CIRANO
Centre interuniversitaire de recherche
en analyse des organisations

Série Scientifique
Scientific Series

96s-12

Model Error in Contingent Claim Models Dynamic Evaluation

Eric Jacquier, Robert Jarrow

Montréal
mars 1996

CIRANO

Le CIRANO est une corporation privée à but non lucratif constituée en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du ministère de l'Industrie, du Commerce, de la Science et de la Technologie, de même que des subventions et mandats obtenus par ses équipes de recherche. La *Série Scientifique* est la réalisation d'une des missions que s'est données le CIRANO, soit de développer l'analyse scientifique des organisations et des comportements stratégiques.

CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère de l'Industrie, du Commerce, de la Science et de la Technologie, and grants and research mandates obtained by its research teams. The Scientific Series *fulfils one of the missions of CIRANO: to develop the scientific analysis of organizations and strategic behaviour.*

Les organisations-partenaires / The Partner Organizations

- Ministère de l'Industrie, du Commerce, de la Science et de la Technologie.
- École des Hautes Études Commerciales.
- École Polytechnique.
- Université de Montréal.
- Université Laval.
- McGill University.
- Université du Québec à Montréal.
- Bell Québec.
- La Caisse de dépôt et de placement du Québec.
- Hydro-Québec.
- Fédération des caisses populaires de Montréal et de l'Ouest-du-Québec.
- Téléglobe Canada.
- Société d'électrolyse et de chimie Alcan Ltée.
- Avenor.
- Service de développement économique de la ville de Montréal.
- Raymond, Chabot, Martin, Paré
- Banque Nationale du Canada

ISSN 1198-8177

Model Error in Contingent Claim Models Dynamic Evaluation*

Eric Jacquier[†], *Robert Jarrow*[‡]

Résumé / Abstract

Résumé

Nous incorporons formellement l'incertitude des paramètres et l'erreur de modèle dans l'estimation des modèles d'option et la formulation de prévisions. Ceci permet l'inférence de fonctions d'intérêt (prix de l'option, biais, ratios) cohérentes avec l'incertitude des paramètres et du modèle. Nous montrons comment extraire la distribution postérieure exacte (de fonctions) des paramètres. Ceci est crucial parce que l'utilisation la plus probable, réestimation périodique des paramètres, est analogues à des échantillons de petite taille et demande l'incorporation d'informations a priori spécifiques. Nous développons des modèles Monte Carlo de chaînes markoviennes afin de résoudre les problèmes d'estimation posés. Nous fournissons des tests de spécification, à la fois pour l'échantillon et le modèle prédictif, qui peuvent être utilisés pour les tests dynamiques et les systèmes de trading en utilisant l'information en coupe transversale et temporelle des données d'option. Finalement, nous généralisons la distribution d'erreurs en tenant compte de la (faible) probabilité qu'une observation ait une plus grande probabilité d'erreur. Cela fournit pour chaque observation la probabilité d'une donnée aberrante et peut aider à différencier erreur de modèle et erreur de marché. Nous appliquons ces nouvelles techniques aux options d'équité. Quand l'erreur de modèle est prise en considération, le Black-Scholes apparaît très robuste, en contraste avec les études précédentes qui, au mieux, incluait l'erreur de paramètre. Après, nous étendons le modèle de base, i.e. Black-Schles, par des fonctions polynomiales des paramètres. Cela permet des tests intuitifs de spécification. Les erreurs en échantillon du B-S sont améliorées par l'utilisation de ces simples modèles étendus, mais cela n'apporte pas d'amélioration majeure dans les prédictions hors-échantillon. Quoi qu'il en soit, les différences entre ces modèles peuvent être importantes parcequ'elles produisent différentes fonctions d'intérêt telles que les ratios et la probabilité d'erreur d'évaluation.

---

* We gratefully acknowledge essential support from the CIRANO (Centre Interuniversitaire de Recherche en Analyse des Organisations), Montréal, and the Kamakura Corporation for Jarrow. This paper has benifitted from discussions with Warren Bailey, David Bates, Peter Carr, Bernard Dumas, René Garcia, Antoine Giannetti, Eric Ghysels, 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.

† Cornell University & chercheur visiteur au CIRANO

‡ Johnson Graduate School of Management, Cornell University

Abstract

We formally incorporate parameter uncertainty and model error in the estimation of contingent claim models and the formulation of forecasts. This allows an inference on any function of interest (option values, bias functions, hedge ratios) consistent with the uncertainty in both parameters and models. We show how to recover the exact posterior distributions of the parameters or any function of interest. It is crucial to obtain exact posterior or predictive densities because the most likely implementation, a frequent updating setup, results in small samples and requires the incorporation of specific prior information. We develop Markov Chain Monte Carlo estimators to solve the estimation problem posed. We provide both within sample and predictive model specification tests which can be used in dynamic testing or trading systems, making use of both the cross-sectional and time series information in the options data. Finally, we generalize the error distribution by allowing for the (small) probability that an observation has a larger error. For each observation, this produces the probability of its being an outlier, and may help distinguish market from model error. We apply these new techniques to equity options. 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 then extend the base model, e.g., Black-Scholes, by polynomial functions of parameters. This allows for intuitive specification tests. The Black-Scholes in-sample error properties can be improved by the use of these simple extended models but this does not result in major improvements in out of sample predictions. The differences between these models may be important however because, as we document it, they produce different functions of economic interest such as hedge ratios, probability 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 (mainly non simultaneity, etc...). The second error is due to the estimation of parameters such as volatility. The third is model error due to the fact that no model perfectly explains prices even in the absence of the first two errors. This is due to the simplifying assumptions necessarily made about the structure of markets and trading. These errors have not been so far integrated in the evaluation of contingent claim models so as to best learn from misspecifications and use necessarily imperfect models. This paper proposes and implements a method to do this.

Researchers, under the null hypothesis of a given model, usually do not incorporate model error in their investigations. Further they often do not fully account for the sampling error due to parameter uncertainty. This is true, for example, with the seminal papers in the empirical option pricing literature. [1] Typical studies generate point estimates of option prices via direct substitution of point estimates for the underlying volatility. Model point estimates are then compared to market prices. They lack the ability to produce specification tests for either in-sample fit or out-of-sample prediction.

The approach recently outlined in Rubinstein (1994) ignores *any* possibility of error and performs exact fitting. This is consistent at the time of calibration, with the no arbitrage framework from which the specific model being implemented stems. But the model is necessarily imperfect as it does not incorporate more general processes for the observables, conditions on approximations of state variables, and ignores market frictions too hard to modelize. The estimation of an imperfect model with no overidentifying restrictions is a case of overfitting: The cross-section is fitted nearly perfectly at time t. The parameters retrieved may be useless at time t+1. This can explain the results in Dumas, Fleming, and Whaley (1995). They fit a binomial model quasi perfectly, only to conclude that it is worthless in out-of-sample performance. A realistic requirement for an acceptable model is that it produced similar out and in of sample performances. For this requirement to be testable, the estimation technique must be able to deliver the uncertainty around model predictions.

A few studies concentrate on parameter uncertainty alone, by incorporating only the underlying data in the likelihood. [2] For the Black-Scholes model, Lo (1986) adopts a Maximum Likelihood (ML) setup and assumes that the option price estimator is normally distributed. The uncertainty of the prediction comes from the sampling variability of the ML estimator of volatility. Lo gives an approximation for the sampling standard deviation of the option price estimator. Jacquier, Polson, and Rossi (1995) show how to construct efficient option price predictors reflecting parameter

[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.

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

uncertainty under stochastic volatility models. Lo soundly *rejects* the Black-Scholes specification. That is, the confidence intervals generated by the option price prediction uncertainty do not cover true values with the expected frequencies. Although it allows for parameter uncertainty and produces a specification test, this approach still forces a zero model error, an unrealistic requirement. More realistically, one would wantto admit a model error that (1) is unrelated to observable model inputs, and (2) has distributional properties preserved out of sample. Much like in any econometric procedures, predictive intervals would follow from both model error (noise) and parameter uncertainty. We will see in our empirical section that when this is done, the the Black-scholes appears much more robust than prviously believed.

Many reasons are given in the literature for the existence of departures from the basic Black-Scholes model. Hull and White (1987), Bailey and Stulz (1988), Heston (1993), Stein and Stein (1993), and others write no arbitrage or equilibrium based models allowing for stochastic volatility. Renault and Touzi (1992) show that a true model with stochastic volatility causes a smile shaped bias in Black-Scholes implied standard deviation. Bossaerts and Hillion (1994b) show that the inability to transact continuously can also cause a smile. Platen and Schweizer (1995) develop an equilibrium model with hedgers and speculators. In their model, a time varying assymetry of the smile (skewness) is caused by the (time varying) demand for hedging out or in the money. Practitioners' common practice of gamma and vega hedging reveals their awareness of model error even though most trading systems do not incorporate an analysis of model error and do not use models other than the Black-Scholes.

The first contribution of this paper is a method of estimation of contingent claim models which incorporates model error and parameter uncertainty. This allows us to obtain a representation of the uncertainty about the estimate or prediction of a quantity of interest. This representation is needed for specification tests such as residual analysis and predictive performance, hedging, dynamic model selection, and identification of likely mispricing. Imperfect models often result in time varying parameters, and need to be implemented in an updating setup allowing for periodic reestimation. Effective sample size may then never be large, but the results of a previous sample may serve as the prior information for the current sample. So the estimation needs to (1) be relevant for small samples and (2) allow the incorporation of prior information. Asymptotic methods based upon the normality of the estimator, e.g., methods of moments or maximum likelihood, are ineffective given these requirements which are best satisfied by the Bayesian framework we adopt.[3]

Another contribution of this paper is the extension of a *basic* model, here the Black-Scholes, with 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, specifically the model inputs. Otherwise, the model implies the possibility of predictable abnormal returns from simple trading strategies, or the resulting price prediction may lead to incorrect inference. [4] First, well known biases, smiles and skewness, mean that a model easily implementable like the Black-Scholes may not meet these error requirements. Second, the econometrics of the

[3]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 purpose.

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

more sophisticated models, e.g., stochastic volatility based, is at a very early stage. The *extended* models which we propose are on the other hand as simple to implement as the Black-Scholes. Finally, the relevance of the extended models is not limited to improving the Black-Scholes only. We will always be in a situation where there exists a *basic* model relatively easy to implement and more complex models are entertained but significantly harder to implement. The extended model approach proposed will then always be a relevant bridge between basic and more complex model.

The final contribution of the paper is the empirical results. We show that test of the Black-Scholes which only account for parameter uncertainty are flawed. They vastly underestimate the spread of the predictive densities and are biased toward rejection. The evidence against the Black-Scholes, once model error is incorporated, is much milder. We use extended models with powers and cross-products of moneyness and time to maturity to nest the Black-Scholes. First, these variables are justified as potential expansions of a more complex unknown model. Second, Lo, Hutchinson, and Poggio (1993) argue that they can approximate the Black-Scholes quite well. Here we use them as approximation of the unknown model over and above the Black-Scholes. We show that the pricing and hedging implications ot the extended-models differ from the Black-Scholes and do improve the in sample specification. However, they have a hard time doing so out of sample.

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

Finally, the flexibility of MCMC estimators allows us to generalize the model. We can allow for heteroskedasticity, and for an intermittent additional mispricing where the error sometimes has a larger variance than usual. This introduces an additional state variable for each option price observation, equal to one if there is an additional error, zero otherwise. We provide the corresponding exact sample estimator, using tools from Markov Chain estimation. The analysis can produce, for each observation, the probability that it is an outlier. [5] This diagnostic may be 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 maybe 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.

The paper is structured as follows. Section 2 introduces the general model and the estimators. Section 3 details the implementation of the method in the case of both *basic* and *extended* Black-Scholes models. Technical issues are concentrated in the appendix. Section 4 is the empirical application of the method to stock option data. Section 5 discusses the updating implementation and generalizations of the error distribution. Section 6 concludes.

[5] At first this seems just like a way to add fat tails. The interest in formulating the mixing variable, is that it produces a quantity of economic interest, the probability of a quote being an outlier.

2 General Framework

2.1 Model and Error Structure

Consider a sequence of discrete observations of a contingent claim's market price, C_t, for t $\in$ 0,1,2,...,N. Assume that there exists an unobservable equilibrium or arbitrage free price for each observation i, 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. One may prefer a log-additive error specification to the standard additive one for two reasons. First, it insures the positivity of C_t, the out of the money intrinsic bound, even with an unbounded distribution for ϵ_t. This does not remove the issue of the in the money intrinsic bound, S-PV(X).. Second, the log-additive specification models relative rather than absolute errors, insuring that out-of-the-money options with low prices are not ignored in the diagnostic. If a given amount is to be invested in a strategy, one may argue that relative pricing errors are a more relevant criterion. The difference between the two forms is an empirical question. We will conduct our empirical tests with both forms.

If the market is always in equilibrium, $\epsilon_t \equiv 0$. Otherwise, mispricings exist. Let $\mathcal{F}_t : i \in 0, 1, ..., N$ be the information set of the researcher studying the system. We assume that C_t is included in $\mathcal{F}_t$, but that both c_t and ϵ_t are not. Let P(.) represent the objective probability measure associated with the system, and let E(.) denote the expectations operator.[6] We also assume that $\mathcal{F}_t$ includes observations of the underlying stock price, denoted S_t. 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 c_t. The model, m depends on vectors of observables x_t, and non-observables θ, i.e., $x_t \in \mathcal{F}_t$ and $\theta \notin \mathcal{F}_t$. [7] We now incorporate in our analysis the view that the model is an approximation even though it was theoretically derived as being exact. It contains an unobservable error, η_t. Formally,

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

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 may not satisfy this condition.

[6] Formally, there is a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ 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.

[7] 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

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

Simplifying assumptions on the structure of trading or the underlying stochastic process made to derive tractable models, often result in the models exhibiting biases, i.e. a non i.i.d error structure. Renault and Touzi (1992), Taylor and Xu (1993), Engle, Kane, and Noh (1993), and Heston (1993) show this within the context of stochastic volatility. Further, Renault (1995) shows that even a small 0.1% error in the underlying price measurement (non synchroneity) can cause skewed implied volatility smiles. Bossaerts and Hillion (1994b) show that the assumption of continuous trading leads to biases in the implementation. Platen and Schweizer (1995) develop a microeconomic model of hedging which may result in time varying skewed smiles. In all the above cases, the errors are related to the inputs of the model.. Finally, 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). [8]

So 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 i.i.d. for most known models. b_t refers to the basic parametric 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 b_t by the introduction of the coefficient β_1 and the linear combination $\beta_2' x_{2t}$. x_{2t} may include functions of the observables x_{1t}, or other relevant variables, and of course an intercept term. x_{2t} can also be one or several models competing with b_t in which case the *extended* model equation allows an estimation nesting competing models. [9] The extended model allows the capture of biases in $\mathrm{E}(\eta_t \mid \mathcal{F}_t)$. It is justified as an approximation of a more general model, unknown or without a closed form representation (see Jarrow and Rudd (1982)), or very costly to implement. The additional cost of the extended model will turn out to be minimal and its intuition similar to the basic model. [10]

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

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

by which 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. η_t and ϵ_t are not identified without further assumptions. In section 6, we propose an error structure to identify outlying quotes which could be suspected to originate from market error. Until then, we assume only one pricing error η_t. We will allow for the heteroskedasticity of η_t as a function of the moneyness. The log model is equivalent to the modelling of relative errors, and one may expect these to be smaller for larger options. Alternatively, one may expect the levels model to generate smaller dollar errors for smaller call values.

[8] 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.

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

[10] In a trading system where the costs of changing model are high, model (4) provides an inexpensive control of model error in hedging a trader's portfolio (book). Then standard portfolio theory to minimize the remaining model error risk. At that stage, a proper specification of the error covariance structure is important.

2.2 Estimation and Prediction with Monte Carlo Methods

For simplicity, we assume that the information set $\mathcal{F}_t$ consists of past and current observations of the contingent claim's market price $\mathrm{C}_{t-\tau}$, the underlying asset price $\mathrm{S}_{t-\tau}$, and other relevant observables $\mathrm{x}_{t-\tau}$, e.g., interest rates, time to maturity. Call $\underline{y}_t$ this finite-dimensional vector of histories at each date t, then $\mathrm{E}(\cdot \mid \mathcal{F}_t) = \mathrm{E}(\cdot \mid \underline{y}_t)$.

We want to derive the posterior distributions of the parameters, of any interesting functions 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, (2) the probability that a given hedge (delta, gamma, vega) will lose money due to model error.

We start the model with a prior distribution $\mathrm{P}(\theta)$, where for convenience θ represents all the parameters including σ_η and β. The contingent claim model used and the distributional assumptions made for the pricing errors and the process of the underlying asset yield the likelihood function $\mathrm{P}(\underline{y}_t \mid \theta)$. By Bayes' theorem, the posterior $\mathrm{P}(\theta \mid \underline{y}_t)$ 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 given a sample of draws of θ.

The general approach to estimate the posterior distributions is as follows: (We will show how to) simulate from the distribution $\mathrm{P}(\theta \mid \underline{y}_t)$. This yields a sample of 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 the parameters 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 its functions of interests have closed form posterior densities. The Monte Carlo approach removes the need to perform numerical integrations, or to resort to the usual delta method approximation for moments computation together with the normality approximation.

Another important deterministic function of the parameters is the residual for each observation. Again, each draw of θ implies a draw of the posterior distribution of η_t for each observation t. The residual is simply $\log C_t - \log m(x_t, \theta)$. The residual 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. The density of $(\eta_f \mid \theta, \underline{y}_t)$ is normal with mean 0 and variance $\sigma^2_{\eta_f}$. So, the predictive density of $(\eta_f \mid \underline{y}_t)$ is

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

Again, the integration in equation (5) is readily performed by Monte-Carlo simulation: 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 $(\eta_f, \theta \mid \underline{y}_t)$. For each such 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 mean or median of the predictive density $P(C_f \mid \underline{y}_t, x_f)$ is a point prediction. Quantiles provide a model-error-consistent 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 methodology followed given draws posterior distribution of the vector of parameters θ. Here we discuss the intuition of two algorithms required to produce draws of θ. More technical discussions of the method are in section 3 and the appendix.

θ is a vector of parameters. We need to be able to make draws from the posterior density of θ. The crucial requirement is that we can write an analytical expression for the kernel of this joint density. This does not yet allow us to draw from it. The standard approach is to rewrite the joint posterior as a product of conditional densities, from which one can draw. In a standard linear regression, a draw of the slopes β and noise standard deviation σ is generated as follows; draw σ from the inverted gamma, then draw $\beta \mid \sigma$ from the normal. The sample of draws of β then follows the standard Student-t distribution.

The model here is more akin to a non linear regression even in the simplest case. It is impossible to use the standard approach to draw from θ. The solution is facilitated by the incorporation of the Gibbs sampling algorithms [11] it solves the following problem. In our case, we want to but can not draw from (σ, σ_η). We can however draw from $(sigma \mid \sigma_\eta)$ 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 distribution σ and σ_η. The algorithm applies to any number of conditionals. It is easy to check for the vanishing of initial values.

We would like to use the Gibbs algorithm but there is one remaining problem. We can draw from the distribution of $(\sigma_\eta \mid \sigma)$ but we can not draw directly from $p(\sigma \mid \sigma_\eta)$. The Metropolis algorithm solves this problem. The kernel of p has an analytical expression but the integration constant does not. In fact we do not need to ever compute the integration constant of p. [12] First, we select a *blanketing* density q with shape similar to p, from which it is easy to make direct draws. Then all we need to know are the shapes of the two distributions p and q. So we only consider the kernel of p. For each draw made from q, the Metropolis algorithm is a probabilistic rule of acceptance and rejection that draw. The rule takes into account the shape difference between p and q. This results in a sample of draws of q which converge in distribution to a sample of draws from p. Even if the shape of q is not close to that of p, the algorithm goes through. The closer the shapes of q and p, the faster the algorithm will generate informative draws on σ. We discuss the Metropolis algorithm in more details in section 3.3

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

[12] This is essential for an algorithm to be feasible. It is theoretically possible but practically infeasible to use any 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.

The combination of these two algorithms constitutes a Markov Chain Monte Carlo estimator. The draws generated converge in distribution to draws of the posteriors of the parameters, under very mild and verifiable conditions.

3 Application to the Black-Scholes Model

3.1 Models and Data

We now illustrate the above technique via an application to the Black-Scholes model. In the Black-Scholes economy, S_t follows a lognormal distribution, i.e., $R_t = \log(S_t/S_{t-1}) \sim N(\mu, \sigma)$. There is an unresolved theoretical question with respect to the heterogeneity of the underlying process with respect to the derivative. In our application the last return used predates the first panel of option price data. We can then assume a zero correlation between the stock returns and model error. In fact, we view the returns data as the basis for the prior on σ.

We also assume that the risk free rate $r_{f,t}$ is known without error. This assumption could be relaxed. A natural alternative would be $r_{f,t} = r^{\star}_{f,t} + \nu_t$, where $r^{\star}_{f,t}$ is the most relevant observed risk free rate, and ν_t a noise. In this case the risk free rate is an unobserved state variable, and our option price estimator nests an optimal signal extraction for the state variable. The same method could be used in cases where the underlying (s_t) is observed with error or unknown. [13]

We select from the Berkeley options database, quotes on calls on the stock of TOYS'R US, from December 89 on. They will be used in the analysis for the remainder of the paper. TOY is traded on the NYSE and is an actively followed stock that does not pay dividends. This allows the use of a european option model for calls. There are commonly between 80 and 300 quotes on TOY calls daily. We will estimate the models over a day, a week, and a month. 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 σ, we also collect TOY stock returns from the CRSP database for the period leading to November 30^{th}, 89.

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

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

where b_t is the Black-Scholes model. 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/PV(X), 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 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.

[13] It is the hierarchical structure of MCMC estimators that allows the extension to unobservable state variables, and makes them so superior to standard methods. We see an example of this in section 6 with the generalized error specification.

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 of 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 far we need to take the expansion of the relevant variables. Model 4 can be seen as a test of the necessity of further expansions beyond the parsimonious models 2 and 3. One could also consider other variables. Liquidity considerations might suggest the bid-ask spread of both the option and the stock as posssible right hand side variables.

3.2 Priors and Posteriors

The extended models with proper prior nest the basic model with diffuse priors. Diffuse priors can be obtained by increasing the variance of the various proper priors. We use the prior

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

Given σ, the joint prior of β and σ_η is the convenient normal-gamma prior used in regression analysis. Apart from σ, the priors are conjugate, and result in similar posteriors given the likelihood. For σ, we use the inverted gamma consistent with a data based prior on the returns time series. This does not add any complication to the diffuse prior case. The inclusion of proper priors may be warranted in this problem. 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 prior conditions no arbitrage conditions, such as the Merton bounds, by truncating the priors to eliminate parameter values that violate the bounds. This can be done with simulation estimators by just rejecting draws of the posterior that violate the bounds. When the sample is updated the previous posterior distribution may be a natural basis for the formulation of the prior distribution.

We also allow the error η_t to be heteroskedastic. Appendix A shows the posterior distributions of the extended model with proper priors.[15] In the logarithm models, η_t is a relative pricing error. Going from small to large 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 size of the call. We can not draw directly from the joint posterior $P(\beta, \sigma, \sigma_\eta \mid \underline{y}_t)$. So

[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.

[15] The appendices do not reflect this heteroskedasticity extension yet.

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 an inverted gamma and a multivariate distribution respectively. Direct draws are possible from both. However direct draws from $\mathrm{P}(\sigma \mid \beta, \sigma_\eta, \underline{y}_t)$ are not possible. Section 3.3 discusses the Metropolis step needed to draw from $(\sigma \mid .)$.

3.3 Posterior Distribution of σ

In this section and appendix B, we discuss the implementation of the MCMC estimator in further details for the interested reader. 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 CDF 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 results. 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.

Even if the shape of q is not close to that of p, the algorithm goes through, albeit inefficiently with many rejections or repeated draws. The closest the shapes of q and p are, the fastest the algorithm will generate 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. Here the blanket q is chosen as a truncated normal with mean the mode of p(σ).

Figure 1 documents the implementation of the estimator for a model with many parameters and a short sample. This serves to demonstrate the reliability of the method even in small sample. 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 of Dec 1, 1989. The prior on σ is flat, based on the last 5 daily returns of 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 where most draws will come from. 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 very case, out of 5000 draws of σ, we had 32 rejections and 28 repeats. These numbers are typical of the efficiency of the algorithm for this problem.

The results we exploit from MCMC algorithms is that the draws converge in distribution to draws of the desired marginal posterior distribution. The reasonable question is then: When have we converged? Figure 1 shows the tools used to answer this question without going into theoretical considerations. The first tool is a time series plot of the draws, shown for $\sigma_{\eta,2}$ on the top right plot. We intentionally started the chain from unrealistic values of the parameters to check how quickly the draws would settle down to a constant regime. Here it took less than 10 draws for the system to settle down. We decide to be conservative to discard the firt 50 draws. A further diagnostic is then to compute sample quantities for different segments of the remaining sample of draws. The three boxplots shown in Figure 1 yield identical conclusions, confirming 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 very quickly confirming that the sequence of remaining draws is stationary.

4 Empirical Application

4.1 Parameters

We now demonstrate the estimation of extended models on the TOY options data. The results presented are based on 3500 draws of the posterior distributions of the parameters.

These posterior distributions are the basis for simple tests of these competing models. We ask whether different models imply different values for common parameters? This question with respect to 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. Let us consider σ and the intercept and slope coefficients of the B-S. 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 clearly 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 an opposite changes from 0.24 to 0.22. One may surprised by the fact that different error specification have opposed effect on a common parameter. This may come from the fact that the levels models concentrate on large prices, i.e., in and at the money options, while the logarithms models treat all option *equally* by concentrating on relative errors. Note that, when the B-S is allowed to be incorrect by the introduction of an error and the extensions, σ does not necessarily estimate the return standard deviation. It is only a free parameter which allows a functional form to fit better. This is specially true here since we use uniformative 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 part in the total model 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 compensated by the fact that σ is going up simultaneously. For the levels model, the changes in parameters do not appear as important. The values of the parameters for the logarithms and levels models can not be directly compared because of the different functional forms.

Finally, it is clear from any of the plots in figure 2 that the parameters of model 4 are estimated with a lot more uncertainty than the other models. This pattern is in fact true for all the 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 shows that model 4 with 14 parameters (3 levels of σ_η) is difficult to estimate precisely. In the presence of parameter uncertainty, additional parameters can be costly and there is a marked difference in uncertainty going from model 3 to model 4.

We now turn to the parameter σ_η. It is of primary interest since it represents the variability of the model error. It is itself a diagnostic of the average error size. the logarithms model error is relative while the levels 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 are clear from the inspection of the plots. First, on average models 2 and 3 significantly reduce model error from about 10.5% to down to 7%. Second there is strong evidence of heteroskedasticity. The standard deviation of model 2 error is 12% out-of, 4.5% at , and 2.5% in the money. In fact, the improvements brought by models 2 and 3 take place out-of and at the money. Models 2 and 3 actually exhibit higher standard deviation for in the money options. However the trade off seems easy to resolve in favor of 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%, and increase in the money from 1.7% to 2.5%.

Figure 3b reproduces these diagnostics for the levels model. Here the evidence of improvement for the homoskedastic models is very marginal. The estimation of heteroskedastic models demonstrates the heteroskedasticity of the error for the levels formulation as well. The pattern is opposite from the logarithms case as expected since these are dollar errors. In fact, 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 dollars. Model 2 and 3 improve the specification for out-of (8 to 3 cents) and in (15 to 13 cents) the money errors. They fail to do so for the at the money errors.

4.2 Hedge Ratios

We now document the ease of construction of the posterior distribution of deterministic functions of the parameters with the study of the hedge ratios. Asymptotic estimation based on the normality assumption for the parameters, would require the use of delta methods to obtain a approximate value of the standard deviation of a hedge ratio. It would then be assumed to be approximately normally distributed. Instead, we easily obtain draws of the posterior distribution of the hedge ratios by direct computation for each model.

Contingent claim models are used for two purposes, pricing and hedging. The previous section gave a first documentation of pricing errors by inspecting the estimates of the model error standard deviation. We want to see whether different models have different policy implications for hedging decisions. Consider the instantaneous hedge ratio Δ. Different models functional forms imply different functional forms for Δ. They are simply calculated by computing the derivatives of the entire model specified in equation (4).

A 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 logarithms models 2 and 3. Figure 4b computes the hedge ratios for the corresponding levels 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 implications, different from the B-S. The difference is statistically significant even though the larger number of parameters yield more spread out Δ distributions. For example, 5 day, 4% out of the money calls have a Δ of 15.4% per models 2 and 3, 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 logarithms B-S in figure 4a with its levels competitor of figure 4b. Their Δ's are not identical. This shows that the distributional assumption of the error alone has an effect on the model parameters.

Figure 4 and 4b show that the extended models have different policy implications with respect to the design of hedged portfolios than the basic Black-Scholes. Note that as noted 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 eliminate the possibility of such ratios. Only reject draws which imply ratios outside the range (0,1).

4.3 Biases

The extended models incorporate additional functions of moneyness and maturity with possibly a fair number of parameters. Rather than inspecting the values of 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 logarithms 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 line or surface this function is. Again a 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 logarithms 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 than that at 60 days shown on the left. The bottom left plot show the bias functions for model 3 at different maturities. There does not appear to be evidence that the bias function is very different at different maturities. Finally we plot the bias function versus time to maturity for models 3 and 4 at 3 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 Dec. 4 to 8, 456 quotes. The bottom plot is for the week of Dec. 11 to 15, 419 quotes. Figure 6 shows that the mispricing implied by a model changes slowly with the period. Parameters vary with time and this results in changing bias functions. Also, 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 consists in the computation of a statistic, e.g., autocorrelation, and a diagnostic based upon an asymptotic sampling distribution for this statistic under a null hypothesis. Bayesian residual analysis uses the exact posterior distribution of the statistic which is readily available as dicussed in section 3. Given a draw of the parameters, we compute for each observation the residual $\log C_t - \log m_t(x_t, \theta)$. Given the sample of draws of the parameters we have a sample of draws of (the posterior distribution of) the vector of residuals. This is the basis for residual analysis, see Chaloner and Brant (1988). Most common diagnostics follow immediately. For a given observation, we can compute mean and quantiles, or plot the histogram of the posterior distribution of the residual, conducting standard outlier analysis. In our case, an outlier may have the intuition 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 can look banal for model 2. Unlike the user of the basic model, the user of model 2 would not undertake any potentially dangerous 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 draws 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. Violations from the Black-Scholes framework, e.g. stochastic volatility, may induce autocorrelation in the pricing errors of the basic model if the option price data cover a span of calendar times. [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 residual or squared residual for different subsamples, for example, long and short maturity, in and out of the money. Before looking at our empirical application, we discuss the out-of-sample tests.

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

4.5 Dynamic Specification Test

The previous discussion pertained to the analysis of quotes or trades belonging to the sample used to estimate the model. We now turn to predictions made for claims not part of the sample. In the absence of 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 model comparison, leading to the selection of the model with the best predictive track record . Period by period, one can follow the predictive performance of a model, across all types of options, or separating by category, e.g., moneyness or maturity. Say that we have available at time t

$$\begin{cases} C_{t-\tau}, \tau = 0, ..., L-1 & \text{Previously observed market realizations} \\ P(C_{t-\tau} \mid \underline{y}_{t-\tau-1}) & \text{Previously formulated predictive densities} \end{cases}$$

These are the basis for simple dynamic model comparisons. The predictive densities can be formulated for several models. The means of the predictive densities $\mathrm{E}\Big(C_{t-\tau} \mid \underline{y}_{t-\tau-1}\Big)$ are then computed. Once $\mathrm{C}_{t-\tau}$ is known, the errors are recorded. For each model there are L realizations of the error. The models can be compared on the basis of their errors and mean squared errors. Time series plots of the error functions can be constructed and updated dynamically for the prediction error of each model. Similar plots of the moving average of the last L errors can be constructed. The same can be done for the squared errors. Finally the bias and squared errors can be computed for subsets of the predictions, e.g., moneyness, maturity.

The above tests concentrate on the computation of a mean as a point estimate. The mean of the prediction determines the direction of a strategy. The predictive density also quantifies the uncertainty of the prediction. This important aspect of the prediction process is often neglected because standard methods do not provide reliable predictive densities. Under risk aversion, the level of uncertainty around a prediction helps determine the amount input into the strategy. 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, it should cover the realization 50% of the time.

The above tests can also be directed to more *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 most useful 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 too large, 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 manage model error risk. Given the model error covariance structure, hopefully diagonal, standard portfolio theory can be applied in this setting 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 $\text{SE}_{t-\tau}, \tau = 0, .. , \text{L-1}$ can serve to rank competing models. [17]

4.6 In and Out of Sample Tests

We now implement some of the tests discussed in the two previous sections, in order to compare the competing models previously discussed. We concentrate on the logarithms heteroskedastic and homoskedastic models as well as the levels heteroskedastic models. Table 1 contains the within-sample tests for the estimation conducted on the first two weeks of december 1989. The parameters are reestimated each week, the errors are then aggregated over the two weeks. This represents a sample of 871 quotes. The first panel, 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 fell outside the bid ask spread. The rationale for this last subsample is economic. Predictions and fitted values that fall inside the bid ask spread may not be the basis of a trading strategy and are not as relevant as those that fall outside.

A quick glance at the first panel of table1 shows that average biases are far smaller than the RMSE's. Extended models appear to improve the bias of the basic models with the exception of the levels model for outside B-A spread quotes. The log. extended models have smaller RMSE than the B-S. Again, although there is evidence that the levels 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 logarithms models in terms of pricing errors. The residuals are relative pricing errors and are not directly comparable to the residuals of the levels models. The pricing errors are shown in the second panel of table1. Note the large pricing errors of model 4 which is not estimated reliably. The RMSE's clearly 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 was only a second moment effect, dispersion of the residual, in the logarithms, becomes incorporated in the mean when we take the exponential to compute the price. Second, the extended models improve upon the B-S significantly. Third, the levels models seem to have marginally better RMSE's, and less bias than the logarithms models. This does not mean that everybody should prefer the levels models. Given a fixed amount to invest, relative error may be a more relevant criterion than dollar error.[18]

The third panel of table 1, distribution analysis, documents the specification of the predictive density distribution. The in-sample predictive density is obtained as discussed in section 2. The first column shows the percentage of quotes which mean of prediction falls outside the B-A spread. The extended models bring a sizeable reduction in this number. It is below 20% for models 2. The third column is the specification test where we compute the percentage of quotes which fall inside the interquartile range. The levels 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. It represents the cover rate when we use, not the predictive

[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.

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

density, but the fit density. The fit density is obtained by drawing from the parameters but setting the model error to zero. This is consistent with studies which have tested the B-S model only allowing for parameter uncertainty, see Lo (1986). As per column 2, one would soundly reject any of the competing models. Their 50% intervals cover the true value only 2% of the time for the B-S, 13% for levels 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 an inappropriate basis for the test. When one uses the predictive density, even the B-S is very well specified for this test.

The final columns of the panel deal with the problem of the intrinsic bounds. No model should generate negative call values, bound B_1=0, of values below B_2=S-PV(X). The logarithms models are immune from the negativity problem. We wanted to know the extent of the problem for the levels models. The column entitled B_1 shows the percentage of quotes which predictive density implies a probability of negative call value larger than 0.1%. There are no more than 3% of such quotes for model 2, 0.3% for model 3. Consider the second bound. Actually 20% of the quote midpoints and 3 % of the asks violate the bound. We left these points in the estimation sample. We computed the number of observation such that the first quartile of their predictive density violated the bound, only considering observations which bid did not violate the bound. After the necessary correction for heteroskedasticity, the logs. and levels models were similar in this respect, see last column. Less than one in a hundred predictions had this undesirable property.

The previous results, even the predictive densities, were in-sample computed for 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 will not detail them. Consider the RMSE's. The first obvious result is that now the extended models do not improve the B-S out of sample RMSE. This is the case for residuals as well as pricing errors, levels or logs. Second, these out of sample errors are close to the within sample numbers of table1. There has not been a significant deterioration of performance going from within to out of sample. Looking more closely at the outside B-A spread RMSE's, one conclude that 1) the B-S have similar performance in and out of sample while the performance of models 2 and 3 has deteriorated to worse than the B-S and 2) the levels models seem to have deteriorated more than the logarithms models.

In any case, this is in sharp contrast with the results of Dumas, Fleming, and Whaley (1995) who report a complete break down of their models in out of sample tests while they had a quasi-perfect fit in sample. This is most entirely due to the fact that we allow for a model error in our estimation. This model error has properties which do not deteriorate quickly out of sample. We do find however that the Black-Scholes 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 shows consistency with the in-sample results. 20% of the in sample means of predictions were outside the spread, 30% of the out-of-sample predictions are outside the spread. We break down the interquartile range (IQR) coverage ratios by in the money and out of the money. 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. The heteroskedasticity brings it down significantly.

5 Extensions

5.1 Updating

In most cases, one will want to reestimate the model regularly and keep predictions for a future time close to the time of estimatio n. This is due to the fact that most misspecifications are likely to result in time varying parameters. This is for example the case when volatility, hedging demand, liquidity are time varying and only accounted through the expansion and the intercept of the extended model. The estimation is then based on updated (daily) cross-sections of option prices. Index these cross-sections by t.

Technically the calibration of priors p_t to the desired mean and variance is easy. For $p_t(\sigma)$ and an inverted gamma prior, it is easy to choose ν_0 and s_0^2 to match the desired mean and variance. The same goes for σ_η, set ν_1 and s_1^2 to obtain the desired prior mean and variance. Finally consider β. We wish to formulate a prior $p_t(\beta) \sim N(\beta_0, \sigma_\eta^2 V_0)$. If we were to use the posterior of t-1, we could use the joint draws 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.

The more interesting question is: Do we want to just use the time t-1 posterior to build the time t prior. This would be the thing to do if the daily parameter estimates were following a random walk on a daily basis. Given the probable cause for the time variation of the parameters, it is unlikely to be the case. Indeed the most persistent of the sources of misspecification, volatility, is known to be stationary though strongly autocorrelated. Other sources of misspecification, changes in hedging demand, liquidity, etc.. are even less likely to induce non stationarity. Clearly, the information on previous, t-2, t-3, .., periods can be important. A natural candidate for the formulation of time t priors will follow from the time series analysis of the previous posteriors and a one step ahead forecast. There are two alternatives to this approach. First keep several panels of options in the likelihood and model the time series variation of the parameters. Second, extend σ by a function of proxies of time varying volatility. The MCMC estimator can accommodate these extensions.

We now document the performance of the competing models through a simplified updating scheme. For 8749 quotes from January 01 to March 30 1990, we reestimate the parameters every second trading day. Given the scant evidence of any improvement brought by the larger models, we restrict the comparison to the B-S and model 2. We allow for heteroskedasticity.

Figure 7 documents the time series variation of some of the parameters. The top left plot shows σ. The top right plot shows σ_η for the homoskedastic models. The parameters of model 2 show clearly more time variation than those of the B-S. The top right plot show that the model 2 error 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. Apart from one day toward the end of the period, the out of the money quotes $(\sigma_{\eta,1})$ have large errors than the in the money quotes.

We compute the pricing errors for models B-S and 2. They are summarized here:

In Sample Performance: Daily Reestimation

Criterion	B-S	Model 2
RMSE	0.31	0.18
MAE	0.18	0.12
RMSE out BA	0.49	0.31
Pred out BA	3200	2260

MAE is the mean absolute error. In sample, model 2 exhibits significantly smaller errors by all the above criteria. Further, there were 627 observations for which model 2 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 BA spread 1632 times, in the BA spread 4922 times. Figure 8 shows where model 2 gained over the B-S. We plot, for each model, the pricing error versus the moneyness. The gaps with no points close to the zero line is because we only plotted the predictions outside the B-A spread. Model 2, top plot, does not suffer from the overpricing of the out of the money quotes as the B-S does.

5.2 Error Specification

We now propose a more general modeling of the error consistent with the presence of intermittent mispricing. Of course, statistical formulation alone can not identify an outlier as being surely a market error. The basic intuition for our formulation is that in most every quote there is no market error, and η_i is the only error. In rare occurrences, an additional error ϵ_t with standard deviation σ_ϵ occurs. We have already indicated how to conduct a residual analysis which may identify this outlying observation. It is more logical, however, to incorporate the possibility of rare errors in the model being estimated. The resulting diagnostics are easier to interpret, e.g., probability of a given observation having the extra error. The estimation and prediction are also more reliable since they incorporate the existence of these errors.

The cost of this extension, the added burden on the estimation, has to be weighted against the likelihood that such errors are indeed present. We model the *market* error as

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

that is,

$$\eta_i + \epsilon_i \begin{cases} \sim & N\left(0,\sigma_\eta^2\right), \text{ with prob. } 1-\pi \\ \sim & N\left(0,\sigma_\eta^2+\sigma_\epsilon^2\right), \text{ with prob. } \pi \end{cases}$$

This formulation is similar to those of the switching literature, see Hamilton (1987), McCulloch and Tsay (1993) and others, with the transition matrix such that the probability of the future state is not a function of the current state. [19] The discrete nature of this formulation captures the belief that the additional error is zero most of the time. The estimation can be conducted in two ways. First, it may be based upon a set value of π, the unconditional probability of mispricing.

[19] This assumption can be relaxed and π made to depend on the information set. This would be more in the spirit of a model specification test.

Second, we can formulate a prior distribution for π and let the data update it. We demonstrate the second approach which subsumes the first.

We specify a prior for π , which can be tightly centered close to zero, yet allow possible departures. We choose a Beta distribution, which parameters generate a great variety of shapes on [0,1], from uniform to as tight as desired. The model can be thought of involving a state variable s_i equal to 1 (market error) or 0 (no market error). Given the 2^n possibilities for the entire state vector S, both traditional Bayesian and maximum likelihood analysis are very complicated. Also, the asymptotic approximation on which maximum likelihood analysis is based may be questionable for conventional sample sizes given the size of the parameter space.

Again the resolution of this problem is facilitated by the incorporation of Gibbs sampling algorithms [20] 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) for data augmentation. Given (S, π, σ_ϵ), the system is that already analyzed. The noises do not have equal variance, but can be standardized given σ_ϵ and S. Given (β, σ_η, S), we can find the posterior distribution of σ_ϵ. Appendix C shows that, for the priors formulated in equation (10), we can 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 sequence of cycles of draws from these conditional posterior distributions converges to draws of the marginal posteriors. Consider the first of these and recall that we do not draw directly from σ_η, but use the Metropolis algorithm. This estimator is a Markov Chain estimator which combines the Gibbs and the Metropolis algorithms. See Tierney (1991) for the conditions required for the convergence of some Markov Chains algorithms. The above algorithm can be shown to converge , see Jacquier, Polson, and Rossi (1994) for a proof. Direct draws can be made from the other N + 2 distributions. For each observation we obtain the posterior probability of a market error. We also obtain the posterior probability of a 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.

5.3 Unobservable Inputs

An estimator for the case where the stock price or the interest rate are not observable can be formulated in a similar fashion. This is of interest because it models the uncertainty introduced by the measurement error directly at the source rather than lumping it in an adhoc external

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

additive error. The resulting model behavior could then be very different. Specifically the input for the stock in the Black-Scholes formulation would be S_i with

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

where S_i^* is the observation in the database. Tight priors can be imposed on $\sigma_n u$ to reflect bid-ask spread limits for example. Given the small size of ν compared to the magnitude of S, the normality assumption is a reasonable start. Simulation evidence (Renault (1995)) shows that synchroneity errors can affect pricing in non trivial ways. This makes the above formulation a worthwhile extension.

6 Conclusion

We introduce a new method to formally incorporate model error in the analysis and implementation of contingent claim models. Given a model, the method allows us to estimate deterministic functions of the parameters, produce the residual of an in-sample observation to assess its abnormality, and produce the predictive density of an out-of-sample claim.

We applying this method to a sample of 10000 quotes on calls on TOYS'R US, and document the behavior of several competing models nesting the B-S. The competing extended models are justified as expansions of a model unknown or too costly to implement. We formulate the error in relative terms (logarithms models) and in dollar terms (levels models). We show that there is indeed evidence of Black Scholes mispricing and by some criteria, the extended models dominate the B-S within sample. They reduce root mean squared errors of pricing and residuals. The improvement is not limitless as models with12 parameters show severe degradation in performance due to the increase in parameter uncertainty. The extended models have different hedging and pricing implications than the B-S. We show that the failure to include model error in specification tests results in very severe biases toward rejection. The interquartile range of a predictive distribution which in fact covers the true value 50% of the time, would be wrongly believed to cover the true value 2% of the time, thus leading to a rejection of the model.

The Black-Scholes appears the most robust when it comes to out of sample performance. Many of the advantages demonstrated by the extended models disappear and they fail to dominate the Black-Scholes drastically. However, most models behave quite well in that their out of sample properties are not very different from their in sample properties. This insight on these models is possible because of the estimation technique which formally incorporates model error. Our results are in strong constrast with those of recent studies who do not allow for model error and conclude to very different in and out-of sample properties. The extended models also imply possibly different hedge ratios than the basic model. Therefore they may have different implications if these difference remain important out of sample. More work remains to be done to test models in an updating setup, based on the method proposed here.

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, the risk free rate or volatility.

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 $x_i' = (\log m(\sigma, x_{1i}),\, x_{2i}')$, $\mathrm{X}' = (x_1, ..., x_N)$, $\beta' = (\beta_1, \beta_2')$, and $\mathrm{Y}' = (\log C_1, ..., \log C_N)$. 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

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

and

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

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

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

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 $\nu s^2(\sigma,\beta) = (Y - X(\sigma))'(Y - X(\sigma))$. 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 the Metropolis algorithm that does not require the computation of integration constants. See Devroye (1987) for the accept-reject method, Metropolis et al. (1953) and Tierney (1991) for the Metropolis algorithm.

The Metropolis algorithm nests a simpler algorithm, the accept/reject, 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 areas, we do not draw often enough from q, and the sample of draws does not reflect the actual 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. Unlike for the accept-reject algorithm, dominance everywhere is not needed, so we have more choice 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 still goes through. 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 in about 10 evaluations of the kernel. The function is well behaved and the optimization is fast. 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 about 10 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$\star$/cq is close to 1 almost everywhere. The intuition of the ratio p$\star$/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 $s_i = 1$. 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

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.

DeGroot, M., (1970): *Optimal Statistical Decisions*, New York, McGraw-Hill.

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.

Geweke, J.F., (1989): "Bayesian Inference in Econometric Models using Monte Carlo Integration" *Econometrica*, 57, 1317-1340.

Geweke, J.F., (1994): "Bayesian Comparison of Econometric Models" Working Paper 532, Federal Reserve Bank of Minneapolis.

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

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.

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., and JARROW, R., (1996): "A Portfolio Approach to Model Error Reduction in Hedging", Unpublished Manuscript, JGSM, Cornell University.

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", Working Paper, Cornell University.

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

JARROW, R.A., and RUDD, A., (1983): *Option Pricing*, Dow-Jones Irwin, Homewood IL.

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.

MELINO, A.,, and TURNBULL S. (1990): " Pricing Foreign Currency Options with Stochastic Volatility", *Journal of Econometrics*, 45, 239-265.

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. (July 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.

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.

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

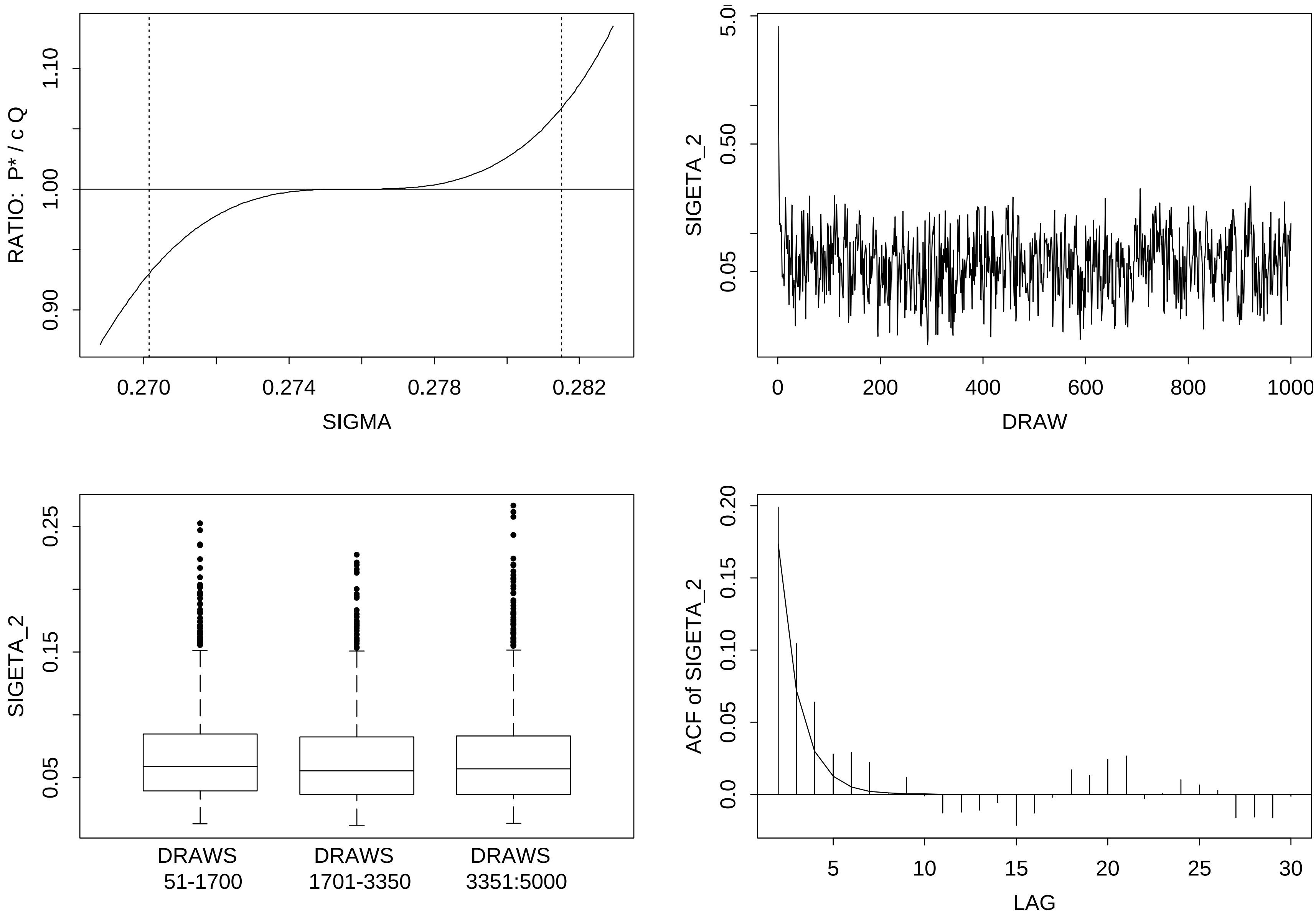


Figure 2: Sigma, Intercept and Slope, Logarithm and Level Models
TOYS'R US, 456 quotes Dec. 04-08, 1989

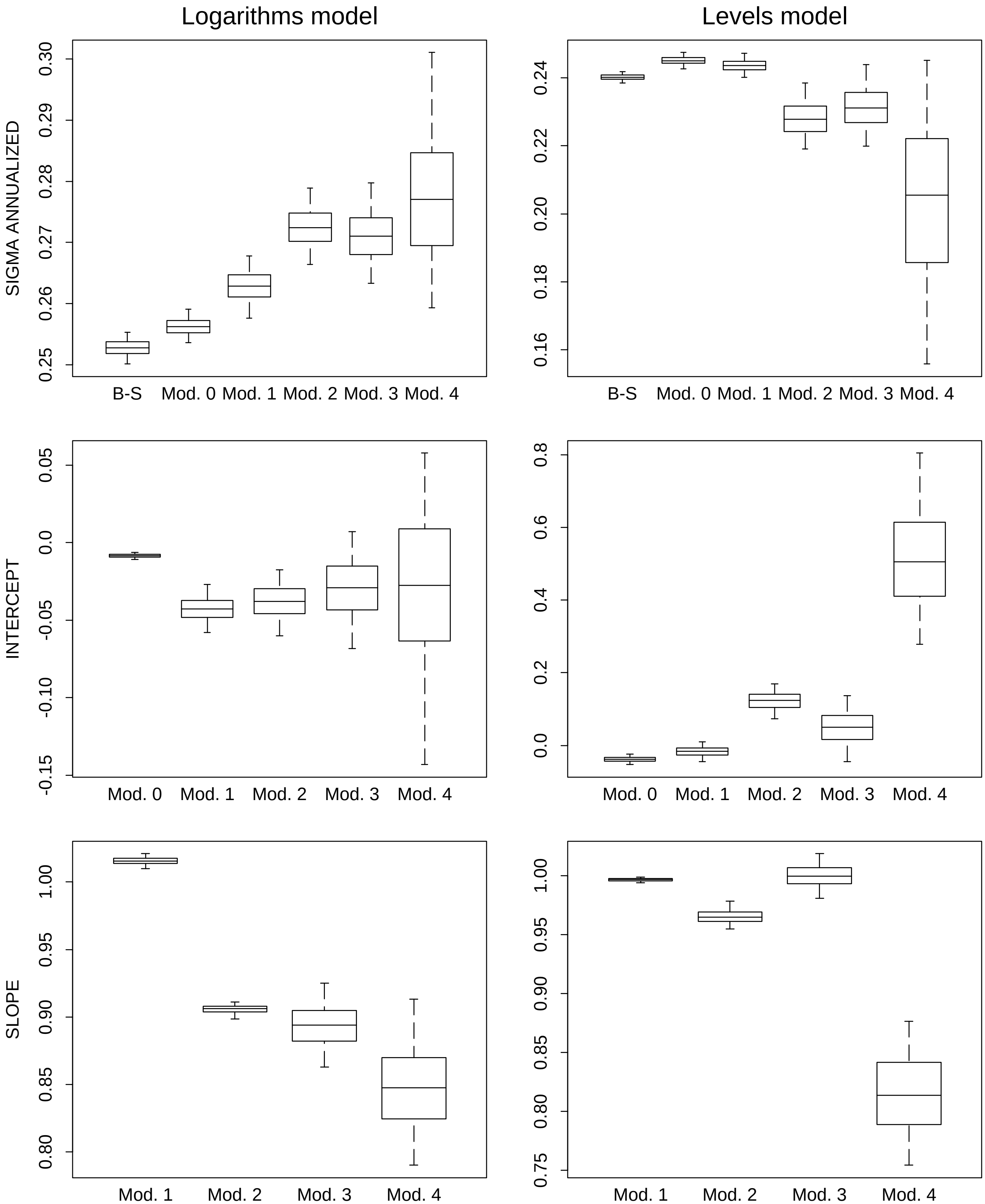


Figure 3a: Model Error and Heteroskedasticity, Logarithms Model
TOYS'R US, 456 quotes Dec. 04-08, 1989

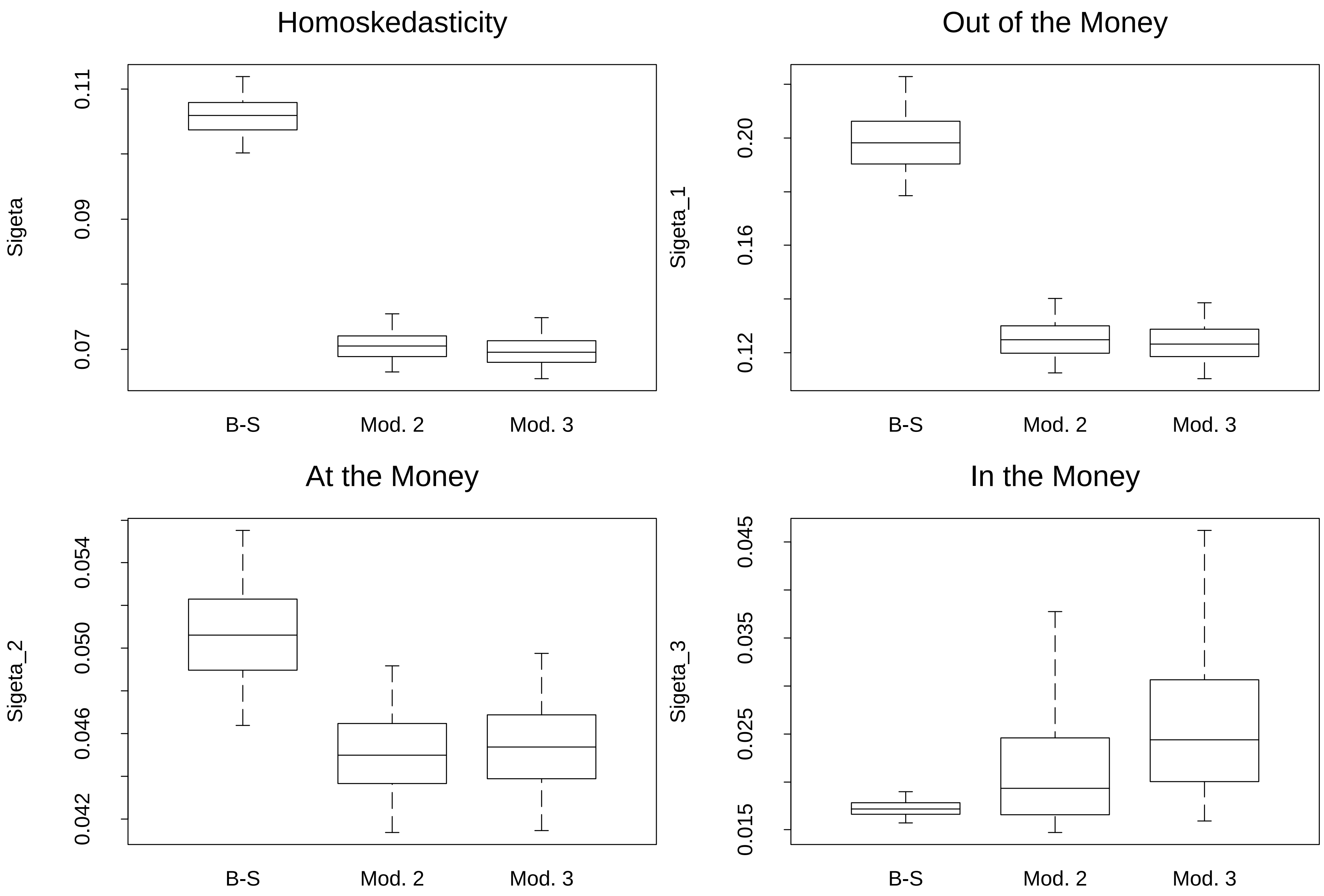


Figure 3b: Model Error and Heteroskedasticity, Levels Model
TOYS'R US, 456 quotes Dec. 04-08, 1989

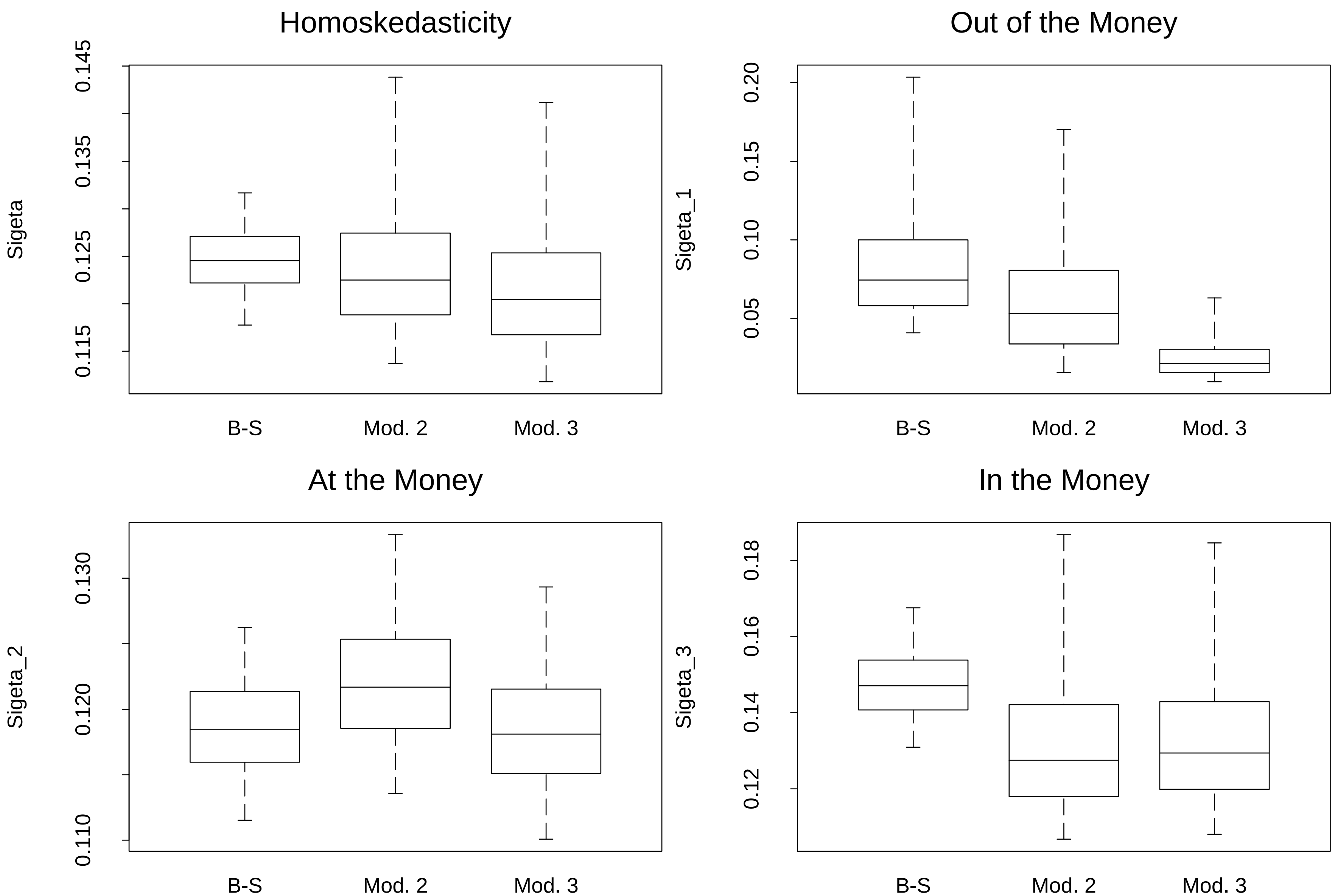


Figure 4a: Hedge Ratios for 3 Logarithms Models, TOYS'R US, Dec 04-08 1989: 452 quotes

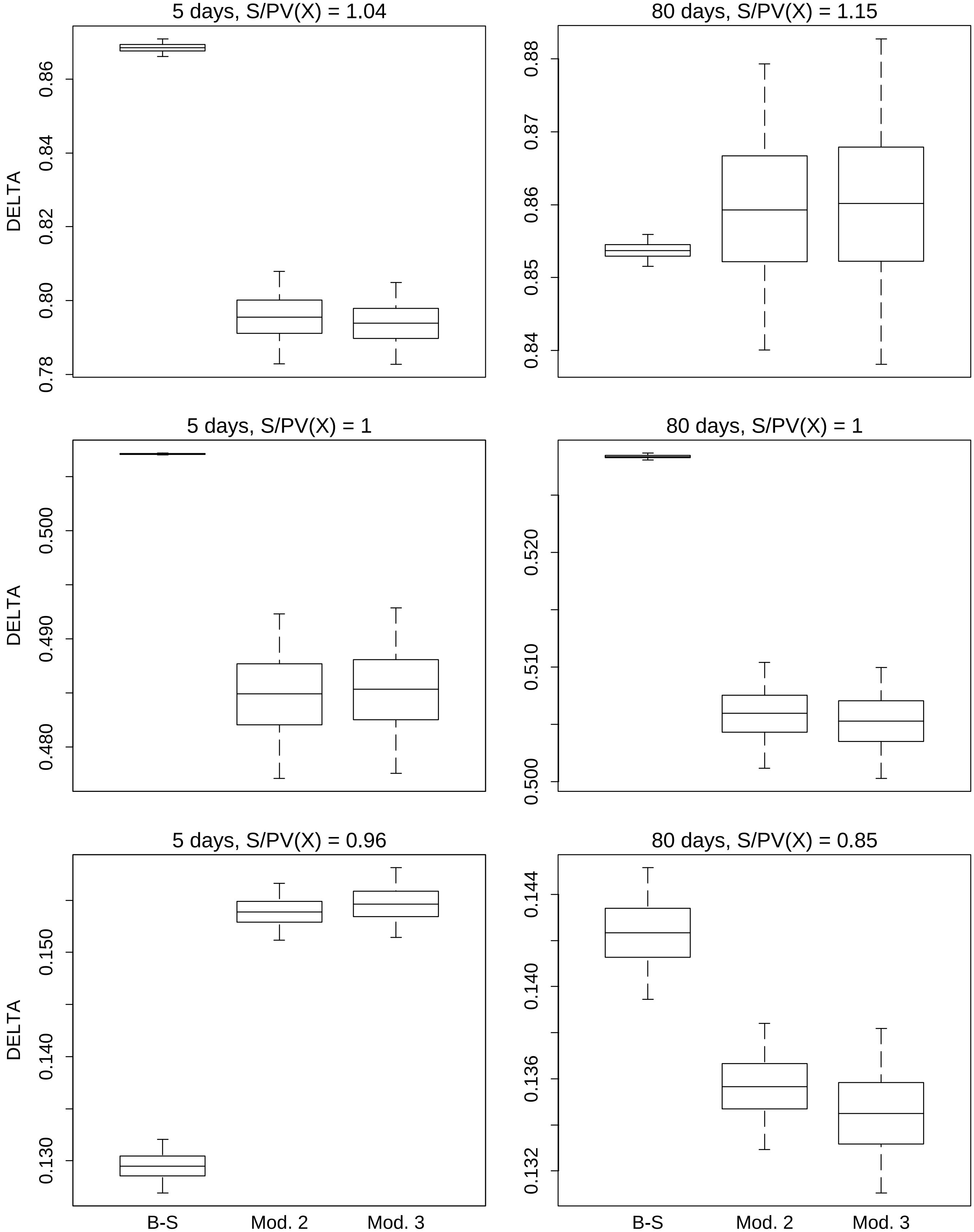


Figure 4b: Hedge Ratios for 3 Levels Models, TOYS'R US, Dec 04-08 1989: 452 quotes

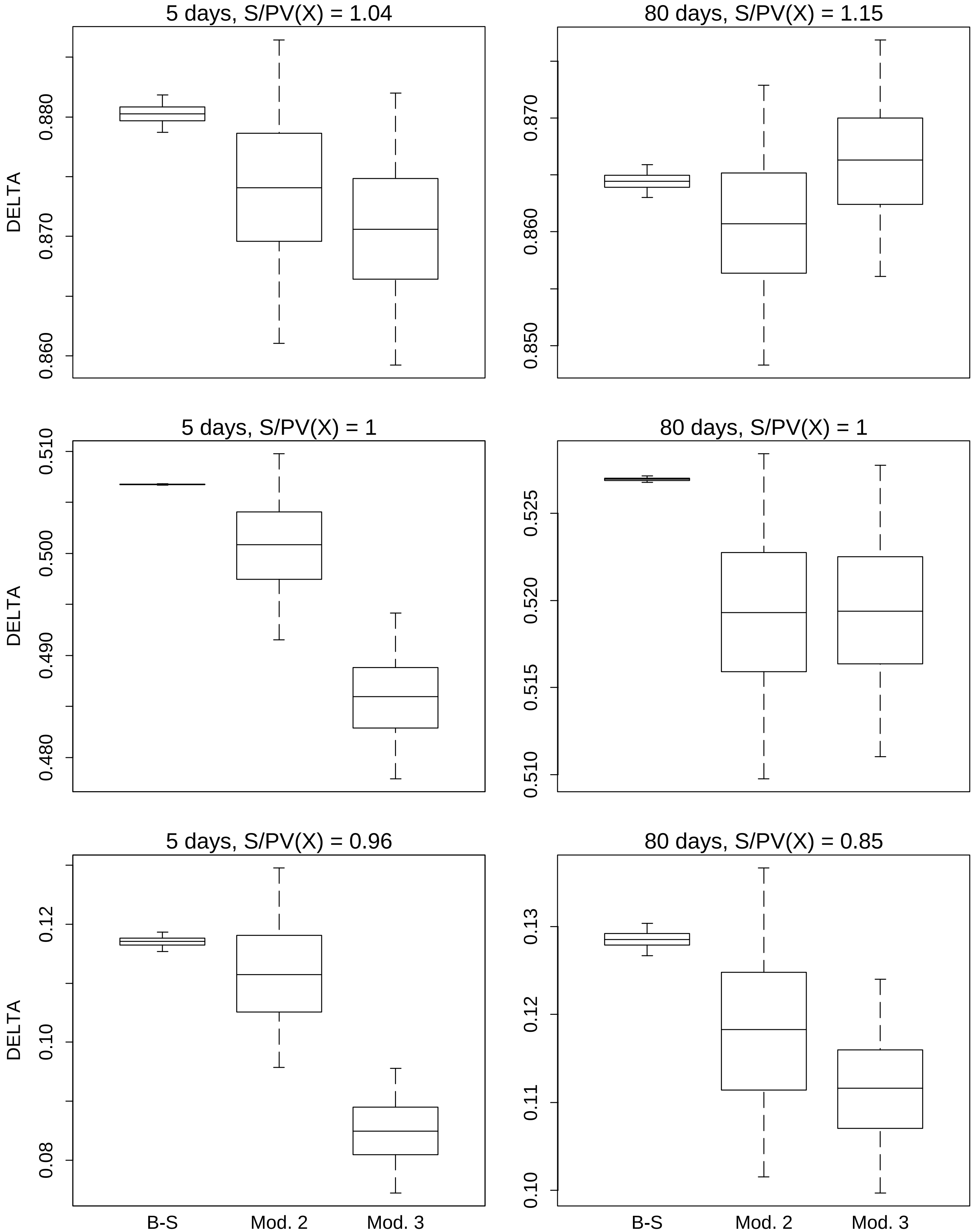


Figure 5: Money and Maturity Biases, Logarithms Models, TOYS'R US, Dec 11-15 89, 419 Quotes

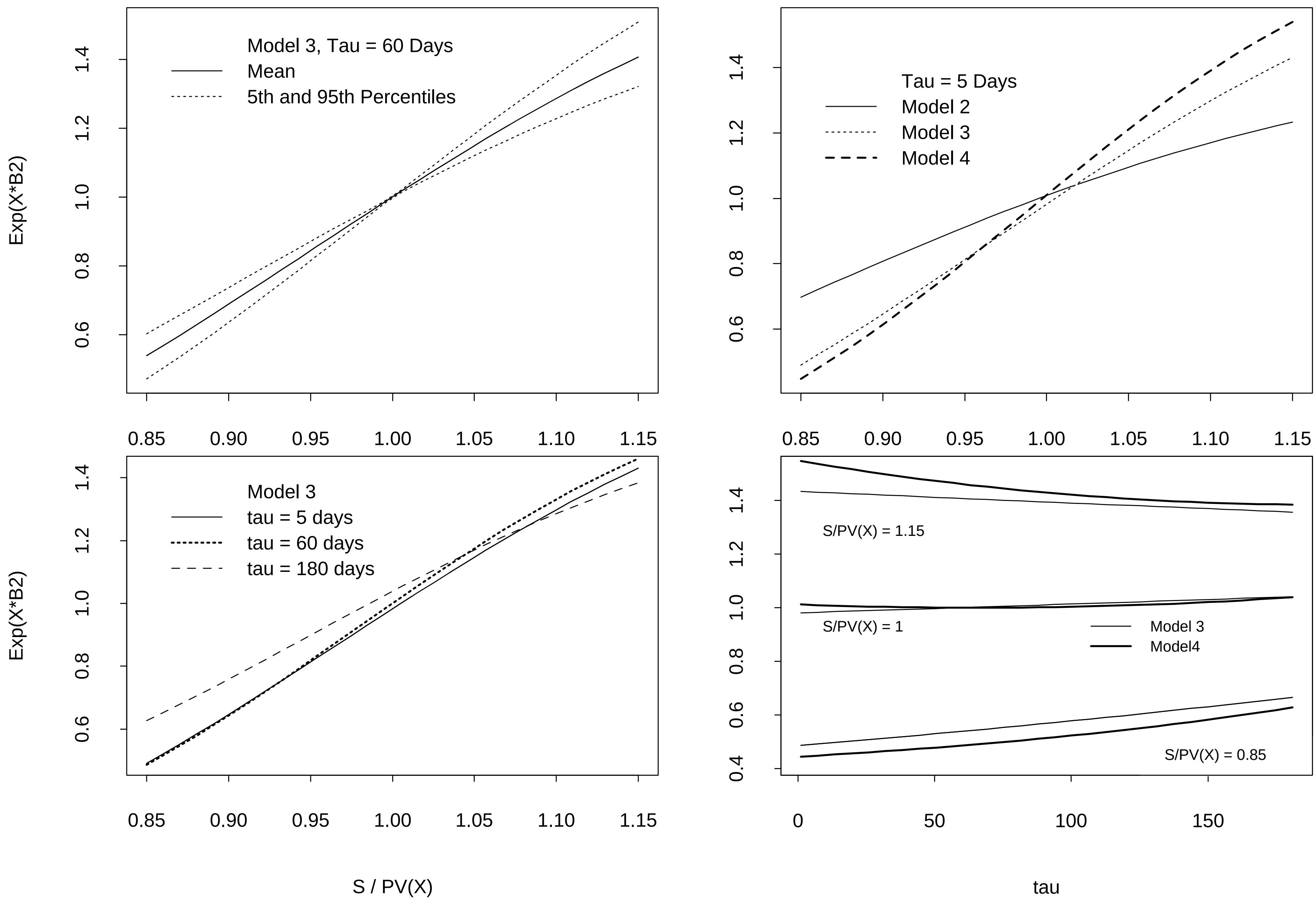
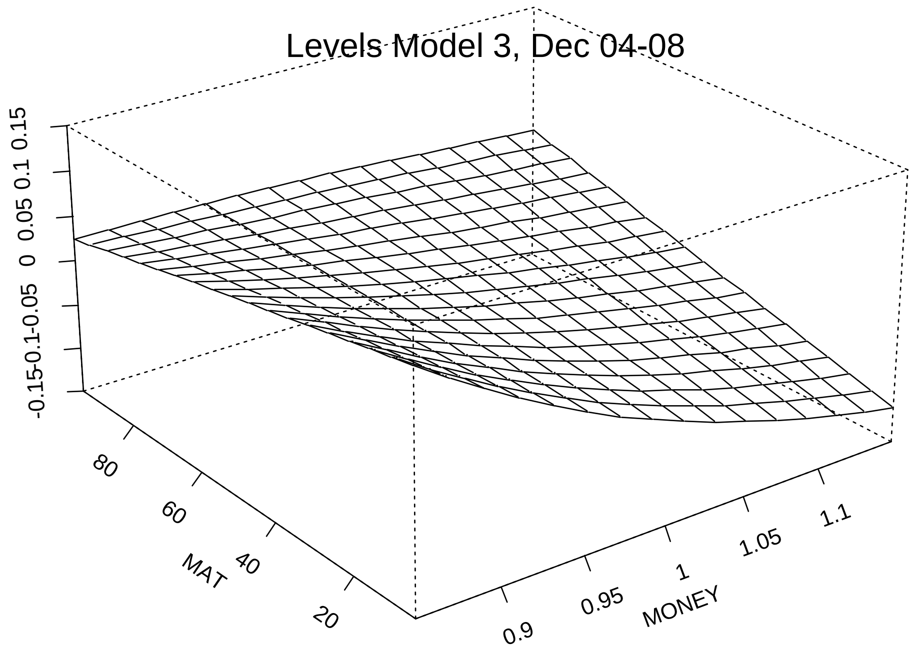


Figure 6: Bias Surfaces, TOYS'R US, Dec 89, 1927 Quotes



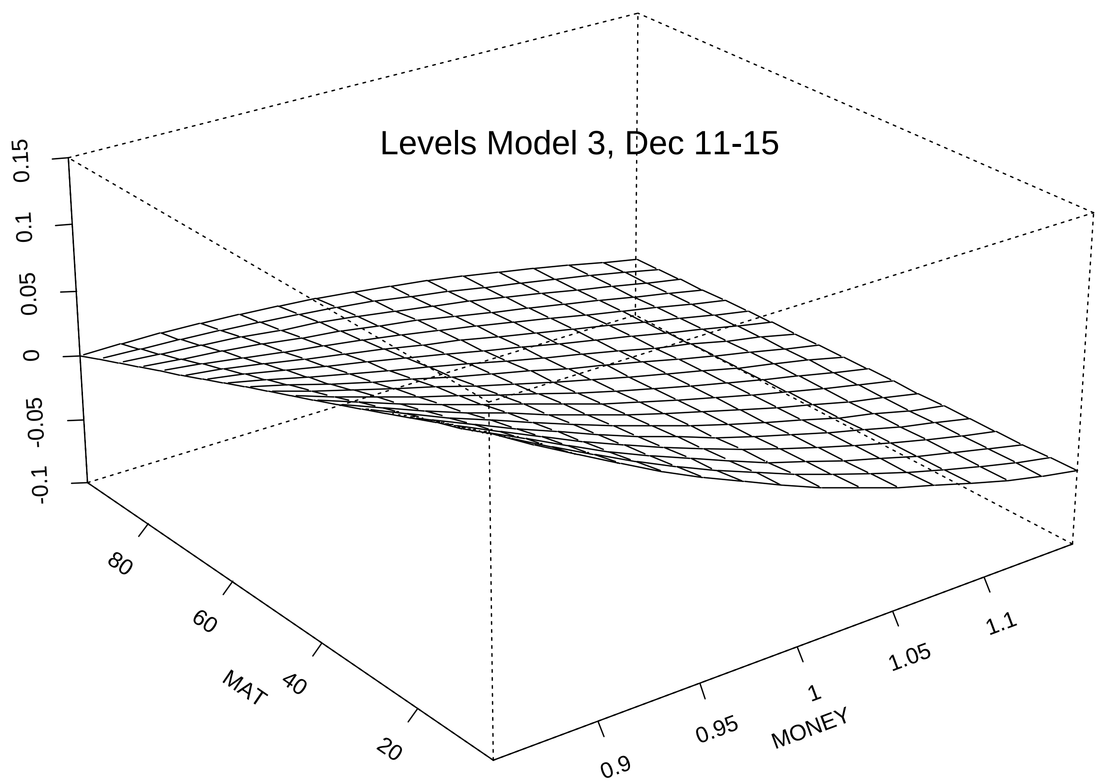


Figure 7: Time Series Plots of Parameters, Logarithms Models, TOYS'R US, January-March 1990

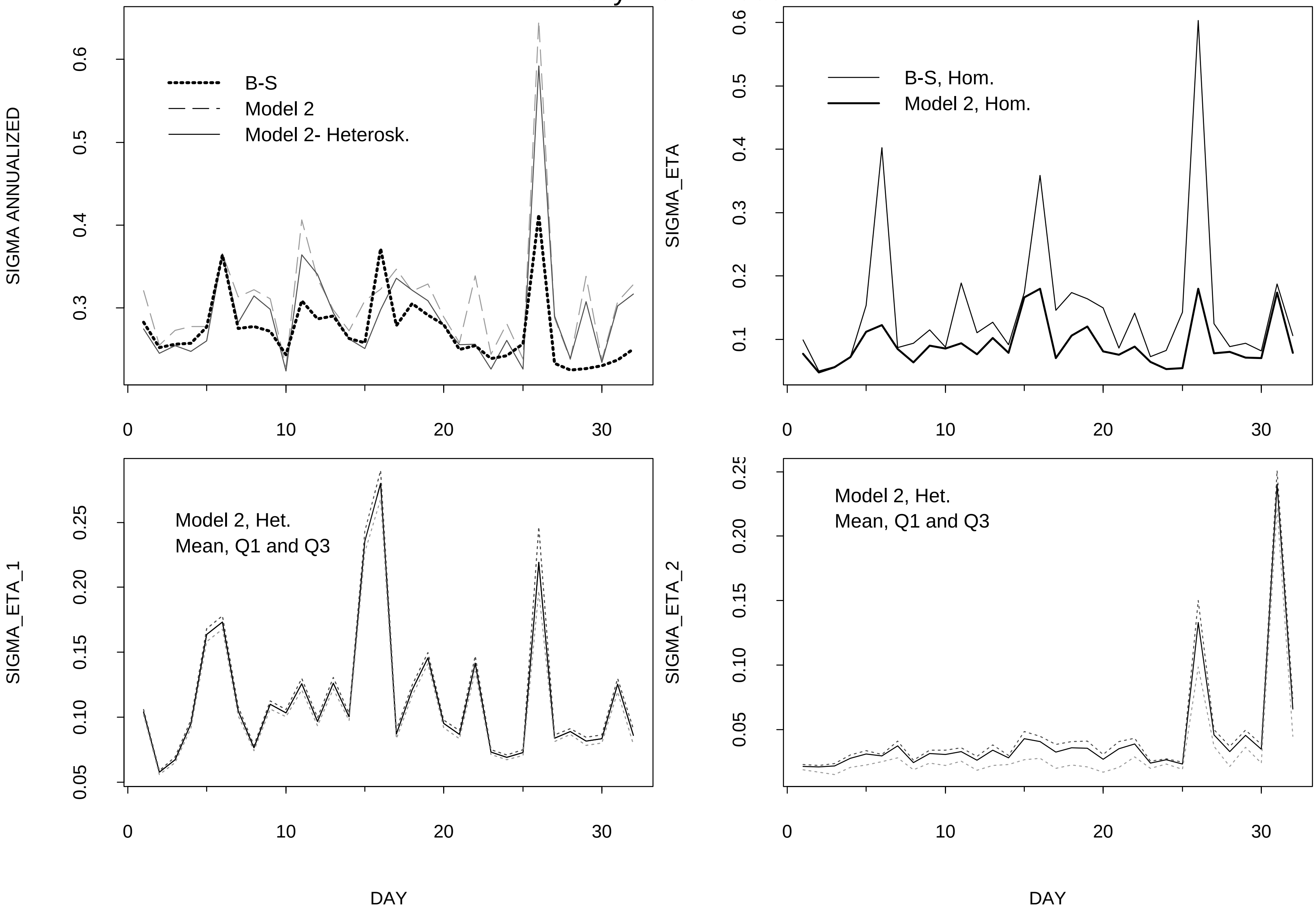
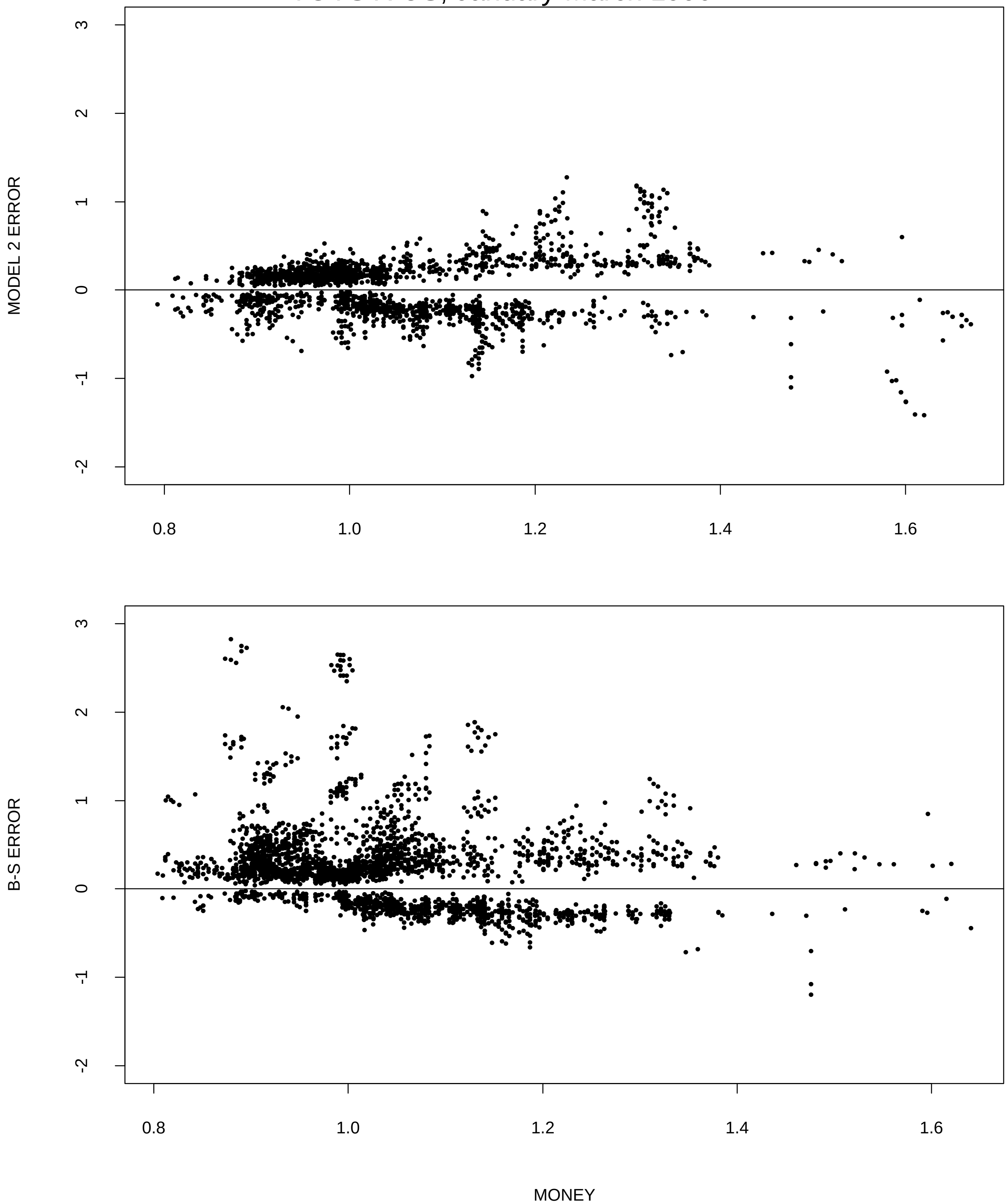


Figure 8: Pricing Error vs Moneyness, Logarithms Models,
TOYS'R US, January-March 1990



Liste des publications au CIRANO *

Série Scientifique / *Scientific Series* (ISSN 1198-8177)

* Vous pouvez consulter la liste complète des publications du CIRANO et les publications elles-mêmes sur notre site World Wide Web à l'adresse suivante :
http://www.cirano.umontreal.ca/publication/page1.html