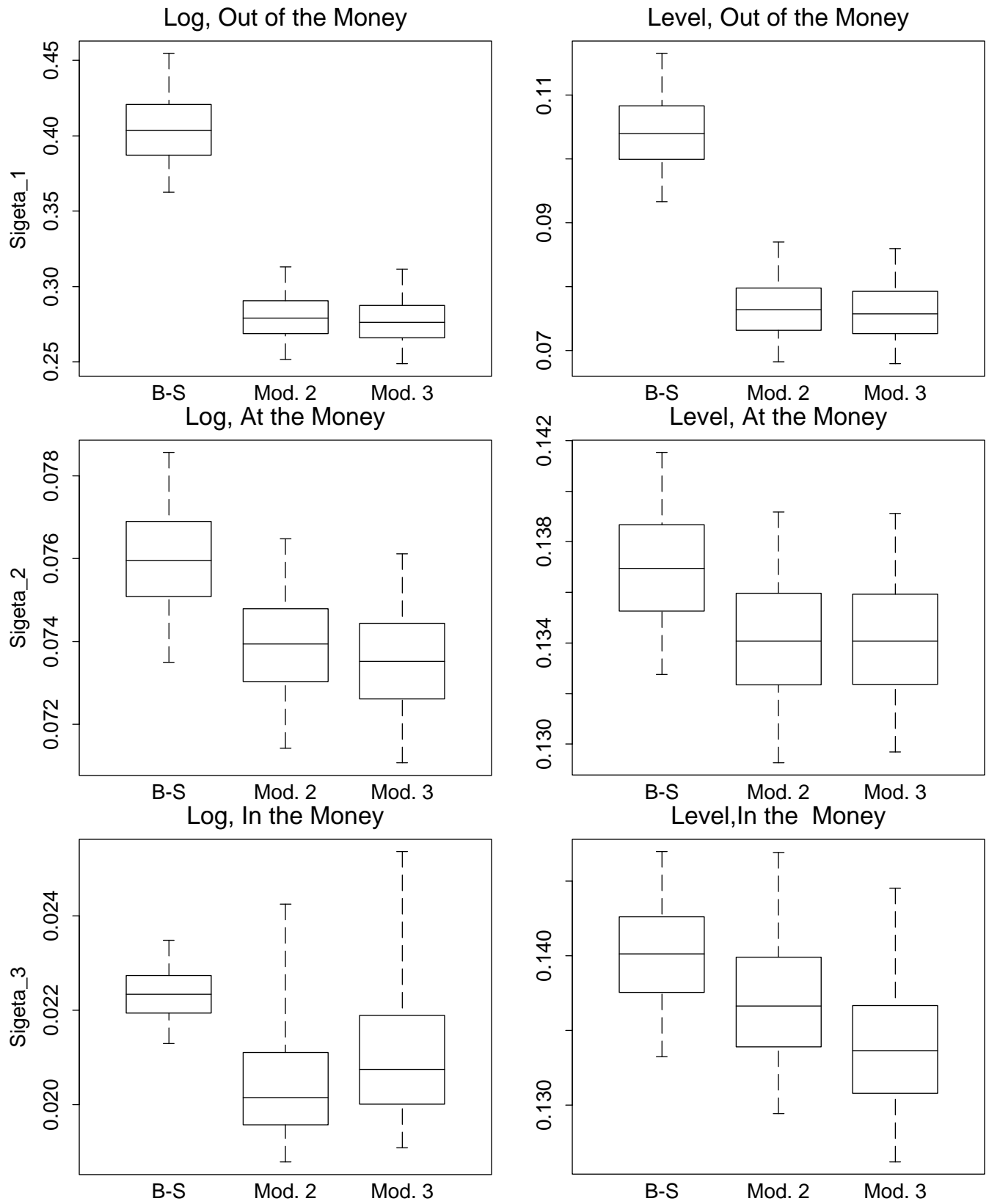Figure 4b: Level Model Error and Heteroskedasticity, 452 quotes Dec. 04-08, 1989

Out of the Money

Homoskedasticity

In the Money

At the Money

Figure 4c: Model Error and Heteroskedasticity, 1923 quotes Dec. 1989

Figure 5: Hedge Ratios for 3 Logarithm models, Dec 04-08 1989: 452 quotes
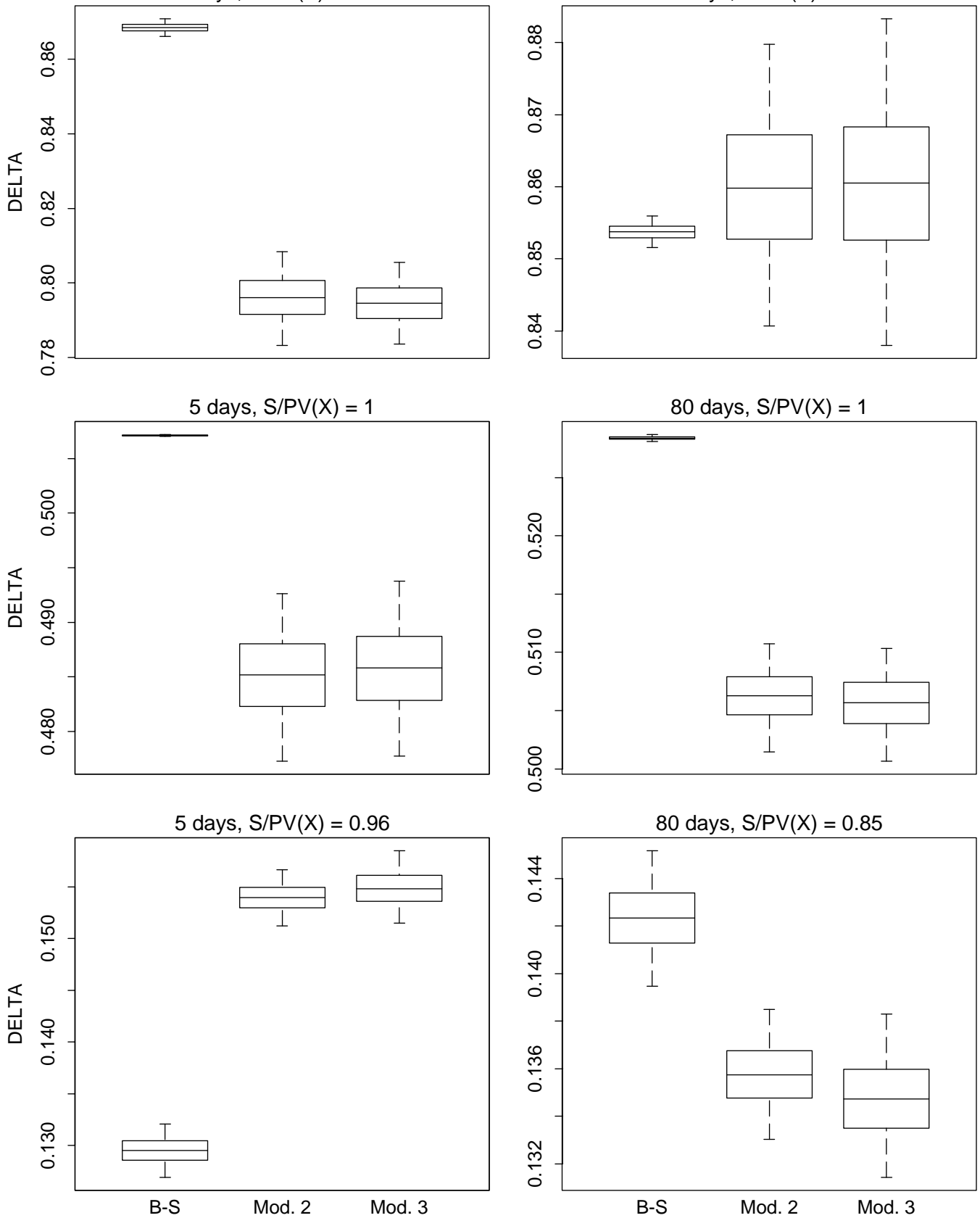
Figure 6: Expansions of the Level Model, Dec 04-08 1989, 452 Quotes

Tau = 5 Days
Model 2
Model 3

Tau = 180 Days
Model 2
Model 3

Model 3, Tau = 100 Days
Mean, +/- 1.64 std. dev.
5th, 50th,95th Percentiles

Model 3
TAU = 5 days
62 days
100 days
180 days

S / PV(X)
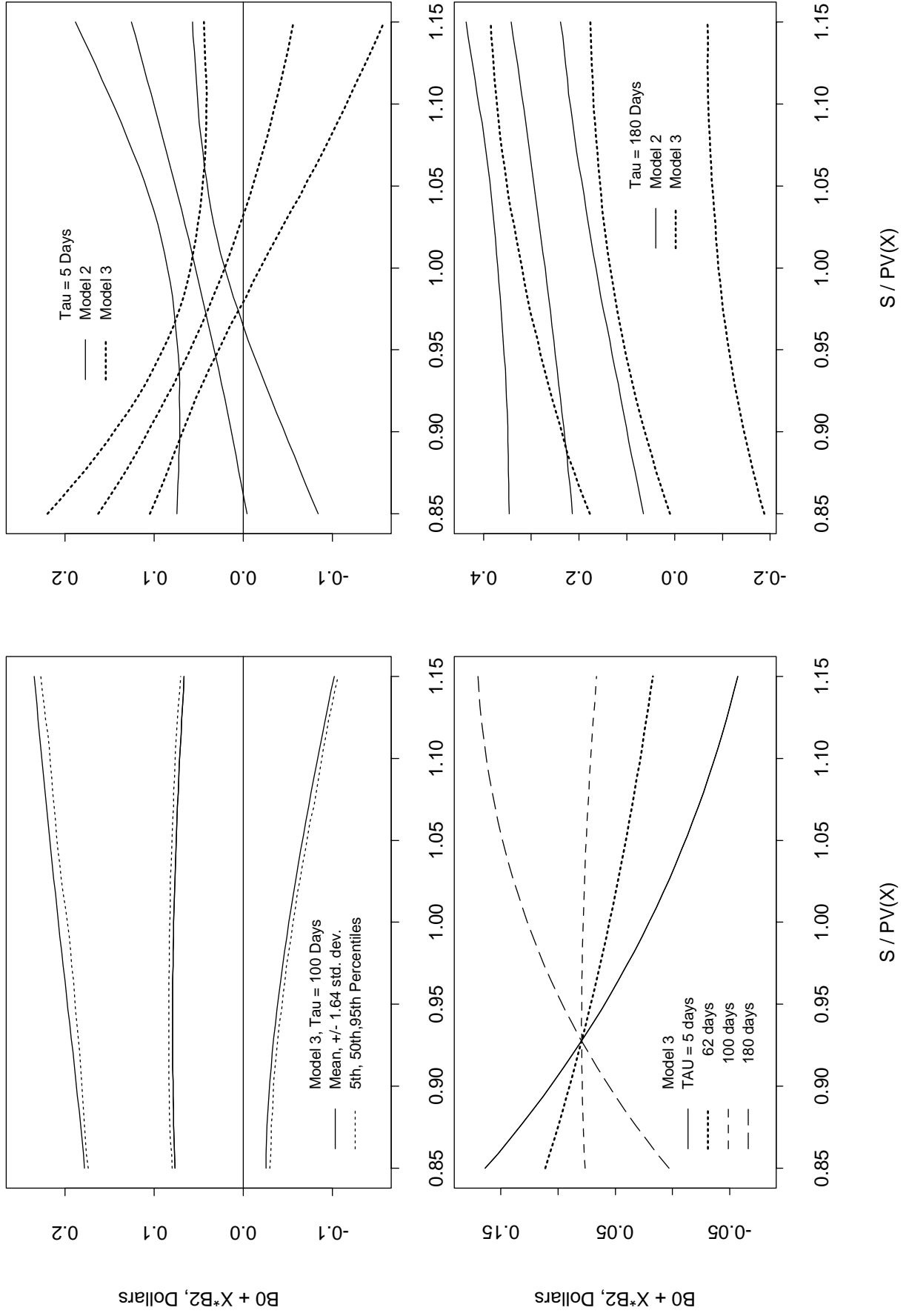
S / PV(X)

B0 + X*B2, Dollars

B0 + X*B2, Dollars

Figure 7: Time Series Plots, Bi-Daily Parameter Means, Level Models, January-March 90