A financial metric for comparing volatility models: Do better models make money?*

Toby Daglish[†] John Maheu[‡] and Tom McCurdy[§]

June 2005

Abstract

This paper proposes a fully-specified equilibrium approach which provides both financial and utility metrics for comparing alternative beliefs about the conditional distribution of a stock price. In this paper, we focus on differences in volatility dynamics which are inputs to investors' assessments of a derivative security. We construct equilibria in which different investors (models) trade a derivative that is sensitive to the volatility of the underlying asset. Our approach can be used to assess the economic importance of parameter uncertainty and model misspecification. Examples using simulated data demonstrate that informed investors (investors with better models) make money and utility gains against uninformed investors. Parameter uncertainty and model uncertainty, in general, both lead to lower profits. Using historical data, we find that GARCH models make significant gains against constant and exponentially weighted moving average specifications of volatility.

*The authors thank Christophe Pérignon, Tony Berrada and participants in the Northern Finance 2003 meeting. Daglish acknowledges support from a University of Iowa Old Gold fellowship. Maheu and McCurdy acknowledge support from the Social Sciences and Humanities Research Council of Canada.

†Corresponding Author: Henry B. Tippie College of Business, University of Iowa, 108 John Pappajohn Business Building, S306, Iowa City, Iowa 52242-1000, tel: (319) 335-1026, fax: (319) 335-3690, e-mail: toby-daglish@uiowa.edu

‡Department of Economics, University of Toronto, e-mail: jmaheu@chass.utoronto.ca

§Joseph L. Rotman School of Management, University of Toronto, and CIRANO, e-mail: tmccurdy@rotman.utoronto.ca

However, it is difficult to discriminate among the various GARCH models we consider. Our results point to the value of modeling time varying volatility, and the smaller gains that additional features such as leverage and fat tails provide.

1 Introduction

One of the great successes of empirical finance has been the time-series modeling of volatility for security returns. The stochastic nature of volatility, and its persistence properties, have significant implications for risk management, portfolio choice and asset pricing. In view of this, it is not surprising that in the past two decades a huge literature has developed that seeks to model the dynamics of volatility. The majority of this literature assumes that better models, as measured by a statistical metric, translate into better models for financial decision making.

However, there are several reasons why a statistical comparison of models can provide a different ranking than one based on economic criteria. First, model forecasts are used as inputs into highly nonlinear objective functions in portfolio choice and asset pricing. Better forecasts of one statistical moment, such as volatility, do not necessarily translate into better diversified portfolios or smaller pricing errors. Second, results which are highly statistically significant can sometimes be of very small magnitude, and economically insignificant. Thirdly, conventional statistical approaches that understate or ignore parameter and model uncertainty can have costly implications for financial decisions (Barberis (2000), and Avramov (2002)). For these reasons, model comparisons based on economic criteria become attractive.

Existing work by Engle, Hong, Kane and Noh (1993) and Noh, Engle and Kane (1994) established that investors using GARCH volatility forecasts can make money by buying straddles when those forecasts anticipate higher volatility than that implied by the market prices. Furthermore, the profits are greater than those for an investor who uses an ARIMA process fitted to implied volatilities to implement a similar trading strategy. However, to price options they use the Black-Scholes model which is not consistent with the volatility process, making it difficult to disentangle the effects of pricing errors versus volatility forecast errors. ?) consider a range of criteria to rank volatility models including pricing options with a Black-Scholes model.

In a Markowitz portfolio framework, West, Edison and Cho (1993) show that agents who are investing in one risky foreign currency will do best (among the models they consider) by choosing a GARCH estimate for volatility. ?) show that different multivariate GARCH models can have very different effects on time-varying betas and hedge ratios. In a multiple risky asset framework, Fleming, Kirby and Ostdiek (2001) also find significant benefits to measuring time-varying covariance matrices. These analyses are restricted to the case of quadratic preferences and/or normal returns for assets.

In contrast to the existing literature, this paper proposes an equilibrium approach which provides both financial and utility-based metrics for comparing alternative beliefs about the conditional distribution of a stock price. Investors are differentiated by their (different) models of the conditional distribution of returns which is an input into their assessment of the value of a traded derivative. We consider location-scale distributions and focus on alternative forecasts of the volatility dynamics associated with the underlying stock returns.

The equilibria for the traded derivative among different investors (models) ensures that all prices are consistent with investors' preferences and heterogeneous beliefs about the volatility process. This involves calculating each investor's expectations which will be specific to his or her assumed conditional distribution for the underlying stock price.

In order to allow a sensitivity analyses of economic gains to different risk aversion levels, we use a preference-based approach to derivative pricing (for example, Garcia, Luger and Renault (2001), Garcia, Luger and Renault (2003), Jackwerth (2000), and Rosenberg and

Engle (2002)). In particular, we consider investors with identical preferences but heterogeneous beliefs concerning the volatility of the underlying security. To study the gains that different models of the conditional variance provide, our empirical examples focus on a butterfly spread.[1] Butterfly options, which are bets on the volatility of the underlying, have high vegas, but, when at the money, have small delta. Differences in profits and utility, will reflect differences in the investors' (models') forecasts of volatility of the underlying.

In our approach to model comparison, investors trade short-term butterfly contracts between themselves and hold them to expiration. Using simulated data, we can separate the gains due to having the correct model from gains due to accurately estimating parameters. That is, we can measure the effects of model misspecification versus parameter uncertainty. Furthermore, since only the underlying asset price is needed to calculate gains, we can compare the performance of the models using actual stock price innovations. Therefore, our approach only requires historical price data on the underlying to conduct both a statistical and economic ranking of volatility models.

We present several examples of our methodology. First, we show that our financial metric is sensitive to model misspecification. That is, informed investors (informed in the sense of knowing the true model) have higher utility and make money against uninformed investors. All of our examples demonstrate significant gains to modeling time-varying volatility. However, the choice of parameterization for the time-varying volatility depends on the precision of the parameter estimates. In other words, there is a tradeoff in terms of model risk between more sophisticated models and parameter uncertainty.

A large literature has documented that higher volatility tends to follow negative returns. We investigate the importance of modeling this so-called leverage effect using the GJR-GARCH

[1] Although straddles are traditionally regarded as the standard position for speculating on an asset's volatility, a butterfly has limited liability for the writer, which makes the calculation of our model equilibrium straightforward.

parameterization proposed by Glosten, Jagannathan and Runkle (1993). Based on a simulation example, we show that if one has sufficient data to minimize the parameter uncertainty, it pays to model the asymmetric volatility (leverage effect).

Finally, we use historical data to estimate the models and calculate the ex post gains from trading a butterfly written on IBM. Using IBM daily equity data, we find that GARCH models make significant gains against constant and exponentially weighted moving average (EWMA) specifications of volatility. That is, we do find value in going beyond a simple RiskMetrics-type model which uses an exponentially-weighted moving-average of past squared returns. However, it is difficult to discriminate among the various GARCH models we consider, even given large estimation samples. Our results point to the value of modeling time varying volatility, and the smaller gains that additional features such as leverage and fat tails provide. This is in direct contrast to statistical measures that favor these additions.

In summary, this paper proposes a new approach to comparing time series models of volatility. There are several benefits to this approach. First, our criteria is based on an equilibrium in which investor trade with one another based on their respective preferences and knowledge of the return process. In this sense, the equilibrium is consistent with our assumptions and avoids the use of incorrect pricing kernels such as Black-Scholes. As a result, it is straightforward to compare models with different time-varying volatility dynamics as well as fat tails and asymmetry. Models can be ranked based on either profits or ex post utility. A significant advantage of our method is that we do not need to collect and sort option data to evaluate the models, rather we only need the underlying return data that the models are estimated from. This means that models can easily be compared over different time periods and assets. Finally, our approach is amiable to simulations. This allows us to study and separate estimation and model risk, something that cannot be done without a fully specified economy.

This paper is organized as follows. Section 2 introduces the trading environment and market equilibrium in which different investors buy and sell derivatives. Section 3 presents some examples which demonstrate how the market can differentiate between model performance. Section 4 contains the results for model comparisons using simulated data which allows us to separate the effects of parameter uncertainty from model misspecification. Section 5 applies the same comparisons to historical return data for IBM. Section 6 concludes.

2 Market Equilibrium

Investors have heterogeneous beliefs about the conditional distribution of the stock price. In this paper, we focus mainly on differences in volatility dynamics.[2] These volatility forecasts are used as inputs to the investors' assessments of the value of the derivative written on that underlying security. As detailed below, the derivative that investors price and trade is a butterfly which is sensitive to the volatility of the underlying. We are interested in comparing the economic differences from alternative models of the return distribution that investors use.

To illustrate our approach, consider an exogenous financial security, such as a share of equity (stock), and a derivative written on that underlying security. To focus on the financial effects of alternative beliefs regarding the volatility of the stock price process, we only allow trading in the butterfly derivative and abstract from trading the underlying security [3] or a risk-free security.

2.1 Heterogeneous Beliefs about Volatility

Formally, define the information set available to investors at the end of time t as $\Phi_t = \{S_t, ..., S_1\}$ where S_t denotes stock price at the end of period t. Then, conditional on Φ_t,

[2] Other features of the conditional distribution such as tail thickness and the conditional mean can also play an important role for pricing.

[3] Therefore, the stochastic process directing the price of the underlying security represents the exogenous forcing process or exogenous uncertainty for the economy.

investor i's statistical model, used in all investment decisions for time $t+1$, is defined as

$$M_{i,t}(\hat{\theta}_i) \equiv \{f_i(S_{t+1}|\Phi_t, \theta_i)|\theta_i = \hat{\theta}_i\}. \tag{1}$$

$f_i(S_{t+1}|\Phi_t, \theta_i)$ is investor i's belief about the conditional distribution of the stock price S_{t+1}, and $M_{i,t}(\hat{\theta}_i)$ is investor i's statistical model or data generating process (DGP) for S_{t+1}, given Φ_t and $\hat{\theta}_i$ available at the end of period t. In practice, it will be convenient to model the distribution of returns (log differences of price levels). In this case the conditional distribution of stock prices is related to returns through $S_{t+1} = S_t \exp(r_{t+1})$, where S_t is known given Φ_t.

Each investor receives a nonstochastic endowment $W_{i,t}$ at the start of each period. Investor i's utility, $U_i(W_{i,t+1})$, depends only on wealth at the end of the period, which is determined by the realized gain from the investor's position in a one-period derivative. The derivative has a stochastic end-of-period payoff $g(S_{t+1})$. We abstract from trading for reasons of consumption smoothing and assume that end-of-period wealth, $W_{i,t+1}$, is fully consumed. This avoids trading for intertemporal transfer of wealth so that any trades will be motivated by heterogeneous beliefs about future distribution of returns, r_{t+1}.

2.2 Market Equilibrium

To find a market equilibrium we need to solve each agent's optimization problem to determine their net demand for the derivative. This involves calculating each investor's expectations which will be specific to his or her assumed conditional distribution for the underlying stock price. Then we need to find the equilibrium that clears the market at each point in time.

2.2.1 Agent i's Optimization Problem

Using information Φ_t, and taking the derivative price $p_{i,t}$ as given, investor i maximizes her expected utility by purchasing $q_{i,t}$ derivatives. The investor's problem is,

$$\max_{q_{i,t}} \quad E(U_i(W_{i,t+1})|M_{i,t}(\hat{\theta}_i), \Phi_t) \tag{2}$$

$$\text{s.t.} \quad W_{i,t+1} = W + q_{i,t}g(S_{t+1}) - p_{i,t}q_{i,t}. \tag{3}$$

The first order condition is

$$E(U_i'(W_{i,t+1})(g(S_{t+1}) - p_{i,t})|M_{i,t}(\hat{\theta}_i), \Phi_t) = 0. \tag{4}$$

Rearranging, we arrive at

$$p_{i,t} = \frac{E(U_i'(W_{i,t+1})g(S_{t+1})|M_{i,t}(\hat{\theta}_i), \Phi_t)}{E(U_i'(W_{i,t+1})|M_{i,t}(\hat{\theta}_i), \Phi_t)}. \tag{5}$$

As indicated by equation (3), $W_{i,t+1}$ is a function of W, $g(S_{t+1})$ and investor i's chosen investment level $q_{i,t}$ for a particular derivative price $p_{i,t}$. In this optimization problem the investor takes the derivative price as given in order to solve for an optimal quantity. However, investor i may redo this optimization at a range of possible $p_{i,t}$. Tracing out the locus of points $(q_{i,t}, p_{i,t})$ that satisfy (5) defines the investor's demand/supply for the derivative. Investor i's price for the derivative will be her expectation of the derivative's payoff, weighted by her marginal utility in the different states.

2.2.2 Solving for a Market Equilibrium

We define a competitive equilibrium as: all investors maximizing utility; supply equaling demand for the derivative, that is, $\sum_i q_{i,t} = 0$; and individual agent prices equaling the market price, that is, $p_t = p_{i,t}$, for all i. The important distinction between investors will be their evaluation of the distribution of the underlying security price which drives their expectations and resulting locus of optimal (q_i, p_i).

To solve for a market equilibrium, we need to calculate investors' expectations. Typically there will be no analytical results so we approximate expectations by Monte Carlo methods. For example, given a p and q and using the budget constraint in equation (3), the numerator in (5) can be calculated as

$$\begin{aligned} & E(U'(W - pq + g(S_{t+1})q)g(S_{t+1})|M_{i,t}(\hat{\theta}_i), \Phi_t) \\ \approx \; & \frac{1}{J}\sum_{j=1}^{J} U'(W - pq + g(s_{t+1}^{(j)})q_1)g(s_{t+1}^{(j)}) \\ & s_{t+1}^{(j)} \sim M_{i,t}(\hat{\theta}_i), \end{aligned}$$

with an analogous result for the denominator. This approximation can be made arbitrarily accurate by increasing J. In our calculations we set $J = 5000$.

We can solve for the competitive equilibrium as a system of nonlinear equations with N unknown variables where N is the number of investors. When N is equal to 2 those equilibrium variables will be $p_t = p_{1,t} = p_{2,t}$ and $q_{1,t} = -q_{2,t}$. We use Newton's method to solve for the equilibrium price and quantities. Note that if both investors shared the same beliefs, the equilibrium would be a zero trade equilibrium. In this special case, $W_{i,t+1} = W$ and hence $p_{i,t} = E(g(S_{t+1}|M_{i,t}(\hat{\theta}_i), \Phi_t))$.

2.3 Derivative Payoff

To study the economic gains that different models of the conditional variance provide we focus on a butterfly spread. Although straddles are traditionally regarded as the standard position for speculating on an asset's volatility, a butterfly has limited liability for the writer, which makes the calculation of our model equilibrium straightforward.

A butterfly spread involves positions in European options, each with different strike prices. Specifically, a butterfly is the purchase of one call option with strike $\underline{k}$, the purchase of another with strike $\bar{k}$, and the sale of $(k - \underline{k})/(\bar{k} - k)$ call options with strike k, where $\underline{k} < k < \bar{k}$. In the following we set $k = (\underline{k} + \bar{k})/2$ and denote the butterfly payoff as $B(S_{t+1}, \underline{k}, \bar{k})$, where S_{t+1} is the terminal price of the underlying stock. The payoff at expiration is

$$B(S_{t+1}, \underline{k}, \bar{k}) = \begin{cases} S_{t+1} - \underline{k} & \text{if } \underline{k} < S_{t+1} < k \\ \bar{k} - S_{t+1} & \text{if } k \leq S_{t+1} < \bar{k} \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Figure 1 is an example of the payoff structure. Included in this Figure are two possible terminal stock price distributions with different variances. An investor who views stock prices as having a large variance will value the butterfly at a lower price relative to an investor who expects a smaller variance. Therefore, the price of a butterfly spread will be sensitive to the maintained conditional variance.

2.4 Example Equilibrium

Much of the intuition for our approach to model comparison can be gleaned from a graphical examination of the demand curves of the two investors. Figure 2 shows an individual's net demand curve for a butterfly with payoff $B(S_{t+1}, 49, 51)$ when the current stock price (S_t) is \$50 and the initial wealth endowment is $W = \$10$. The agent in question has utility $U = \frac{W^{1-\gamma}}{1-\gamma}$, with relative risk aversion parameter $\gamma = 5$, and believes correctly that stock returns will be distributed normally with mean zero and variance 0.0003. In this case, $M_{1,t}(\hat{\theta}_1) = \{N(0, .0003)\}$ for all t. If the time period is defined as a day, the standard deviation corresponds

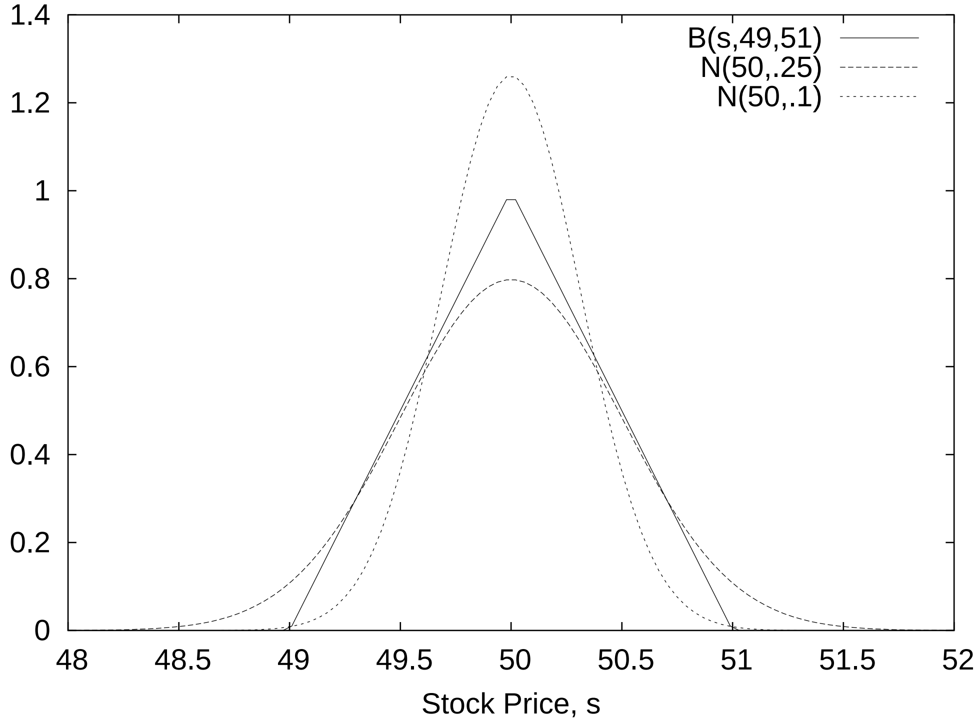

Figure 1: Butterfly Payoff $B(s, 49, 50)$ and Alternative Distributions for the Underlying

to an annual volatility of 27%. We label this individual as the informed investor. If all investors in the economy were identical to this agent, the market would clear when $q = 0$. This would lead to a price of $p = 0.416$ for the butterfly, as illustrated in Figure 2. Note that one advantage of the use of the CRRA utility function is that the equilibrium holdings (and therefore profits) will be homogeneous with respect to the wealth of the agent. Thus our results in this paper can be adjusted for a desired level of wealth by simple scaling.

Note the asymptotic behavior of the investor's demand curve. As the price of the butterfly approaches 1, the investor's demand approaches $-\infty$. This is because the butterfly has a maximum payoff of \$1 - hence at a price of \$1, the investor can make an arbitrage profit by selling the security. Conversely, the butterfly's payoff always equals or exceeds zero, so as the price of the butterfly approaches zero, the investor's demand tends to $+\infty$.

Figure 3 shows what happens when a second (uninformed) agent with different beliefs is introduced into the economy. Figure 3 illustrates the second investor's *supply* curve, that is,

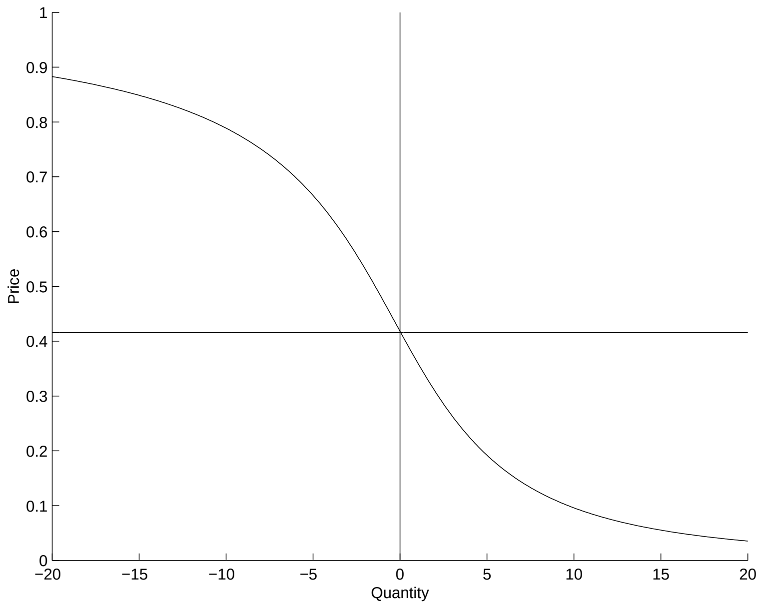

Figure 2: Demand curve for a butterfly with payoff $B(s, 49, 51)$ from an investor with CRRA utility and relative risk aversion parameter $\gamma = 5$. The current stock price is \$50 and the investor believes that the daily return on the stock will have a mean of zero and a variance of 0.0003 (corresponding to a yearly volatility of 27%). The horizontal line shows the market clearing price ($p = 0.416$) for the butterfly if all investors were identical and thus agreed about this distribution of returns for the underlying stock.

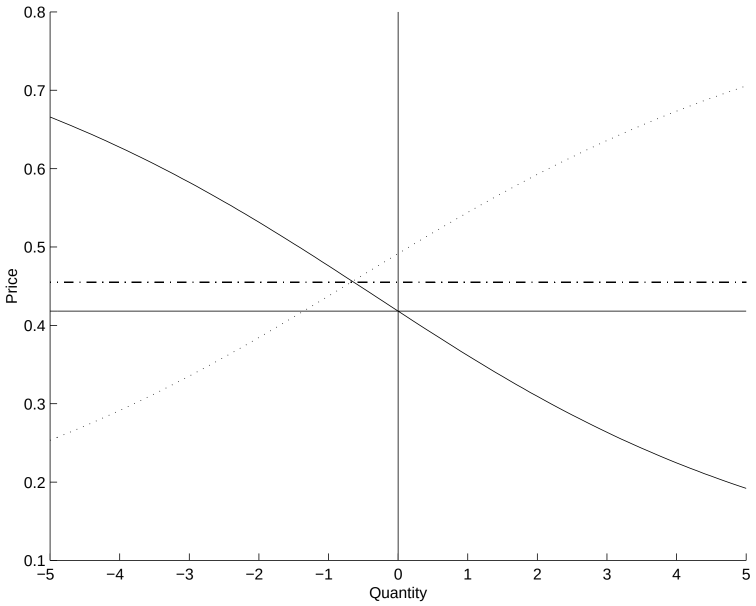

Figure 3: Butterfly demand and supply curves for two investors with different beliefs about the variance of the underlying stock. Both investors have CRRA utility with relative risk aversion parameter $\gamma = 5$. The current stock price is \$50 and the butterfly has payoff $B(s, 49, 51)$. The first investor's demand curve (solid line) is derived from the belief that returns of the underlying will be distributed $N(0, 0.0003)$. The upward sloping dotted line is the second investor's supply (negative demand) curve based on the belief that returns will be distributed $N(0, 0.0002)$. Equilibrium is characterized by the intersection of the demand and supply curves, at $p = 0.455$, $q = -0.636$.

$-q_{2,t}$. The second investor is identical to the first, except that he or she believes that the return variance will be 0.0002, that is, $(M_{2,t}(\hat{\theta}_2) = \{N(0, .0002)\}$ for all t). This investor values the butterfly higher than the first, since the butterfly's expected payoff is higher when prices are less volatile.

Equilibrium is determined by the intersection of the informed investor's demand curve (solid line) and the uninformed investor's supply curve (dotted line). Since the uninformed investor's belief about volatility of the underlying results in a higher valuation for the butterfly, in equilibrium the informed investor sells derivatives to the uninformed investor. As illustrated in Figure 3, this equilibrium corresponds to a price of 0.455 for the derivative, with the informed investor selling 0.637 butterflies to the uninformed investor. That is, the informed investor is shorting the security.

3 Model Comparison Examples

The following two examples illustrate how our financial metric for comparing alternative volatility models works and demonstrate in a controlled simulation setting that the results are what one would expect. That is, forecasting models which are based on the true data generating process for the underlying will make money against models that have imperfect information; and forecasts that track the true dynamics of volatility will typically perform better than those that assume that volatility is constant.

3.1 Better models make money

Consider the case of an 'informed' investor who knows the true process for the stock price. If the true return generating process is $N(0, 0.0003)$, then we should see this investor make money (on average) when trading against an investor who has incorrect beliefs about the data generating process. For instance, suppose the uninformed investor has a model $M_{2,t}(\hat{\theta}_2) = \{N(0, .0003\alpha)\}$, where α represents the proportional error this uninformed investor makes about the variance of stock prices. The expected profit $(E(q_{1,t}(B(S_t, 49, 51) - p_t)))$ for the informed investor is plotted in Figure 4 as a function of α.

Intuitively, when $\alpha = 1$, expectations are identical so no trade takes place and no profits are made. When $\alpha > 1$, the uninformed investor sells the derivative to the informed investor. Conversely, when $\alpha < 1$, the uninformed investor buys it from the informed investor. As α moves away from one, the error the uninformed investor makes about the variance of stock prices becomes larger so the profits earned by the informed investor increase.

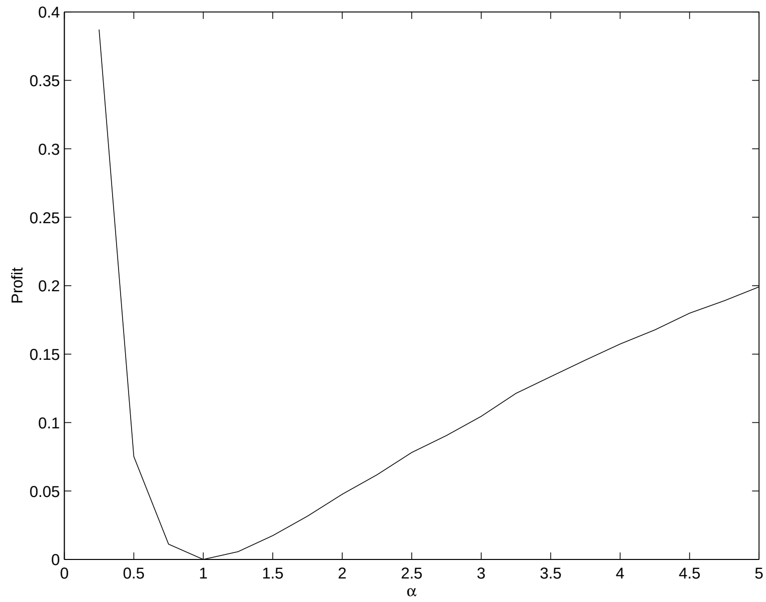

Figure 4: Expected profits for the informed investor as a function of α which is the proportional error the uninformed investor makes about the variance of stock returns. The informed investor uses model $M_{1,t}(\hat{\theta}_1) = N(0, 0.0003)$ which coincides with the true return generating process. The second (uninformed) investor uses model $M_{2,t}(\hat{\theta}_2) = N(0, 0.0003\alpha)$. The initial stock price is \$50 and the derivative payoff is $B(s, 49, 51)$.

3.2 Time-varying volatility

We now turn to an example that illustrates the importance of modelling time-varying volatility versus assuming it is constant when pricing and trading butterfly derivatives. Consider two investors trading butterflies written on an equity stock. As in the previous example, both investors have CRRA utility and per period endowment of \$10. A GARCH model is an empirically realistic data generating process for stock returns that display a changing conditional variance. Therefore, we assume that the exogenous stock price follows a process consistent with

$$r_t = 0.0005 + \epsilon_t, \quad \epsilon_t \sim N(0, h_t) \tag{7}$$

$$h_t = 0.000002 + 0.0579\epsilon_{t-1}^2 + 0.9376h_{t-1}, \tag{8}$$

where r_t is the continuously compounded return.[4]

The informed investor has model $M_{1,t}(\hat{\theta}_1)$ equal to (7)-(8). This investor is informed in the sense that he or she knows the true parameters for the DGP in equations (7) and (8), and hence the true distribution of the stock return at each time. On the other hand, the uninformed investor assumes returns have a variance equal to the unconditional variance associated with the process (8). This is $M_{2,t}(\hat{\theta}_2) = \{N(.0005, 0.0004)\}$. At each time, the investors trade butterflies with wings at 2% above and below the currently observed stock price which implies that the butterfly has payoff $B(s_{t+1}, 0.98s_t, 1.02s_t)$ in period $t+1$.

We simulate a price series of 5000 observations from (7)-(8) and calculate the sample mean and standard error of profits earned by the informed investor. We also calculate the average value of derivative holdings ($p_t q_{1,t}$) by the informed investor. Finally, we calculate the utility received by each investor. Two levels of risk aversion are considered, $\gamma = 2$ and 5.

The results are reported in Table 1. On average, for the less risk averse ($\gamma = 2$) case, the informed investor takes long positions, makes money, and hence achieves higher utility levels than the uninformed investor. Note that for this sample size of 5000, our estimate of the average profits is statistically different from zero (with a t-statistic of 14.6), as are the differences between the two expected utilities. Although returns are not well defined for short positions, the average profit of 0.1894 based on an average holding of 0.5744, suggests the informed investor is making substantial gains. For the more risk averse case of $\gamma = 5$, we see fewer butterflies traded, and less variable (although still positive) profits for the informed investor. In this economy, the two investors are more reluctant to bet on their beliefs.

We repeat the simulation with a data generating process that has reduced volatility per-

[4]The model has been calibrated to IBM daily stock returns so that our simulation results are comparable to the results in section 5 below in which we apply our model comparison technique to historical data.

Table 1: **Comparisons for Informed versus Uninformed Investors**

Risk aversion parameter	$\gamma = 2$		$\gamma = 5$	
High Persistence	Mean	s.e.	Mean	s.e.
Profits	0.1894	(0.0130)	0.0774	(0.0053)
Holdings	0.5744	(0.0164)	0.2290	(0.0067)
U_1	-0.0990	(1.355e-4)	-2.459e-5	(5.528e-8)
U_2	-0.1028	(1.392e-4)	-2.616e-5	(5.611e-8)
Low persistence	Mean	s.e.	Mean	s.e.
Profits	0.0061	(0.0021)	0.0024	(8.476e-4)
Holdings	0.0151	(0.0022)	0.0060	(8.658e-4)
U_1	-0.1000	(2.129e-5)	-2.498e-5	(8.519e-9)
U_2	-0.1001	(2.115e-5)	-2.503e-5	(8.462e-9)

Profits refer to the informed investor. Holdings refer to the dollar value invested in the derivative by the informed investor. U_1 is the informed investor's realized utility, while U_2 is the uninformed investor's realized utility. Numbers in parentheses are standard errors. Data for the underlying stock price is simulated from a GARCH model which has been calibrated to historical IBM data. The informed investor uses the correctly specified GARCH model while the uninformed investor uses the unconditional mean implied by the GARCH model. The traded butterflies have wings 2% above and below the current price. Low persistence refers to data simulated with a lower rate of persistence ($\alpha + \beta$).

sistence ($\alpha + \beta$),

$$h_t = 0.00009 + 0.0579\epsilon_{t-1}^2 + 0.7376h_{t-1}. \tag{9}$$

The parameters are chosen so that the unconditional volatility, and thus the uninformed investor's model, is identical to the previous simulation. Once again we assume the informed investor has perfect knowledge of the data generating process. The second panel of Table 1 shows similar results to the high persistence case, although the returns and variation of returns are smaller for the informed investor. Average profits for the informed investor are significantly different from zero (t-statistic is 2.9 for a risk aversion $\gamma = 2$ and 2.8 for risk aversion $\gamma = 5$). The size of the holdings show the informed investor makes significant gains against the uniformed investor.

We conclude that the gains to modeling volatility are larger when volatility is persistent and when investors have low risk aversion.

4 Model Risk

Building on the above examples, we now report the results from a simulation which compares the profitability of five different volatility forecasting models. We examine the effects of both model uncertainty and parameter uncertainty.

One stylized fact of volatility evolution is that good news has a different effect on market volatility relative to bad news. A simple parametrization of this effect is the GJR-GARCH model (Glosten et al. (1993)), that is,

$$r_t = \mu + \epsilon_t, \;\; \epsilon_t \sim N(0, h_t), \tag{10}$$

$$h_t = \omega + (\alpha + \delta I(\epsilon_{t-1} > 0))\epsilon_{t-1}^2 + \beta h_{t-1}, \tag{11}$$

where ω, α, β and δ are constants. Note that I is an indicator variable which is equal to unity if $\epsilon_{t-1} > 0$ and 0 otherwise. Therefore, $\delta > 0$ results in a model in which market volatility increases more in response to good news ($\epsilon_{t-1} > 0$), while $\delta < 0$ implies higher volatility forecasts in response to bad news ($\epsilon_{t-1} < 0$). Clearly, if volatility *did* behave in this manner, failing to incorporate the asymmetry term would result in poor volatility forecasts. Misspecified volatility dynamics, as we have seen, will lead to mispricing of butterfly securities, and we should see these agents lose money when trading against an agent with better volatility forecasts.

The data generating process (DGP) that we simulate as our reference model is equations (10)-(11) using parameters estimated from the IBM data discussed in section (5). We set the DGP parameters to those reported in the first row of Table 2.

In the following two subsections, we use this model to analyze the relative importance of

Table 2: Models for different agents

Model	μ	ω	α	β	δ	ν
DGP	0.311e-3	0.202e-5	0.08631	0.9378	-0.05672	
T-GARCH	0.397e-3	0.203e-5	0.05751	0.9381		87.115
GARCH	0.408e-3	0.202e-5	0.05737	0.9383		
EWMA	-0.548e-4		0.03000	0.9700		
Constant Vol	-0.548e-4	0.04594				

model choice, versus parameter uncertainty, for volatility forecasting and butterfly trading.

4.1 Parameter Uncertainty

In this section we consider parameter uncertainty assuming that agents have the correct model. Since the model is nonlinear, the impact on profits from parameter uncertainty will differ over parameters. Therefore we focus on the effect that uncertainty has on one parameter at a time.

The DGP is equations (10)-(11). Using this DGP, a path of 5000 returns and true conditional variances are generated. The informed agent is assumed to know the true DGP, and therefore the true conditional variance and return distribution. The uninformed agent, in contrast has all but one of the parameters correctly specified. The remaining parameter is estimated. As a result, misspecification in the return distribution comes only from parameter uncertainty associated with this one parameter. We then calculated the informed agent's ex-ante expected profit (using 20 000 simulations of possible outcomes for S_{t+1}) for each of the 5000 observations. These were averaged to obtain an estimate of the expected profit for the informed agent.

By varying the extent to which the uninformed agent misjudged the parameter, we are able to gain a measure of the cost of parameter uncertainty. To give an idea of the expected dispersion of the parameter estimate in finite samples, we simulated 100 observations of data and estimated the model parameter. Repeating this for 1000 repetitions gives an empirical

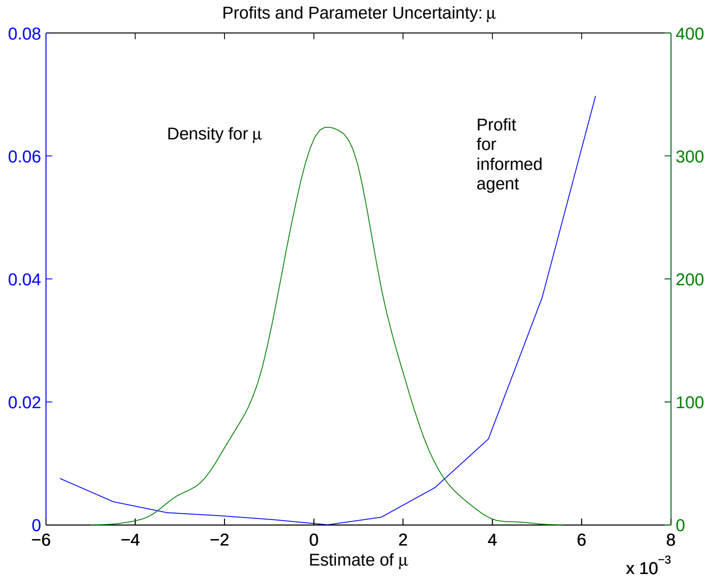

Figure 5: Profit to informed agent as a function of uninformed agent estimate of μ. The left axis measures the profit, while the right axis measures density of the uninformed agent's estimates.

distribution of the estimator in finite samples. The density plot for each parameter represents this distribution and provides a measure of the likely range (and thus uncertainty associated with) that parameter in small samples.

Figures 5 to 9 illustrate the effects of parameter uncertainty on profitability of butterfly trades. In each figure, we show the finite sample distribution for the estimator of the unknown parameter based on 100 observations; as well as the expected payoff to the informed agent, given the uninformed agent's estimates of the unknown parameter.

Although this analysis focuses on the marginal contribution of each parameter estimate to profits, it is difficult to separate some joint effects of parameters. For example, an agent who over-estimates β but underestimates α so that their unconditional variance is correct may well fare better than an agent who over-estimates β and correctly estimates α.[5]

[5] In joint estimation with 500 observations, we found the asymptotic estimates for the GJR-GARCH model parameters to be somewhat more dispersed than the densities shown in our figures 5 to 9. As such, these results represent an understatement of the costs of parameter uncertainty from the joint estimation of the GJR-GARCH

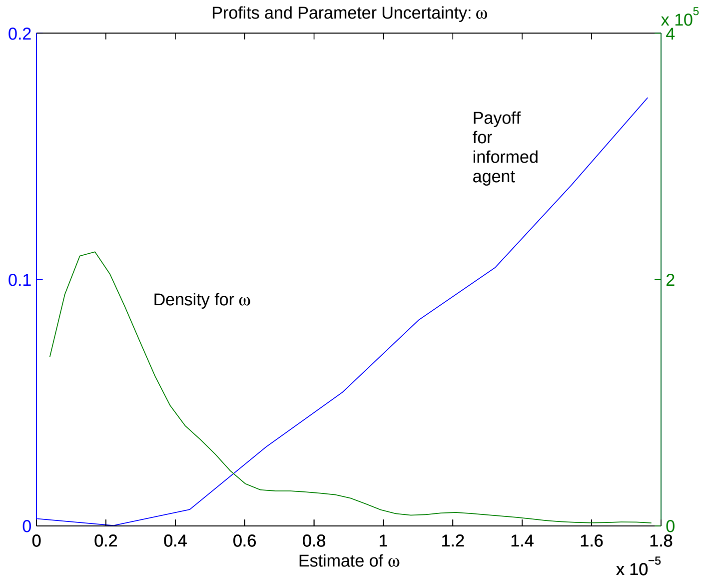

Figure 6: Profit to informed agent as a function of uninformed agent estimate of ω. The left axis measures the profit, while the right axis measures density of the uninformed agent's estimates.

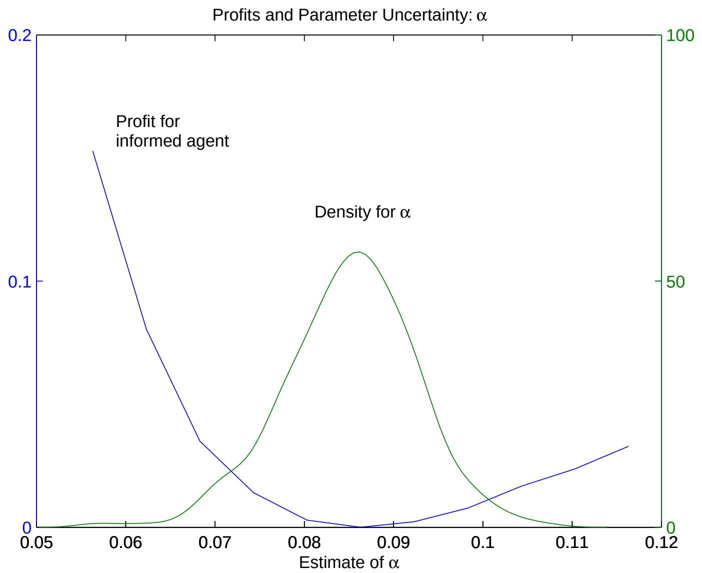

Figure 7: Profit to informed agent as a function of uninformed agent estimate of α. The left axis measures the profit, while the right axis measures density of the uninformed agent's estimates.

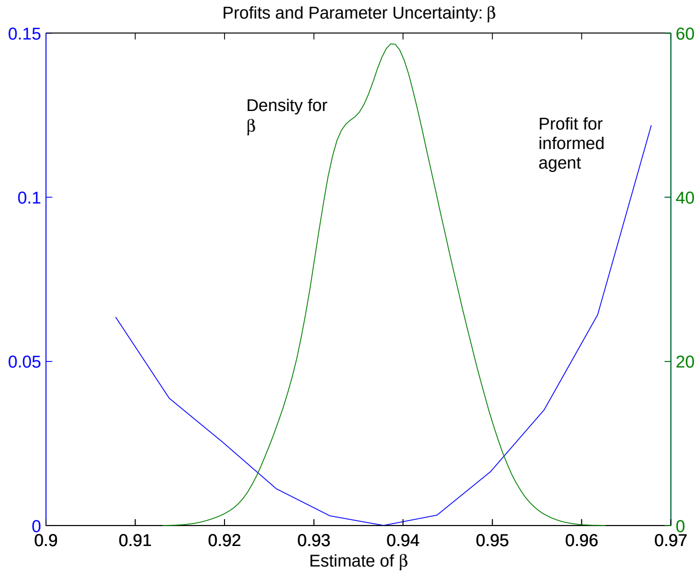

Figure 8: Profit to informed agent as a function of uninformed agent estimate of β. The left axis measures the profit, while the right axis measures density of the uninformed agent's estimates.

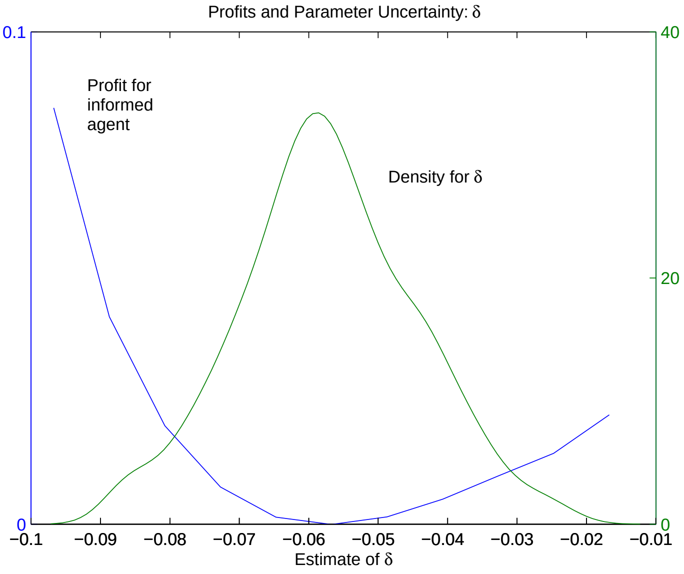

Figure 9: Profit to informed agent as a function of uninformed agent estimate of δ. The left axis measures the profit, while the right axis measures density of the uninformed agent's estimates.

4.2 Model Uncertainty

Model uncertainty results from the fact that we do not, in reality, know the true data generating process for asset returns, and as a result, must choose from a range of potentially incorrect models.

In order to obtain a range of reasonable models to compare model misspecification risk, we consider six competing models:

- A GJR model with the true parameters: $(\mu, \omega, \alpha, \beta, \delta)$.
- A T-GARCH model

$$
\begin{aligned}
r_t &= \mu + z_t h_t^{1/2}, z_t \sim t(\nu) \\
h_t &= \omega + \alpha \epsilon_{t-1}^2 + \beta h_{t-1}.
\end{aligned}
$$

  with estimated parameters: $(\hat{\mu}, \hat{\omega}, \hat{\alpha}, \hat{\beta}, \delta = 0, \hat{\nu})$.

- A GARCH model with estimated parameters: $(\hat{\mu}, \hat{\omega}, \hat{\alpha}, \hat{\beta}, \delta = 0)$.
- An EWMA model: $(\hat{\mu}, \omega = 0, \alpha = 0.03, \beta = 0.97, \delta = 0)$.
- A constant volatility model:$(\hat{\mu}, \hat{\omega}, \alpha = 0, \beta = 0, \delta = 0)$.

In order to derive an economic metric for model specification, we have an informed agent who uses the true GJR-GARCH model, as well as four other agents who use the misspecified models T-GARCH, GARCH, EWMA and constant volatility respectively, to derive their variance forecasts and price the butterfly option in each period.

To minimize parameter uncertainty and focus on model misspecification in this section, we estimated each of the misspecified models with 10 000 drawn from the GJR-GARCH DGP.

model in finite samples.

This gives us an accurate estimate of the probability limit of the parameters for the misspecified models under the true DGP. The parameters for each model are reported in Table 2 and are taken as given for the entire trading period.

We use the parameter estimates in Table 2, along with each agent's maintained model, to generate out-of-sample forecasts for the return distribution and price the butterfly option for each period. Each agent (model) trades against each of the others for a total of 100,000 trading periods from which we compute average profits and average holdings. The butterflies traded have wingspans covering from 99% of the current stock price to 101% of the current stock price.[6]

Table 3 presents average profits and holdings, as well as standard errors, associated with agents using models listed on the vertical axis when trading against a competing agent using models listed on the horizontal axis. For example, an agent using the true GJR-GARCH model makes very significant profits trading against the EWMA or the constant volatility models. However, while profits are positive trading against GARCH or T-GARCH models, they are not significantly different from zero. T-GARCH and GARCH users have quite close performance, with both making money from EWMA users and constant-volatility traders.

4.3 Both Model and Parameter Uncertainty

In practice, agents do not have the luxury of being able to separate model risk from parameter risk. In Tables 4 and 5, we present results from comparisons between the performance of agents using estimated models, as compared to an agent who has knowledge of the true model in Section 4.2. In Table 4 we generate 100 series of 5100 observations. For each data series, using the first 5000 observations, we estimate a GJR model, a GARCH model, a T-GARCH

[6]The choice of a somewhat narrower wingspan makes the numerical implementation of our equilibrium somewhat easier. For wider wingspans it is sometimes possible to observe an agent whose estimate of volatility is so low that they do not acknowledge the possibility of a zero payoff. A 1% daily shift, although large, lies well within the forecast range of most models, and so this problem does not occur.

Table 3: **Model Misspecification: trading results based on simulated GJR-GARCH data**

A. **Profits**

	Competing Models			
Model	T-GARCH	GARCH	EWMA	Const. vol.
True GJR	0.0026	0.0026	0.0126	0.1660
	(0.0022)	(0.0023)	(0.0037)	(0.0132)
T-GARCH		0.0004	0.0101	0.1650
		(0.0010)	(0.0029)	(0.0129)
GARCH			0.0100	0.1654
			(0.0029)	(0.0127)
EWMA				0.1527
				(0.0134)

B. **Holdings**

	Competing Models			
Model	T-GARCH	GARCH	EWMA	Const. vol.
True GJR	0.0103	0.0302	-0.0153	0.5191
	(0.0017)	(0.0017)	(0.0029)	(0.0082)
T-GARCH		0.0196	-0.0253	0.5096
		(0.0007)	(0.0023)	(0.0079)
GARCH			-0.0458	0.4902
			(0.0023)	(0.0078)
EWMA				0.5337
				(0.0085)

Panels A and B report average period profits and holdings of models listed on the vertical axis when trading against a competing model on the horizontal axis. Standard errors are reported in parenthesis.

model, an EWMA model and a constant volatility model, and use these parameter estimates to generate out-of-sample forecasts for volatility for each of the last 100 time periods.

Using these out-of-sample forecasts, we allow each model to trade against each of the others for each period. In all cases, the agents have risk-aversion parameters of two. Since there are 100 out-of-sample periods and 100 replications, there are 10,000 outcomes for each comparison from which we compute average profits or losses and average holdings. The butterflies traded by the agents are identical to those considered in Section 4.2.

When the sample size is 5000 (see Table 4), the fitted GJR model estimates the parameters almost exactly. An informed agent who knows the true parameters agrees almost exactly with the agent who has to estimate the parameters. As a result, very little trade takes place and profits are quite low. In contrast, both agents outperform agents with GARCH models, T-GARCH models, EWMA models and constant volatility models. T-GARCH and GARCH users have quite close performance, with both making money from EWMA users and Constant Volatility traders. EWMA users only make money from constant volatility users. Our ranking of the five approaches would be: true model, fitted GJR model, GARCH or T-GARCH, EWMA, and finally Constant Volatility.

When the estimation sample is shortened to 1000 observations (see Table 5) the results across models become very close. The fitted GJR model has almost indistinguishable performance when compared to the GARCH model or T-GARCH model. Even its performance when compared to the EWMA model is statistically insignificant. It would be impossible to reject the hypothesis that the fitted GJR model had no economic gains on average when trading with any of the misspecified models except the constant volatility model. Note that the T-GARCH does slightly worse than the GARCH model, as it does not use the correct (Gaussian) distribution for returns.

In summary, using an estimation sample of 1000 observations results in no benefit to

Table 4: **Trading results for volatility models based on simulated GJR-GARCH data. T=5000.**

A. **Profits**

	Competing Models				
Model	Fitted GJR	T-GARCH	GARCH	EWMA	Const. vol.
True GJR	0.0009	0.0044	0.0045	0.0079	0.1853
	(0.0012)	(0.0021)	(0.0021)	(0.0030)	(0.0135)
Fitted GJR		0.0028	0.0030	0.0070	0.1836
		(0.0020)	(0.0020)	(0.0032)	(0.0134)
T-GARCH			0.0011	0.0033	0.1810
			(0.0009)	(0.0027)	(0.0132)
GARCH				0.0033	0.1818
				(0.0027)	(0.0131)
EWMA					0.1690
					(0.0150)

B. **Holdings**

	Competing Models				
Model	Fitted GJR	T-GARCH	GARCH	EWMA	Const. vol.
True GJR	0.0075	0.0143	0.0160	-0.1145	0.4892
	(0.0010)	(0.0017)	(0.0017)	(0.0027)	(0.0087)
Fitted GJR		0.0067	0.0098	-0.1206	0.4832
		(0.0016)	(0.0016)	(0.0030)	(0.0086)
T-GARCH			0.0048	-0.1291	0.4769
			(0.0007)	(0.0025)	(0.0085)
GARCH				-0.1333	0.4731
				(0.0025)	(0.0084)
EWMA					0.6042
					(0.0104)

Panels A and B report average period profits and holdings of models listed on the vertical axis when trading against a competing model on the horizontal axis. Standard errors are reported in parenthesis. Each agent obtains model estimates from a rolling window of $T = 5000$ observations.

Table 5: **Trading results for volatility models based on simulated GJR-GARCH data. T=1000.**

A. **Profits**

Model	Fitted GJR	T-GARCH	GARCH	EWMA	Const. vol.
	Competing Models				
True GJR	0.0047	0.0075	0.0077	0.0107	0.1186
	(0.0027)	(0.0029)	(0.0029)	(0.0027)	(0.0121)
Fitted GJR		0.0020	0.0015	0.0063	0.1158
		(0.0025)	(0.0025)	(0.0038)	(0.0115)
T-GARCH			-0.0004	0.0038	0.1148
			(0.0010)	(0.0030)	(0.0112)
GARCH				0.0053	0.1174
				(0.0030)	(0.0112)
EWMA					0.1054
					(0.0132)

B. **Holdings**

Model	Fitted GJR	T-GARCH	GARCH	EWMA	Const. vol.
	Competing Models				
True GJR	0.0133	0.0221	0.0280	-0.0878	0.3831
	(0.0023)	(0.0023)	(0.0023)	(0.0021)	(0.0077)
Fitted GJR		0.0086	0.0130	-0.1003	0.3668
		(0.0021)	(0.0021)	(0.0031)	(0.0074)
T-GARCH			0.0058	-0.1089	0.3563
			(0.0007)	(0.0024)	(0.0070)
GARCH				-0.1147	0.3519
				(0.0024)	(0.0070)
EWMA					0.4734
					(0.0086)

Panels A and B report average period profits and holdings of models listed on the vertical axis when trading against a competing model on the horizontal axis. Standard errors are reported in parenthesis. Each agent obtains model estimates from a rolling window of $T = 1000$ observations.

using the fitted GJR model as compared to using a EWMA model for this application. Since this was not the case when a longer sample was used to estimate the GJR parameters, the greater variability in accuracy of the volatility forecasts due to the higher degree of parameter uncertainty almost eliminates the disadvantage of using the incorrect model. We also note that, conversely, inclusion of an erroneous parameter, as in the T-GARCH model, has little effect for large samples, but can lead to inferior performance for small samples.

It may be useful to consider one of the model comparisons in greater detail. Consider the case of an agent with a GJR-GARCH model competing with an agent who has access to an EWMA model. We simulate 5100 observations of GJR-GARCH data, generated according to (10)-(11). Both agents fit their models to the first 5000 observations. They then use their models to compute out-of-sample forecasts of volatility for each of the last 100 observations. Using these volatility beliefs, each computes a demand curve, and the market is cleared, resulting in a price for the butterfly, and an equilibrium quantity of butterflies for each agent. Given these outcomes, it is possible to compute the payoff for the agents. We then repeat this experiment 100 times to generate 10,000 equilibria and payoffs.

Figure 10 shows the distribution of holdings for the GJR-GARCH agent. This agent invests up to 67 cents in long positions, and takes short positions of at most \$2.09. The mean holding for this agent is a *short* position of value 12 cents, with a standard error of 0.3 cents. Hence the standard deviation of holdings is $0.003 \times \sqrt{10000}$ or 29 cents. These holdings result in the GJR-GARCH agent registering profits as high as \$2.09 and losses as high as \$2.54. Overall, the GJR agent does profit, however, making \$0.007 profit on average per day on an average (short) position of \$.12. Since the market is a zero sum game, the EWMA agent will have an average *long* position of \$0.12, and an average *loss* of \$0.007. These positions are scalable. To give an idea of the average return in percentage terms, note that the uninformed investor's average loss is 3.54% of the average (absolute) position per day.

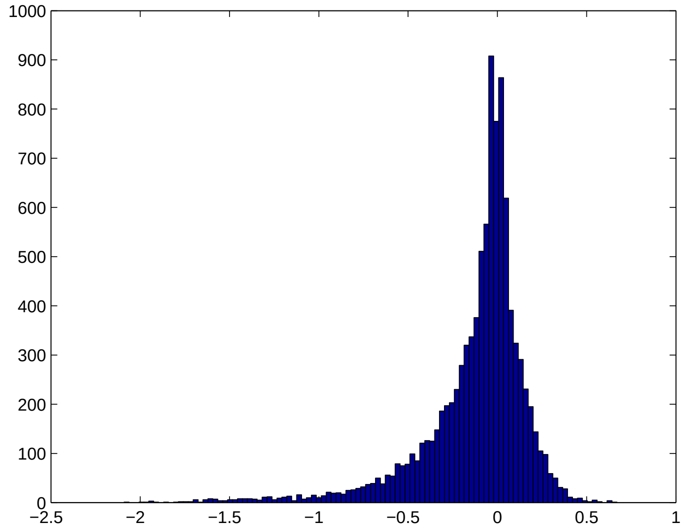

Figure 10: Holdings (in dollars) of agent with GJR model, when trading against an agent using a EWMA model. Data was generated using the GJR model (10)-(11). Agents were allowed 5000 observations to calibrate their models.

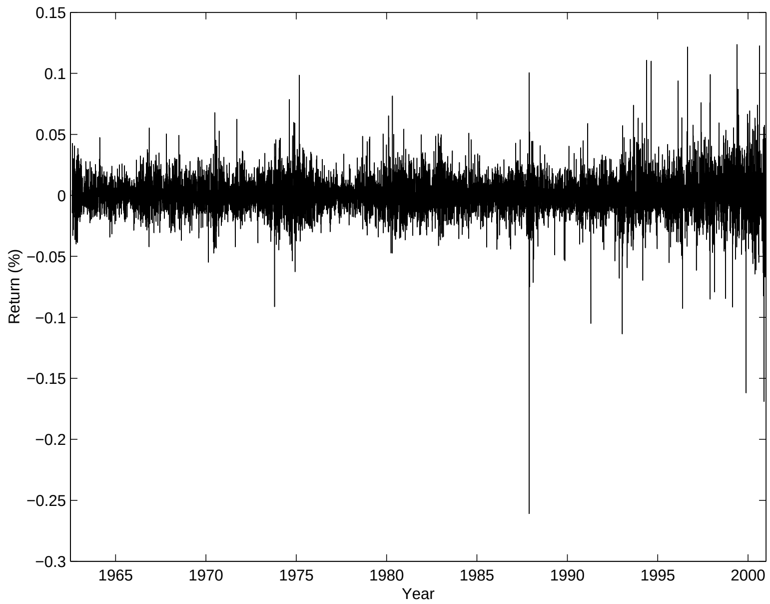

Figure 11: Returns on IBM stock, July 1962 to December 2000.

5 Empirical Example

We now perform our model comparisons using historical data for IBM returns. In this case, the true data generating process is not known. Therefore, we calculate gains and losses to butterfly trades using realized historical prices. Figure 11 plots IBM daily returns from July 1962 to the end of December 2000. As with many financial time series, IBM's returns have exhibited substantial volatility clustering: large positive and negative returns tend to occur in succession. Although we know that for equity returns large negative returns are typically followed by an increase in conditional variance, given the tradeoff between model risk and parameter uncertainty associated with small sample sizes, it is not immediately clear that the GJR GARCH model or T-GARCH model should outperform the GARCH model (or indeed, the EWMA model).

We allow each model a rolling window of 1,000 (alternatively 5,000) observations to estimate parameters, and then use these parameters to forecast volatility for the next day. Each model is estimated for each new time period. We perform pairwise comparisons of the models' performance according to our financial metric.

The results are reported in tables 6 and 7. In both cases, a ranking emerges - T-GARCH, GJR-GARCH, GARCH, EWMA and finally constant volatility. Each of the GARCH specifications makes significant profits against the constant volatility and the EWMA models. However, there are no significant differences among the profits when a GARCH model trades against another GARCH model. This is in direct contrast to statistical measures which generally indicate the importance of fat-tails and leverage. Finally, there is the widening of the spread between T-GARCH and GJR-GARCH and between GJR-GARCH and GARCH as the in-sample estimation period increases from 1000 to 5000.

Table 6: **Trading results for volatility models based on historical IBM data. T=1000.**

A. **Profits**

Model	Competing Models T-GARCH	GARCH	EWMA	Const. vol.
GJR-GARCH	-0.0013	0.0026	0.0102	0.0413
	(0.0058)	(0.0035)	(0.0058)	(0.0071)
T-GARCH		0.0081	0.0125	0.0414
		(0.0055)	(0.0064)	(0.0091)
GARCH			0.0083	0.0379
			(0.0050)	(0.0067)
EWMA				0.0288
				(0.0096)

B. **Holdings**

Model	Competing Models T-GARCH	GARCH	EWMA	Const. vol.
GJR-GARCH	-0.2890	0.0278	-0.0659	0.2018
	(0.0032)	(0.0023)	(0.0046)	(0.0051)
T-GARCH		0.3117	0.2302	0.5203
		(0.0029)	(0.0041)	(0.0055)
GARCH			-0.0885	0.1731
			(0.0042)	(0.0050)
EWMA				0.2877
				(0.0076)

Panels A and B report average period profits and holdings of models listed on the vertical axis when trading against a competing model on the horizontal axis. Standard errors are reported in parenthesis. Each agent obtains model estimates from a rolling window of $T = 1000$ observations.

Table 7: **Trading results for volatility models based on historical IBM data. T=5000.**

A. **Profits**

Model	Competing Models T-GARCH	 GARCH	 EWMA	 Const. vol.
GJR-GARCH	-0.0026	0.0042	0.0108	0.0272
	(0.0048)	(0.0025)	(0.0034)	(0.0080)
T-GARCH		0.0067	0.0140	0.0262
		(0.0047)	(0.0047)	(0.0084)
GARCH			0.0076	0.0271
			(0.0022)	(0.0080)
EWMA				0.0182
				(0.0099)

B. **Holdings**

Model	Competing Models T-GARCH	 GARCH	 EWMA	 Const. vol.
GJR-GARCH	-0.2642	0.0251	-0.0310	0.0162
	(0.0026)	(0.0017)	(0.0029)	(0.0061)
T-GARCH		0.2784	0.2325	0.3248
		(0.0021)	(0.0022)	(0.0068)
GARCH			-0.0498	-0.0076
			(0.0021)	(0.0060)
EWMA				0.0833
				(0.0081)

Panels A and B report average period profits and holdings of models listed on the vertical axis when trading against a competing model on the horizontal axis. Standard errors are reported in parenthesis. Each agent obtains model estimates from a rolling window of $T = 5000$ observations.

6 Conclusions

This paper introduces a new metric for comparing the performance of volatility forecasting models. By comparing the ability of models to make or lose money when trading butterfly positions, we hope to evaluate whether differences in forecasting accuracy are *economically* significant. Our methodology offers a fast, easy to implement, means of ranking alternative forecasting models and evaluating the economic implications of parameter uncertainty and model misspecification.

The applications presented in this paper demonstrate the use of our approach for comparing economic losses due to parameter uncertainty and model misspecification. In trading based on IBM data, we find significant gains to going beyond the basic exponentially-weighted moving-average model of volatility and using GARCH parameterizations. However, additional features such as fat-tails and leverage appear to have a smaller impact on profits in our applications, unless one has sufficient data to minimize parameter uncertainty.

References

Avramov, D. (2002), 'Stock return predictability and model uncertainty', *Journal of Financial Economics* **64**, 423–458.

Barberis, N. (2000), 'Investing for the long run when returns are predictable', *Journal of Finance* **55**(1), 225–264.

Engle, R. F., Hong, C., Kane, A. and Noh, J. (1993), *Advances in Futures and Options Research* , Vol. 6, JAI Press, chapter Arbitrage Valuation of Variance Forecasts with Simulated Options, pp. 393–415.

Fleming, J., Kirby, C. and Ostdiek, B. (2001), 'The economic value of volatility timing', *Journal of Finance* **56**, 329–352.

Garcia, R., Luger, R. and Renault, E. (2001), Asymmetric smiles,leverage effects and structural parameters, CIRANO Working Paper 2001s-01, University of Montreal.

Garcia, R., Luger, R. and Renault, E. (2003), 'Empirical assessment of an intertemporal option pricing model with latent variables', *Journal of Econometrics* pp. 49–83.

Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993), 'On the relation between the expected value and the volatility of the nominal excess return on stocks', *Journal of Finance* **48**(5), 1779–1801.

Jackwerth, J. (2000), 'Recovering risk aversion from option prices and realized returns', *Review of Financial Studies* **13**(2), 433–451.

Noh, J., Engle, R. F. and Kane, A. (1994), 'Forecasting Volatility and Option Prices of the S&P Index', *Journal of Derivatives* pp. 17–30.

Rosenberg, J. E. and Engle, R. F. (2002), 'Empirical pricing kernels', *Journal of Financial Economics* **64**, 341–372.

West, K., Edison, H. and Cho, D. (1993), 'A utility-based comparision of some models of exchange rate volatility', *Journal of International Economics* **35**, 23–45.