Implied and Realized Volatility: Empirical Model Selection

Lan Zhang*
University of Illinois at Chicago

May 1, 2009

Abstract

The paper studies the nonparametric connection between realized and implied volatilities. No-arbitrage identities and comparison inequalities are found. We formulate the multi-factor trading system on the volatility scale. To empirically determine the number of factors, we develop a high frequency analysis for sequential F-testing. We also design a cross validated estimate of quadratic variation.

Keywords: cross validation, discrete observation, F-testing, implied volatility, Itô process, leverage effect, model selection, realized volatility,

JEL Codes: C02; C13; C14; C22; D52; D81; G11; G13

*Lan Zhang, Department of Finance, The University Illinois at Chicago, Chicago, IL 60607, U.S.A.; lanzhang@uic.edu; http://tigger.uic.edu/~lanzhang/.

1 Introduction

It is commonly known that volatility can either be estimated through the *past* variation of the price process of an underlying asset, or be imputed through a derivative pricing model. The former approach delivers the *realized* or *historical* volatility,[1] and the latter gives the *implied* volatility.[2] In the Black and Scholes (1973)-Merton (1973) model, both volatilities represent the same object, the variations of the returns in the underlying security. However, it is rare that realized and implied volatility agree in practice.

In this paper we study the connection between realized volatility and implied volatility. We first establish (Section 2) a conceptual relationship between implied and realized volatility in a nonparametric framework. We here argue that it is natural to think of implied volatility on the *cumulative* scale, and that there is a tradable part of the implied volatility. One- and multi-factor structures for implied volatility are discussed in Section 3, where we discuss trading schemes. Since our framework is nonparametric, and since the data are only observed at discrete times, there is a need for statistical model selection in the multi-factor structure. We therefore end the paper by proposing two procedures for such model selection: sequential F-testing, and cross validation (Section 4).

Conceptually, the current paper is close to Renault and Touzi (1996) in that implied volatilities are used to facilitate hedging. The difference between the two papers is that Renault and Touzi (1996) uses a probability model to determine the hedging strategy, while the current paper uses non-parametric inference.

Implied volatility has been studied from several other angles. First of all, there is a rich econometric literature investigating which (realized or implied) volatility provides a better forecast for the subsequent variation of the returns. Substantial empirical work suggests that implied volatilities contain information about future variability in the stock in a way that the past realized data cannot capture.[3] Second, there is also an important mathematical finance literature on the concept of local implied volatility[4], and on the asymptotic behavior of implied volatility near expiration or at extreme strikes.[5] Third, there is a large literature on variance swaps.[6] Finally, inversion of the

[1]The estimation of volatility from high frequency data has yielded a huge literature in recent year, see, for example, Jacod (1994), Foster and Nelson (1996), Comte and Renault (1998), Jacod and Protter (1998), Andersen, Bollerslev, Diebold, and Labys (2001, 2003), Barndorff-Nielsen and Shephard (2001, 2002), Dacorogna, Gençay, Müller, Olsen, and Pictet (2001), Zhang (2001, 2006), Zhang, Mykland, and Aït-Sahalia (2005), and Mykland and Zhang (2006, 2007).

[2]Beckers (1981), Engle and Mustafa (1992), Bick (1995), Pena, Rubio, and Serna (1999), and Rubinstein (1994)).

[3]See Latane and Rendleman (1976); Schmalensee and Trippi (1978); Day and Lewis (1992); Scott (1992); Canina and Figlewski (1993); Lamoureux and Lastrapes (1993); Sheikh (1993); Derman and Kani (1994); Jorion (1995); Amin and Ng (1997), Christensen and Prabhala (1998), Blair and Taylor (2001)), Kang, Zhang, and Chen (2009), and many others.

[4]Derman and Kani (1998), Lee (2001)

[5]Berestycki, Busca, and Florent (2000, 2002, 2004), Lee (2004). A nice survey is given in Lee (2005).

[6]Carr, Geman, Madan, and Yor (2005), Carr and Lee (2007).

Black-Scholes-Merton formula can be used in connection with bounds on options prices.[7]

2 Implied and Realized Volatility

2.1 Setup and Notation

We consider a non-dividend paying stock $\{S_t\}$ and a zero coupon bond $\{\Lambda_t\}$, on a time interval $[0,T]$, where $\Lambda_T = 1$. The discounted stock price is given by $\tilde{S}_t = S_t/\Lambda_t$, and, in general, for other instruments with price V_t, $\tilde{V}_t = V_t/\Lambda_t$. In the special case of a non-random short rate r_t, the price of the zero coupon bond at time t can be expressed as $\Lambda_t = \exp\{-\int_t^T r_u du\}$. For simplicity, we shall develop theory for discounted prices,

We assume that under the actual physical distribution P, the discounted stock price follows the equation

$$d\tilde{S}_t = \mu_t \tilde{S}_t dt + \sigma_t \tilde{S}_t dW_t \tag{1}$$

where the drift term μ_t and the diffusion term σ_t can be stochastic and time-varying, and W follows a standard Brownian Motion. Under a a risk neutral distribution, (1) becomes $d\tilde{S}_t = \sigma_t \tilde{S}_t dW_t^*$, where W^* is the standard Brownian Motion under that risk neutral measure. It will be assumed that all quantities in (1) are adapted to an underlying filtration $(\mathcal{F}_t)$, which is not necessarily generated by S or W. However, we do assume that W is an $(\mathcal{F}_t)$-Wiener process. We also suppose that S is an $(\mathcal{F}_t)$-*Itô process*, which, in addition to the above, requires $|\mu_u|$ and σ_u^2 to be integrable (a.s.) on the interval $[0,T]$. Finally, we shall suppose that $\sigma_t^2 > 0$ for all t.

Consider a European option with payoff $f(S_T)$ at expiration T, its price from the standard Black and Scholes (1973)-Merton (1973) formula (abbreviated with BS) at time t can be written as $C(S_t, -\log(\Lambda_t), \sigma^2(T-t))$, where

$$C(S,R,\Xi) = \exp(-R)Ef(S\exp(R - \Xi/2 + \sqrt{\Xi}Z)) \tag{2}$$

The option maturity T and payoff form f are given by the option contract, and Z is standard normal (see, for example, Chapter 6 of Duffie (1996)). The most common instance would be the call option, where $f(s) = (s-K)^+$ with pre-determined strike price K. In general, for simplicity, we assume the following about the payoff function f:

Condition $C1$. The function f is assumed to satisfy: (i) $f : (0,+\infty) \to \mathbb{R}$, (ii) f is convex, and f is not an affine function on $(0,+\infty)$; and (iii) $E|f(\exp(U))| < \infty$ for any normal random variable U. □

From Condition $C1$, it follows that C in (2) is well defined and infinitely many times differen-

[7] Mykland (2000, 2003a,b).

tiable, with $C_\Xi > 0$ [8], for $\Xi > 0$, $S > 0$, and any R. Also, set $C(S, R, \infty) = \lim_{\Xi \to \infty} C(S, R, \Xi)$. This latter quantity is typically infinity, but not always.

Suppose that the actual market price of this option is given by V_t, and also suppose that $\tilde{V} = V/\Lambda$ is an $(\mathcal{F}_t)$-Ito process. We emphasize that the Brownian motion driving V need not be the same as the one driving S, although one expects that in general the two will be connected. The Itô process assumption is natural for both S and V, since their discounted values are martingales under the risk neutral measure, and since the original measure P and the risk neutral measure are mutually absolutely continuous under no-arbitrage assumptions (see Delbaen and Schachermayer (1995a,b, 1998) for details about the no arbitrage argument).

Note that from standard no-arbitrage considerations,

$$\forall t \in [0, T), \quad C(S, -\log(\Lambda_t), \infty) > V_t > C(S_t, -\log(\Lambda_t), 0) \quad \text{a.s.} \tag{3}$$

Thus, the following is well defined.

Definition 1. Under Condition C1, the *cumulative implied volatility (CIV) at time* t is defined to be the unique solution Ξ_t of

$$V_t = C(S_t, -\log(\Lambda_t), \Xi_t). \tag{4}$$

□

In the following sections, we discuss the relationship between implied and realized volatility when Ξ demonstrates different level of smoothness. The realized volatility is either on its instantaneous (σ_t^2) or its integrated quadratic variation $(\langle \log \tilde{S}, \log \tilde{S} \rangle_t = \int_0^t \sigma_u^2 du)$ form.

Remark 1. Dividing both sides of $V_t = C(S_t, -\log(\Lambda_t), \Xi_t)$ by Λ_t yields,

$$\begin{aligned} \tilde{V}_t &= \frac{1}{\Lambda_t} C(S_t, -\log \Lambda_t, \Xi_t) \\ &= C(\tilde{S}_t, 0, \Xi_t) \end{aligned} \tag{5}$$

In equation (5), using zero-coupon bond as numeraire reduces the dimensions of model price C of the option, where C becomes only a function of the futures price $\tilde{S}_t$ and of the CIV Ξ_t. In particular, Ξ_t is also an Ito process by the Implicit Function Theorem, in view of our assumptions on f (and hence on C) and because $\tilde{S}$ and $\tilde{V}$ are Itô processes. □

2.2 Cumulative vs. Instantaneous Implied Volatility, and The Zero Factor Model

Our notion of implied volatility is on the *cumulative* scale, from t to expiration T. This is in contrast to much of the literature (cited above), which considers implied instantaneous volatility. We argue

[8] C_Ξ is defined as $\partial C / \partial \Xi$.

that *instantaneous implied volatility almost never exists, and even if it does, it must equal to the instantaneous historical volatility* σ_t^2. We explain in the following.

THEOREM 1. *Assume Condition C1. Suppose* $\tilde{S}$ *and* $\tilde{V}$ *are Ito processes, with* $\tilde{S}$ *satisfying (1). Let the cumulative implied volatility* Ξ *be given by (4). Assume that there is (at least one) measure* P^**, equivalent to* P*, under which* $\tilde{S}_t$ *and* $\tilde{V}_t$ *are martingales. (In particular, there is no arbitrage). If* Ξ_t *is absolutely continuous as a function of* t *on the interval* (t_1, t_2)*,* $\xi_t = -\frac{d}{dt}\Xi_t$*, then for the same interval, under both the physical probability measure* P *and the risk neutral measure* P^**,*

$$\begin{aligned} \sigma_t^2 &= \xi_t, \text{ and} && (6)\\ d\tilde{V}_t &= \Delta_t d\tilde{S}_t, \text{where} && (7)\\ \Delta_t &= C_{\tilde{S}}(\tilde{S}_t, 0, \Xi_t) && (8) \end{aligned}$$

The good news is that in this situation, the implied volatility leads to an exact "delta hedge" (7)-(8).

However, the theorem's prediction as in (6) occurs rarely in reality, in other words, "instantaneous implied volatility" $(-\frac{d}{dt}\Xi_t)$ will typically not exist. Intuitively, the reason is as follows: If Ξ_t is absolutely continuous on the whole interval until expiration $(0, T)$, then subject to regularity conditions, (6) leads to $\int_0^T \sigma_t^2 dt = -\Xi_T + \Xi_0$. Since $\Xi_T = 0$, this implies that $\int_0^T \sigma_t^2 dt$ will be known and equal to Ξ_0 at time zero. Such a happy state of affairs is not typical.

The model described above will be called the *zero factor model*, since there is no impact of any factor structure in the volatility on the hedging of the derivative security. This will be consistent with our multi-factor structure below.

REMARK 2. (Empirical testing of the zero factor model). To check empirically whether Ξ_t is absolutely continuous, one can proceed as follows. From Remark 1, Ξ_t is an Itô process. Hence, Ξ_t is absolutely continuous if and only if its quadratic variation $\langle \Xi, \Xi \rangle$ is zero. This is testable on the basis of high frequency financial data. A general discussion of methods is given in Section 4 below, for the case where there is no microstructure noise. If noise is present, one can estimate $\langle \Xi, \Xi \rangle$ (over a desired interval) with an estimator such as the two- or multi-scale realized volatility (Zhang, Mykland, and Aït-Sahalia (2005), Zhang (2006)). One can then use the asymptotic mixed normality of the estimator to test whether $\langle \Xi, \Xi \rangle$ is zero. □

REMARK 3. Most literature on implied volatility define the instantaneous volatility at time t to be $\Xi_t/(T-t)$, which by itself is well defined, however, this seems to somewhat defy the notion of "instantaneous". □

Proof of Theorem 1: This result is a corollary to the more general Theorem 3 below. No arbitrage implies (3). As Ξ_t is absolutely continuous, $\Xi_t^{MG} = 0$, $\Xi_t = \Xi_t^{DR}$. Thus, the last two terms in equation (11) both become zero. Meanwhile, (14) and (12) reduce to equations (6) and (7). □

2.3 A Comparison Result for Implied Volatilities

We here provide a further connection between realized and implied volatilities. Let $\Xi^- < \Xi^+$, and consider the set

$$\Xi^- \;\leq\; \int_0^T \sigma_t^2 dt \;\leq\; \Xi^+ \tag{9}$$

We have the following result:

THEOREM 2. *Assume Condition C1. Suppose $\tilde{S}$ and $\tilde{V}$ are Ito processes, with $\tilde{S}$ satisfying (1). Let the cumulative implied volatility Ξ be given by (4). Assume that there is (at least one) measure P^*, equivalent to P, under which $\tilde{S}_t$ and $\tilde{V}_t$ are martingales. (In particular, there is no arbitrage). Finally, asssume that (9) has probability one. Then the set*

$$\forall t \in [0,T] : \Xi^- - \int_0^t \sigma_u^2 du \;\leq\; \Xi_t \;\leq \Xi^+ - \int_0^t \sigma_u^2 du \tag{10}$$

also has probability one.

Proof of Theeorem 2. Let $\epsilon > 0$. Suppose $\Xi_t + \int_0^t \sigma_u^2 du$ hits $\Xi^+ + \epsilon$ at time τ ($\in (0,T)$)). At this time τ, sell one unit of V_τ, and start a trading strategy at time τ with initial (discounted) value $C(\tilde{S}_\tau, 0, \Xi^+ - \int_0^\tau \sigma_u^2 du)$ and hedge ratio (delta) in the stock of $C_{\tilde{S}}(\tilde{S}_t, 0, \Xi^+ - \int_0^t \sigma_u^2 du)$ at times t in $[\tau, T]$. From p. 667 of Mykland (2000), this strategy produces a payoff at time T equal to $C(\tilde{S}_T, 0, \Xi^+ - \int_0^T \sigma_u^2 du) \geq V_T$. Hence, the overall strategy makes a positive profit at time τ, and cannot lose money at time T. This shows the upper bound. The lower bound follows similarly. □

2.4 General No Arbitrage Identities

From Section 2.2, it is clear that one has to expect the implied volatility to be non-differentiable with respect to t, and hence have non-zero quadratic variation. We here explore this more general case.

Using Ito's Lemma on (5), one gets

$$\begin{aligned} d\tilde{V}_t &= C_{\tilde{S}} d\tilde{S}_t + C_\Xi d\Xi_t \\ &\quad + \frac{1}{2} C_{\tilde{S}\tilde{S}} d\langle \tilde{S}, \tilde{S} \rangle_t \\ &\quad + \frac{1}{2} C_{\Xi\Xi} d\langle \Xi, \Xi \rangle_t + C_{\tilde{S}\Xi} d\langle \tilde{S}, \Xi \rangle_t \end{aligned} \tag{11}$$

Now define a smooth process through

$$\begin{aligned} d\Xi_t^{DR} &= -\frac{1}{C_\Xi} \left\{ \frac{1}{2} C_{\tilde{S}\tilde{S}} d\langle \tilde{S}, \tilde{S} \rangle_t + \frac{1}{2} C_{\Xi\Xi} d\langle \Xi, \Xi \rangle_t + C_{\tilde{S}\Xi} d\langle \tilde{S}, \Xi \rangle_t \right\} \\ &= -\sigma_t^2 dt - \frac{1}{C_\Xi} \left\{ \frac{1}{2} C_{\Xi\Xi} d\langle \Xi, \Xi \rangle_t + C_{\tilde{S}\Xi} d\langle \tilde{S}, \Xi \rangle_t \right\} \end{aligned} \tag{12}$$

since $C_\Xi = C_{\tilde{S}\tilde{S}}\tilde{S}^2/2$. Also, let the remainder Ξ_t^{MG} be given by

$$d\Xi_t = d\Xi_t^{DR} + d\Xi_t^{MG}. \tag{13}$$

Equation (11) therefore becomes

$$d\tilde{V}_t = C_{\tilde{S}} d\tilde{S}_t + C_\Xi d\Xi_t^{MG}, \tag{14}$$

and, in particular, Ξ^{MG} is a martingale under any equivalent martingale measure. As importantly for our purposes, Ξ^{MG} can be seen as the (discounted) value of a traded security.

To summarize the above, we state the following:

THEOREM 3. *Assume Condition C1. Suppose $\tilde{S}$ and $\tilde{V}$ are Ito processes, with $\tilde{S}$ satisfying (1). Let the cumulative implied volatility Ξ be given by (4). Assume that there is (at least one) measure P^*, equivalent to P, under which $\tilde{S}_t$ and $\tilde{V}_t$ are martingales. Then, under P and any equivalent measure, the equations (12) and (14) will hold.*

EXAMPLE 1. (Logarithmic Payoffs). Take $f(s) = \log(s)$. This is the standard situation of the variance swap (Carr, Geman, Madan, and Yor (2005), Carr and Lee (2007)). Here, $C(\tilde{S}, 0, \Xi) = \log(\tilde{S}) - \frac{\Xi}{2}$, and so $C_{\tilde{S}} = \frac{1}{\tilde{S}}$, $C_{\tilde{S}\tilde{S}} = -\frac{1}{\tilde{S}^2}$, and $C_\Xi = -\frac{1}{2}$. Hence,

$$d\Xi_t^{DR} = -\frac{1}{C_\Xi}[\frac{1}{2}C_{\tilde{S}\tilde{S}} d\langle\tilde{S},\tilde{S}\rangle_t] = -\sigma_t^2 dt \tag{15}$$

and

$$d\Xi^{MG} = -2(d\tilde{V}_t - C_{\tilde{S}} d\tilde{S}_t). \tag{16}$$

This is an unusual situation in that (15) holds despite the possibility that Ξ^{MG} is nonzero. This is another way of describing why the variance swap can be hedged in S and the log option. □

3 Hedging Implied Volatility

3.1 Hedging in the Underlying Process

3.1.1 General Considerations; "Gamma" and "Vega" Risk; Leverage Effect

We saw in Section 2.2 that if the implied volatility is absolutely continuous, the above result provides a perfect hedge. In the more general case, suppose that

$$d\Xi_t = \rho_t d\tilde{S}_t + dZ_t, \tag{17}$$

where $\rho_t = d\langle\Xi,\tilde{S}\rangle_t / d\langle\tilde{S},\tilde{S}\rangle_t$, for $t \in (0,T)$. One can view equation (17) as a local regression of Ξ on $\tilde{S}$, with ρ being the regression coefficient and Z being the residual. Z may or may not be absolutely continuous. In this case, a minimal martingale hedge (Föllmer and Sondermann (1986), Föllmer and Schweizer (1991), Schewizer (1990, 1991)) gives the evolution of the price of the derivative in the following form:

$$d\tilde{V}_t = \Delta_t d\tilde{S}_t + C_\Xi dZ_t^{MG}, \text{ where} \tag{18}$$

$$\Delta_t = C_{\tilde{S}} + \rho_t C_\Xi \tag{19}$$

where $dZ_t^{MG} = d\Xi_t^{MG} - \rho_t d\tilde{S}_t$. Hence, (19) provides a correction term to the "implied" delta hedge which improves trading on the basis of implied volatility only. In analogy with common parlance, we can think of this as a minimization of exposure to "Gamma" and "Vega" risk (see Hull (2008)).

Similarly,

$$d\Xi_t^{DR} = -\sigma_t^2 \left(1 + \frac{C_{\Xi\Xi}}{C_{\tilde{S}\tilde{S}}} \rho_t^2 + 2\frac{C_{\tilde{S}\Xi}}{C_{\tilde{S}\tilde{S}}} \rho_t\right) dt - \frac{1}{2} \frac{C_{\Xi\Xi}}{C_\Xi} d\langle Z, Z\rangle_t \tag{20}$$

again since $C_\Xi = C_{\tilde{S}\tilde{S}} \tilde{S}^2/2$.

REMARK 4. (Leverage effect.) The coefficient ρ_t is a form of *leverage effect* on the implied volatility, and is empirically expected to have a negative sign. Since C_Ξ is positive for convex payoffs, we obtain that the delta hedge in (19) is smaller than the Black-Scholes implied delta $C_{\tilde{S}}$. The latter will thus tend to overhedge the position. This is similar to the findings of Renault and Touzi (1996) in their setting. □

REMARK 5. Obviously, the hedge in (19) can be obtained by directly using the prices of market traded (European) options $\{V_t\}$, as $C_{\tilde{S}} + \rho_t C_\Xi = d\langle \tilde{V}, \tilde{S}\rangle_t / d\langle \tilde{S}, \tilde{S}\rangle_t$. The advantage of proceeding via the implied volatility Ξ rather than the raw price V is that implied volatilities move on the same scale across strike prices, thus share similar statistical properties whereas raw prices of at-the-money options behaves quite differently from those out-of-the-money and in-the-money ones. This consideration motivates the use of statistical techniques in Section 4 below. □

If one hedges according to this scheme, it is of substantial interest to control the size of the residual $C_\Xi dZ_t^{MG}$. This can be done empirically by using the Analysis of Variance/Variation (ANOVA) techniques developed in Zhang (2001) and Mykland and Zhang (2006). Below, in Section 4.2, we develop a safer estimator of $\langle Z, Z\rangle$, which is based on cross validation.

3.1.2 One Factor Structure on Ξ

Suppose the *CIV* Ξ has a one-factor structure, by which we mean that $\tilde{S}_t$ exhausts the martingale variation of Ξ_t^{MG}. The following result then gives the form of the implied volatility, and asserts that we have a perfect hedge.

Corollary 1. *Assume the conditions of Theorem 3. Also suppose that the cumulative implied volatility Ξ_t process satisfies equation (17) as well as*

$$dZ_t = -\zeta_t dt \tag{21}$$

then

$$\zeta_t = \sigma_t^2 + \frac{C_{\Xi\Xi}}{C_{\tilde{S}\tilde{S}}}\rho_t^2\sigma_t^2 + 2\frac{C_{\Xi\tilde{S}}}{C_{\tilde{S}\tilde{S}}}\rho_t\sigma_t^2, \text{ and} \tag{22}$$

$$d\tilde{V}_t = \Delta_t d\tilde{S}_t, \text{ where} \tag{23}$$

Δ_t is given by (19), under both the physical probability measure P and any equivalent measure.

The result follows directly from (18)-(20) since $dZ^{MG} \equiv 0$.

3.2 Multifactor Structure on Ξ: Hedging as Regression

In general, one cannot expect to be in the one factor situation as in (21). One can then seek to complete the market by adding market traded instruments until one has to a great extent completed the market. Suppose that we have p different options, with prices $V^{(1)}, ..., V^{(p)}$, written on the underlying stock. Each option $V^{(i)}$ has an implied volatility process $\Xi^{(i)}$. In extension of model (17), one can then hope to hedge the first option in subset of the remaining options, say

$$d\Xi^{(1)MG} = \rho_t d\tilde{S}_t + \sum_{k=2}^{q} \rho_t^{(k)} d\Xi^{(k)MG} + dZ_t^{MG} \tag{24}$$

where the aim is for $\langle Z, Z\rangle$ to be as small as possible. In view of the relationship (14), this is the same as constructing a hedge for $\tilde{V}_t^{(1)}$ in S and $\tilde{V}_t^{(2)}, ..., \tilde{V}_t^{(q)}$, with error $C_\Xi dZ_t$. Since we shall now proceed to treating the problem statistically, it is preferable to have the relevant quantities as much as possible on the same scale. See Remark 5 for elaboration.

The question now arises how to select the subset of options to be used in making the market more complete. As a probabilistic problem, with a well defined model and continuously observed process paths, this is straightforward. In particular, if the underlying filtration $(\mathcal{F}_t)$ is generated (under some equivalent martingale measure) by r independent Brownian motions, then $q = r - 1$. For example, in a Heston (1993) model, $q = 1$.

However, when a model or continuous observation is absent, we need to solve the problem with statistical methods. This motivates the next section.

4 Empirical Model Selection for Options Hedging

We are dealing with a system on the form

$$dY_t = \sum_{k=1}^{p} \beta_t^{(k)} dX_t^{(k)} + dZ_t, \text{ with } \langle X^{(k)}, Z\rangle_t = 0 \text{ for all } t \text{ and } k. \tag{25}$$

where Itô processes $X_t^{(1)}, ..., X_t^{(p)}$ and Y_t are observed (and traded) at discrete times.

In terms of the development above, we take $Y_t = \Xi^{(1)MG}$, $X_t^{(1)} = \tilde{S}_t$, and $X_t^{(k)} = \Xi^{(k)MG}$. We write X and Y to emphasize the general nature of the formulation.

Remark 6. One can equally well take $Y_t = \Xi^{(1)}$, $X_t^{(1)} = \tilde{S}_t$, and $X_t^{(k)} = \Xi^{(k)}$, in view of the asymptotic negligibility of "*dt*" type terms (Section 2.2 of Mykland and Zhang (2007)). Apart from the loss of scale comparability, the analysis can also be done on the original system, with $Y_t = \tilde{V}_t^{(1)}$, $X_t^{(1)} = \tilde{S}_t$, and $X_t^{(k)} = \tilde{V}_t^{(k)}$ for $k \geq 2$. Obviously, the formulation also extends to several underlying securities S, etc. □

The regression coefficients $\beta_t^{(k)}$ can be estimated at any given time through a rolling regression on the preceding data (Foster and Nelson (1996), Comte and Renault (1998), Hayashi and Mykland (2005), Mykland and Zhang (2008)). Hence, an approximate statistical trading strategy is well defined. The development in Zhang (2001) and Mykland and Zhang (2006) was concerned with measuring $\langle Z, Z\rangle$ in the model of form (25).

We now consider model selection for this system. Suppose that one observes this system for, say, a day, and then want to apply the conclusions to trading for the following day, assuming reasonable stability of the system.

For ease of notation, we shall think of the day under observation as being a time span from 0 to $\mathcal{T}$. Observations are made at times $t_{n,j} = j\mathcal{T}/n$, $j = 0, ..., n$. Take $\Delta Y_{t_{n,j+1}} = Y_{t_{n,j+1}} - Y_{t_{n,j}}$, and similarly for $\Delta X_{t_{n,j+1}}^{(k)}$. We divide the sampling times into blocks of size $M > p$ (the last block may be longer). The number of blocks is K_n, where K_n is the largest integer $\leq n/M$. We denote the block boundaries by $\tau_{n,i} = t_{n,Mi}$, except for the last block.

4.1 Model Selection through F-testing

A standard way of selecting a model is through (possibly sequential) F-testing (see, for example, Weisberg (1985), Chapter 4.4 and 8.7). We here present a version of such testing which applies to the system (25). Specifically, we consider the following nested sequence of models ($0 = q_0 < q_1 <$

$\ldots < q_\nu = p$):

$$\begin{aligned}
\text{Model}_0: &\quad dY_t = dZ_t^{(0)} \\
\text{Model}_1: &\quad dY_t = \sum_{k=1}^{q_1} \beta_t^{(k)} dX_t^{(k)} + dZ_t^{(1)} \\
\text{Model}_2: &\quad dY_t = \sum_{k=1}^{q_2} \beta_t^{(k)} dX_t^{(k)} + dZ_t^{(2)} \\
&\quad \ldots \\
\text{Model}_\nu: &\quad dY_t = \sum_{k=1}^{q_\nu} \beta_t^{(k)} dX_t^{(k)} + dZ_t^{(\nu)} \qquad (26)
\end{aligned}$$

where $\langle X^{(k)}, Z^{(1)} \rangle_t = 0$ for all t and $k \leq q_1$, $\langle X^{(k)}, Z^{(2)} \rangle_t = 0$ for all t and $k \leq q_2$, and so on.

We want to test if the variation in Y is significantly better explained by adding regressees (hedging instruments). Finding the minimal number of hedging instruments will provide the most parsimonious and also cheapest strategy in the sense of incurring least transaction cost.

The procedure is as follows. For each block i, form the following residual sums of squares (RSS)

$$RSS_{i,q} \;=\; \text{RSS in regression of } \Delta Y_{t_{n,j}} \text{ on } \Delta X_{t_{n,j}}^{(1)}, \Delta X_{t_{n,j}}^{(2)}, \ldots, \Delta X_{t_{n,j}}^{(q)}, Mi < j \leq M(i+1), \qquad (27)$$

and similarly for the last block. We denote $RSS_{i,0} = \sum \Delta Y_{t_{n,j}}^2$ over block i. To avoid having too many indices, for the moment we are suppressing the dependence on n. The regressions are in all cases without intercept. The sum of squares additionally explained (in block i) by model number m (going from q_{m-1} to q_m regressors) is given by $RSS_{i,q_{m-1}} - RSS_{i,q_m}$. An F-statistic to test the significance of this is given by

$$F_{i,m} = \frac{(RSS_{i,q_{m-1}} - RSS_{i,q_m})/(q_m - q_{m-1})}{RSS_{i,p}/(M-p)}. \qquad (28)$$

To cumulate across blocks, write

$$F_m = \frac{1}{K_n} \sum_{i=0}^{K_n - 1} F_{i,m}. \qquad (29)$$

The heuristic of the testing procedure would now be that if $\beta_t^{(k)} \equiv 0$ for $q_r < k \leq p$ (*i.e.*, additional models $r+1$ to ν contribute nothing), then, for $r < m \leq \nu$, F_m should be close to

$$f_{M-p} = \frac{M-p}{M-p-2}, \qquad (30)$$

where $f_b = b/(b-2)$ is the expected value of an F random variable with a and b degrees of freedom. We now investigate this heuristic rigorously.

Define stable convergence as in Mykland and Zhang (2007). Under the regularity conditions of that paper, we obtain

THEOREM 4. *Let $M > p + 4$. Suppose that the quadratic covariation matrix $\langle X, X\rangle_t'$ is of full rank for all $t \in [0, \mathcal{T}]$. Assume that $\beta_t^{(k)} \equiv 0$ on $[0, \mathcal{T}]$ for $q_r < k \leq p$. Then, as $n \to \infty$, $K_n^{1/2}(F_{r+1} - f_{M-p}, ..., F_\nu - f_{M-p})$ converges stably to a normal distribution with mean zero and covariance matrix with elements $(m_1, m_2 = r+1, ..., \nu)$:*

$$\alpha_{m_1,m_2} = \frac{2(M-p)^2}{(M-p-2)^2(M-p-4)} + \delta_{m_1m_2}\frac{1}{q_{m_1} - q_{m_1-1}}\frac{2(M-p)^2}{(M-p-2)(M-p-4)}, \tag{31}$$

where $\delta_{m_1m_2}$ is the Kronecker delta.

One can then test the null hypothesis that the trading strategy of Model r is the best attainable ($\beta_t^{(k)} \equiv 0$ on $[0, \mathcal{T}]$ for $q_r < k \leq p$) by comparing F_r with a normal $\mathrm{N(f_{M-p}, \alpha_{r+1,r+1}/K_n)}$ upper quantile. The multivariate result in Theorem 4 is useful for the sequential selection procedure.

Proof of Theorem 4. Let the distribution Q_n be as in Mykland and Zhang (2007). Since asymptotics is unaffected, take $n = MK_n$. Let $\mathcal{Y}_{n,i}$ be the information at time $\tau_{n,i}$. Note that from standard (normal distribution based) regression theory, $(RSS_{i,q_{m-1}} - RSS_{i,q_m})$, $m = r+1, ..., \nu$, and $RSS_{i,p}$ are independent given $\mathcal{Y}_{n,i}$ and under Q_n, with distributions $\chi^2 \times \langle Z, Z\rangle_{\tau_{n,i}}' \Delta t$, where the χ^2 distributions have $q_m - q_{m-1}$, $m = r+1, ..., \nu$, and $M - p$ degrees of freedom, respectively. Thus, conditionally on $\mathcal{Y}_{n,i}$, the vector $(F_{i,r+1}, ...F_{i,\nu})$ has mean $f_{M-p}(1, ..., 1)$, and covariance matrix given by $\mathrm{Cov}(F_{i,m_1}, F_{i,m_2} \mid \mathcal{Y}_{n,i}) = \alpha_{m_1,m_2}$, since $E((\chi_b^2)^{-2}) = 1/(b-2)(b-4)$. By standard martingale CLT considerations, this yields the result of the theorem under Q_n. The validity of the result under P follows Theorem 4 (and Proposition 1) of Mykland and Zhang (2007) (there is no adjustment from these theorems). □

4.2 Cross Validation

Another way to approach the question of model selection is through cross validation (see, for example, Allen (1974), O'Sullivan (1986), and Breiman and Spector (1992)). We shall consider this in its simplest form, where we compute the residual quadratic variation $\langle Z, Z\rangle_\mathcal{T}$ by predicting an increment $\Delta Y_{t_{n,j}}$ based on $\Delta X_{t_{n,j}}^{(1)}$, $\Delta X_{t_{n,j}}^{(2)}$, ..., and $\Delta X_{t_{n,j}}^{(q)}$, where the coefficients β were estimated without the j-th observation. One do this estimate for a number of models (like (26), except that there is now no need for nesting), without fear of being overly optimistic due to overfitting). A good model with have a low (though not necessarily minimal) value of $\widehat{\langle Z, Z\rangle}_\mathcal{T}$.

We proceed using model (25). Of course, the number and choice of regressors can vary, and this give rise to different estimates.

The cross validated sum of squares (CVSS) is given, again using blocks in terms of cross validated residuals

$$\widehat{\Delta Z}_{t_{n,j}}^{CV} = \Delta Y_{t_{n,j}} - \sum_{k=1}^{p} \hat{\beta}_{(-j)}^{(k)} \Delta X_{t_{n,j}}^{(k)} \tag{32}$$

where $(\hat{\beta}^{(1)}_{(-j)}, ..., \hat{\beta}^{(p)}_{(-j)})$ is the least squares estimate in the block i containing j $(Mi < j \leq M(i+1))$ using all observations in the block except observation # j. The sum of squares is then

$$\text{CVSS} = \sum_{j=1}^{n} \left(\widehat{\Delta Z}^{CV}_{t_{n,j}}\right)^2, \tag{33}$$

A computationally efficient method for computing the residual $\widehat{\Delta Z}^{CV}_{t_{n,j}}$ is given in Weisberg (1985), equation (8.23) (p. 217), so computation does not require the fitting of n models.

Subject to the conditions of Mykland and Zhang (2007), we give the theoretical properties of the CVSS for high frequency data. It turns out that to achieve asymptotic unbiasedness, the estimator needs a multiplicative adjustment:

$$\widehat{\langle Z, Z\rangle}^{\text{CV}}_{\mathcal{T}} = \frac{M-p-2}{M-p-1}\text{CVSS} \tag{34}$$

To describe the asymptotic variance, define the matrix random variable $H = U(U^*U)^{-1}U^*$, where U is an $n \times p$ matrix consisting of iid standard normal random variables, and "$*$" denotes transposition. We now need the quantity

$$\begin{aligned}\gamma_{M,p} = {} & 3ME\left(\frac{1}{(1-H_{11})^2}\right) + M(M-1)E\left(\frac{1}{(1-H_{11})(1-H_{22})}\right) \\ & + 2M(M-1)E\left(\left(\frac{-H_{12}}{(1-H_{11})(1-H_{22})}\right)^2\right) - \left(\frac{M-p-2}{M-p-1}\right)^2 M^2.\end{aligned} \tag{35}$$

THEOREM 5. *Let $M > p$. Suppose that the quadratic covariation matrix $\langle X, X\rangle'_t$ is of full rank for all $t \in [0, \mathcal{T}]$. Define the cross validated quadratic variation (CVQV) of Z by (34). Then, as $n \to \infty$, $n^{1/2}(\widehat{\langle Z, Z\rangle}^{\text{CV}}_{\mathcal{T}} - \langle Z, Z\rangle_{\mathcal{T}})$ converges stably in law to*

$$\text{N}(0,1) \times \left(\left(\frac{\text{M}-\text{p}-2}{\text{M}-\text{p}-1}\right)^2 \gamma_{\text{M,p}} \frac{\mathcal{T}}{\text{M}} \int_0^{\mathcal{T}} (\langle \text{Z}, \text{Z}\rangle'_{\text{t}})^2 \text{dt}\right)^{1/2}, \tag{36}$$

where the normal random variable is independent of the underlying filtration $(\mathcal{F}_t)$.

In view of this result, one can use the cross validated sum of squares to set confidence intervals for the residual quadratic variation, in analogy with the results in Zhang (2001) and Mykland and Zhang (2006).

Proof of Theorem 5. As in the proof of Theorem 4, let the distribution Q_n be as in Mykland and Zhang (2007). Since asymptotics is unaffected, take $n = MK_n$. Let $\mathcal{Y}_{n,i}$ be the information at time $\tau_{n,i}$. Set $\Delta t = \mathcal{T}/n$, and also let let $\mathcal{Y}'_{n,i}$ be the smallest sigma-field containing $\mathcal{Y}_{n,i}$ and $\sigma(\Delta X_{t_{n,j}},\ \tau_{n,i} < t_{n,j} \leq \tau_{n,i+1})$.

For the moment, consider block # i. Let (h_{jl}) be the "hat matrix" (Weisberg (1985), p. 109) in the regression of ΔY or the ΔX's. Conditionally on $\mathcal{Y}'_{n,i}$, and under Q_n, the vector

$(\widehat{\Delta Z}^{CV}_{t_{n,j}}, Mi < j \leq M(i+1))$ is (*ibid,* (5.5) p. 110 and (8.23) p. 217) conditionally normal with mean zero and has a covariance matrix with elements

$$\langle Z, Z\rangle'_{\tau_i} \Delta t \frac{\delta_{jl} - h_{jl}}{(1-h_{jj})(1-h_{ll})}. \tag{37}$$

Hence, if we set

$$\text{CVSS}_i = \sum_{j=Mi+1}^{M(i+1)} \left(\widehat{\Delta Z}^{CV}_{t_{n,j}}\right)^2, \tag{38}$$

we obtain

$$\begin{aligned}
E_{Q_n}\left(\text{CVSS}_i \mid \mathcal{Y}'_{n,i}\right) &= \langle Z, Z\rangle'_{\tau_i} \Delta t \sum_{j=Mi+1}^{M(i+1)} \frac{1}{(1-h_{jj})}, \text{ and} \\
\text{Var}_{\text{Q}_\text{n}}\left(\text{CVSS}_\text{i} \mid \mathcal{Y}'_{\text{n,i}}\right) &= \sum_{j,l=Mi+1}^{M(i+1)} 2\text{Cov}_{Q_n}\left(\widehat{\Delta Z}^{CV}_{t_{n,j}}, \widehat{\Delta Z}^{CV}_{t_{n,l}} \mid \mathcal{Y}'_{n,i}\right)^2 \\
&= 2\left(\langle Z, Z\rangle'_{\tau_i} \Delta t\right)^2 \sum_{j,l=Mi+1}^{M(i+1)} \left(\frac{\delta_{jl} - h_{jl}}{(1-h_{jj})(1-h_{ll})}\right)^2.
\end{aligned}$$

Now observe that the hat matrix is invariant to linear transformations of the data. Hence, by the nondegeneracy of quadratic covariation matrix $\langle X, X\rangle'_{\tau_i}$ the law of the matrix (h_{jl}) given $\mathcal{Y}_{n,i}$ is independent of $\mathcal{Y}_{n,i}$, and equals to the law of (H_{jl}). Hence, by iterated conditioning, and symmetry,

$$\begin{aligned}
E_{Q_n}\left(\text{CVSS}_i \mid \mathcal{Y}_{n,i}\right) &= \langle Z, Z\rangle'_{\tau_i} \Delta t M E_{Q_n}\left(\frac{1}{1-h_{MiMi}} \mid \mathcal{Y}_{n,i}\right) \\
&= \langle Z, Z\rangle'_{\tau_i} \Delta t M E\left(\frac{1}{1-H_{11}}\right), \text{ and similarly} \\
\text{Var}_{\text{Q}_\text{n}}\left(\text{CVSS}_\text{i} \mid \mathcal{Y}_{\text{n,i}}\right) &= \left(\langle Z, Z\rangle'_{\tau_i} \Delta t\right)^2 \gamma_{M,p}, \text{ where} \\
\gamma_{M,p} &= M\text{Var}\left(\frac{1}{1-\text{H}_{11}}\right) + \text{M}(\text{M}-1)\text{Cov}\left(\frac{1}{1-\text{H}_{11}}, \frac{1}{1-\text{H}_{22}}\right) \\
&\quad + 2ME\left(\frac{1}{(1-H_{11})^2}\right) + 2M(M-1)E\left(\left(\frac{-H_{12}}{(1-H_{11})(1-H_{22})}\right)^2\right)
\end{aligned}$$

To further calculate the result, note that $E(1/(1-H_{11})) = (M-p-1)/(M-p-2)$ (using (5A.3) (p. 291) in Weisberg (1985), along with p. 85 and 456 in Mardia, Kent, and Bibby (1979)). This also yields that $\gamma_{M,p}$ has the form in (35).

It follows that under Q_n, $\widehat{\langle Z, Z\rangle}_{\mathcal{T}}^{\text{CV}} - \langle Z, Z\rangle_{\mathcal{T}}$ is to adequate approximation the end point of a martingale with discrete-time predictable quadratic variation

$$\left(\frac{M-p-2}{M-p-1}\right)^2 \gamma_{M,p} \sum_i \left(\langle Z, Z\rangle'_{\tau_i} \Delta t\right)^2 \;=\; n^{-1}\left(\frac{M-p-2}{M-p-1}\right)^2 \gamma_{M,p} \frac{\mathcal{T}}{M} \int_0^{\mathcal{T}} (\langle Z, Z\rangle'_t)^2 dt + o_p(n^{-1}).$$

By standard martingale CLT considerations, this yields the result of the theorem under Q_n. The validity of the result under P follows Theorem 4 (and Proposition 1) of Mykland and Zhang (2007) □

5 Conclusion

In the above, we have described an implied volatility approach to completing markets. To implement the approach in practice, model selection methods are developed. Important open problems include how to carry out the model selection in the presence of market microstructure and asynchronous observations.

REFERENCES

Allen, D. M. (1974): "The Relationship between Variable Selection and Data Augmentation and a Method for Prediction," *Technometrics*, 16, 125–127.

Amin, K. and V. Ng (1997): "Inferring future volatility from the information in implied volatility in Eurodollar options: a new approach," *The Review of Financial Studies*, 10 (2), 333–367.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001): "The Distribution of Exchange Rate Realized Volatility," *Journal of the American Statistical Association*, 96, 42–55.

——— (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

Barndorff-Nielsen, O. E. and N. Shephard (2001): "Non-Gaussian Ornstein-Uhlenbeck-Based Models And Some Of Their Uses In Financial Economics," *Journal of the Royal Statistical Society, B*, 63, 167–241.

——— (2002): "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society, B*, 64, 253–280.

Beckers, S. (1981): "Standard deviations implied in option prices as predictors of future stock price variability," *Journal of Banking and Finance*, 5, 363–382.

Berestycki, H., J. Busca, and I. Florent (2000): "An Inverse Parabolic Problem Arising in Finance," *C.R.Acad.Sci.Paris Sér. I Math.*

——— (2002): "Asymptotics and Calibration of Local Volatility Models," *Quant. Finance*, 2, 61–69.

——— (2004): "Computing the Implied Volatility in Stochastic Volatility Models," *Communications on Pure and Applied Mathematics*, 62, 1–22.

Bick, A. (1995): "Quadratic-variation-based dynamic strategies," *Management Science*, 41 (4), 722–732.

Black, F. and M. Scholes (1973): "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637–654.

Blair, B.J., S. P. and S. Taylor (2001): "Forecasting S & P 100 volatility: the incremental information content of implied volavilities and high-frequency index returns," *Journal of Econometrics*, 105 (1), 5–26.

Breiman, L. and P. Spector (1992): "Submodel Selection and Evaluation in Regression. The X-Random Case," *International Statistical Review*, 60, 291–319.

Canina, L. and S. Figlewski (1993): "The informational content of implied volatility," *The Review of Financial Studies*, 6 (3), 659–681.

Carr, P., H. Geman, D. B. Madan, and M. Yor (2005): "Pricing Options on Realized Variance," *Finance and Stochastics*, 4.

Carr, P. and R. W. Lee (2007): "Realized Volatility and Variance: Options via Swaps," *RISK*, 20, 76–83.

Christensen, B. and N. Prabhala (1998): "The relation between implied and realized volatility," *Journal of Financial Economics*, 50 (2), 125–150.

Comte, F. and E. Renault (1998): "Long memory in continuous-time stochastic volatility models," *Mathematical Finance*, 8, 291–323.

Dacorogna, M. M., R. Gençay, U. Müller, R. B. Olsen, and O. V. Pictet (2001): *An Introduction to High-Frequency Finance*, San Diego: Academic Press.

Day, T. E. and C. M. Lewis (1992): "Stock-Market Volatility and The Information-Content of Stock Index Options," *Journal of Econometrics*, 52 (1-2), 267–287.

Delbaen, F. and W. Schachermayer (1995a): "The Existence of Absolutely Continuous Local Martingale Measures," *Annals of Applied Probability*, 4, 926–945.

——— (1995b): "The No-arbitrage property under a change of Numeraire," *Stochastics and Stochastics Reports*, 53 (3-4), 213–226.

——— (1998): "The Fundamental Theorem of Asset Pricing for Unbounded Stochastic Processes," *Mathematische Annalen*, 312, 215–250.

Derman, E. and I. Kani (1994): "Riding on a Smile," *Risk*, 7 (2), 32–39.

——— (1998): "Stochastic Implied Trees: Arbitrage Pricing with Stochastic Term and Strike Structure of Volatility," *Internat. J. Tehoret. Appl. Finance*, 1, 61–110.

Duffie, D. (1996): *Dynamic Asset Pricing Theory*, Princeton, New Jersey: Princeton University Press.

Engle, R. F. and C. Mustafa (1992): "Implied ARCH models for option prices," *Journal of Econometrics*, 52, 289–311.

Föllmer, H. and M. Schweizer (1991): "Hedging of Contingent Claims under Incomplete Information," in *Applied Stochastic Analysis*, ed. by H. Davis and R. Elliot, New York: Gordon and Breach, 389–414.

Föllmer, H. and D. Sondermann (1986): "Hedging of Redundant Contingent Claims," in *Contributions to Mathematical Economics*, ed. by W. Hildebrand and A. Mas-Colell, Amsterdam, The Netherlands: North Holland, 205–223.

Foster, D. and D. Nelson (1996): "Continuous record asymptotics for rolling sample variance estimators," *Econometrica*, 64, 139–174.

Gwilym, O. and M. Buckle (1997): "Forward/forward volatilities and the term structure of implied volatility," *Applied Economics Letters*, 4 (5), 325–328.

Hayashi, T. and P. Mykland (2005): "Evaluating hedging errors: An asymptotic approach," *Mathematical Finance*, 15, 1931–1963.

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

Hull, J. C. (2008): *Options, Futures, and Other Derivatives*, New York: Prentice Hall, 7th ed.

Jacod, J. (1994): "Limit of Random Measures Associated with the Increments of a Brownian Semimartingale," Tech. rep., Université de Paris VI.

Jacod, J. and P. Protter (1998): "Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations," *Annals of Probability*, 26, 267–307.

Jorion, P. (1995): "Predicting Volatility in the Foreign-Exchange Market," *Journal of Finance*, 50 (2), 507–528.

Kang, R., L. Zhang, and R. Chen (2009): "Forecasting Return Volatility in the Presence of Microstructure Noise," Working Paper, University of Illinois at Chicago.

Lamoureux, C. and W. Lastrapes (1993): "Forecasting stock return variance: Toward an understanding of stochastic implied volatilities," *The Review of Financial Studeis*, 6, 293–326.

Latane, H. and R. Rendleman (1976): "Standard Deviations of Stock Price Ratios Implied in Options Prices," *Journal of Finance*, 31, 369–381.

Lee, R. W. (2001): "Implied and Local Volatilities under Stochastic Volatility," *International Journal of Theoretical and Applied Finance*, 4, 45–89.

——— (2004): “The Moment Formula for Implied Volatility at Extreme Strikes,” *Mathematical Finance*, 14, 469–480.

——— (2005): “Implied Volatility: Statics, Dynamics, and Probabilistic Interpretation,” in *Recent Advances in Applied Probability*, ed. by R. Baeza-Yates, New York: Springer-VErlag.

Mardia, K. V., J. Kent, and J. Bibby (1979): *Multivariate Analysis*, London: Academic Press.

Merton, R. C. (1973): “The Theory of Rational Option Pricing,” *Bell journal of Economics and Management Science*, 4, 141–183.

Mykland, P. A. (2000): “Conservative delta hedging,” *Annals of Applied Probability*, 10, 664–683.

——— (2003a): “Financial options and statistical prediction intervals,” *Annals of Statistics*, 31, 1413–1438.

——— (2003b): “The interpolation of options,” *Finance and Stochastics*, 7, 417–432.

Mykland, P. A. and L. Zhang (2006): “ANOVA for Diffusions and Itô Processes,” *Annals of Statistics*, 34, 1931–1963.

——— (2007): “Inference for continuous semimartingales observed at high frequency (revised version),” Tech. rep., to appear in *Econometrica*.

——— (2008): “Inference for Volatility Type Objects and Implications for Hedging,” *Statistics and its Interface*, 1, 255–278.

O’Sullivan, F. (1986): “A Statistical Perspective on Ill-Posed Inverse Problems A Statistical Perspective on Ill-Posed Inverse Problems,” *Statistical Science*, 1, 502–518.

Pena, I., G. Rubio, and G. Serna (1999): “Why do we smile? On the determinants of the implied volatility function,” *Journal of Banking and Finance*, 23 (8), 1151–1179.

Renault, E. and N. Touzi (1996): “Options hedging and implicit volatilities in a stochastic volatility model,” *Mathematical Finance*, 6, 279–302.

Rubinstein, M. (1994): “Implied binomial trees,” *Journal of Finance*, 49 (3), 771–818.

Schewizer, M. (1990): “Risk-minimality and Orthogonality of Martingales,” *Stochastics*, 30, 123–131.

——— (1991): “Option Hedging for Semimartingales,” *Stochastic Processes and Applications*, 37, 339–363.

Schmalensee, R. and R. R. Trippi (1978): “Common Stock Volatility Expectations Implied by Option Premia,” *Journal of Finance*, 31 (1), 129–147.

Scott, L. (1992): “Information content of prices in derivative security markets,” *International Monetary Fund. Staff paper*, 39 (3), 596–616.

Sheikh, A. (1993): "The behavior of volatility expectations and their effects on expected returns," *Journal of Business*, 66 (1), 93–116.

Weisberg, S. (1985): *Applied Linear Regression*, New York: Wiley, second ed.

Zhang, L. (2001): "From Martingales to ANOVA: Implied and Realized Volatility," Ph.D. thesis, The University of Chicago, Department of Statistics.

——— (2006): "Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach," *Bernoulli*, 12, 1019–1043.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005): "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411.