2001

# Estimation of security excess returns from derivative prices and testing for risk-neutral pricing

Gurupdesh S. Pandher,
*University of Windsor*

*Econometric Theory*, **17**, 2001, 785–819. Printed in the United States of America.

# ESTIMATION OF EXCESS RETURNS FROM DERIVATIVE PRICES AND TESTING FOR RISK NEUTRAL PRICING

Gurupdesh S. Pandher
*DePaul University*

This paper develops an econometric framework for (i) estimating excess returns of the security process from high frequency derivative prices, (ii) testing for risk neutral pricing, and (iii) measuring premiums outside the no-arbitrage pricing model. The estimator is constructed by applying quasi-likelihood and Feynman–Kac theory to the risk neutral contingent claims pricing model to generate the optimal orthogonality restriction. The strong consistency and asymptotic normality of the estimator are established in the context of a nonstationary underlying state process. These results further imply that the estimator is robust to distributional assumptions on the underlying asset process. The proposed approach is applicable to any arbitrary derivative security, does not require estimation of the risk neutral probability measure, and has application to spot rate bond pricing models. A controlled diagnostic study based on generating the S&P500 index and calls verifies the ability of the estimators to correctly estimate security excess returns and test for risk neutral pricing. The estimator is invariant to call strikes, and larger samples constructed by cycling over shorter maturity options can be used to reduce its variance.

## 1. INTRODUCTION

The risk neutral valuation model for pricing derivative securities is based on the principle of finding a unique equivalent risk neutral probability measure that renders the underlying discounted asset process (e.g., stock, bond, index) a martingale and valuing contingent claims as expectations. This paper uses quasi-likelihood estimation and risk neutral martingale theory to develop an econometric framework for (i) estimating excess returns of the underlying security

This paper is based on my dissertation at Cornell University, and I thank the chairman of my graduate committee, Professor Robert Jarrow. I also thank Fred Arditti, Ren-Raw Chen, Rick Durret, Thomas Epps, Ali Fatemi, Robert Grauer, Yongmiao Hong, Dongcheol Kim, Robert Lind, Carl Luft, and Lorne Switzer for useful comments and discussions. Special thanks also to seminar participants at Concordia University (Montreal), Cornell University, DePaul University, Simon Fraser University, University of Virginia, Rutgers University, the 1999 European Finance Association Meeting (Paris), and the 1999 Southern Finance Association Meetings (Key West, Florida). Address correspondence to: Gurupdesh S. Pandher, Department of Finance, DePaul University, 1 East Jackson Blvd., Chicago, IL 60604, USA; e-mail: gpandher@mozart.fin.depaul.edu.

 0266-4666/01 $9.50 

from high frequency derivative prices, (ii) testing for risk neutral pricing, and (iii) measuring premiums outside the no-arbitrage pricing model. The strong consistency and asymptotic normality of the estimator are established in the context of a nonstationary underlying state process. The asymptotic properties further imply that the proposed estimator is robust and estimation holds when distributional assumptions on the underlying asset process assumed in the risk neutral model (e.g., Brownian motion) are relaxed. A diagnostic study is undertaken to resolve sample design issues such as impact of the strike level, strike replication, and shorter maturity cycles on estimation of excess returns.

The estimation framework exploits the relationship between an arbitrary claim's partial differential equation and probabilistic representations (Feynman–Kac theory) and uses continuous risk neutral pricing and quasi-likelihood theory to identify the optimal orthogonality condition for estimating excess returns from derivative prices sampled at discrete intervals. This estimate can be compared with excess returns estimated directly from the underlying asset price process. Significant departures from equivalence imply the existence of additional premiums in derivative prices outside the no-arbitrage pricing model.

Beyond the risk neutral application of the paper, the proposed estimation framework also has interesting empirical derivative pricing applications that are being explored. Market prices of risk can be readily constructed from derivative excess returns and volatility. Contingent claims can then be empirically priced with the risk neutral density derived from Girsanov's change of measure formula. Pandher (2000) extends the framework to estimate the volatility of the security process from high frequency derivatives prices.

Quasi-likelihood estimators for excess returns (and their variance) are obtained for both the derivative price process and the underlying asset price process (e.g., stock, bond, or index). The results on strong consistency and asymptotic normality of the estimator are distribution free and derived under a milder conditional second moment assumption. This is satisfied by a large class of stochastic processes with finite conditional second moments (or finite variation). Therefore, the proposed estimator is robust to the distributional assumptions of risk neutral martingale theory where the stochastics are driven by Brownian motion. However, when the underlying state process is close to being a Brownian motion, the quasi-likelihood estimator offers optimal and efficient estimation. Moreover, the final feasible estimator developed is a discretized version of the estimator implied by the continuous risk neutral pricing framework. The convergence results show that in addition to possessing the robustness property, the feasible estimator offers consistent estimation and is asymptotically normal.

There are a number of further implications of this econometric framework for derivatives. First, the methodology is very general and applicable to any arbitrary traded derivative including calls, futures, and swaps. Second, the estimation procedure inherits the optimality properties of the quasi-likelihood or estimation function (EF) framework (Godambe, 1960; Godambe and Heyde, 1987; Thavaneswaran and Thompson, 1986), ensuring that the statistical equations used to estimate the implied market rate of return are (i) unbiased and (ii) of minimum variance in the class of all linear estimating equations.

Third, much of the finance related stochastic processes literature has focused on estimation of parameters (e.g., drift, volatility) from the state process $X$ (e.g., index, stock, bond) using maximum likelihood, moment conditions, and nonparametric methods (Broze, 1997; Dohal, 1987; Hansen and Scheinkman, 1995; Florens-Zmirou, 1993; and others). This paper considers the quasi-likelihood estimation of the excess return parameter from the derivative process $V(X)$ overlying the state process (and also from $X$). Furthermore, much of the quasi-likelihood literature deals with estimation in a purely discrete or continuous context. Here, the setting is mixed where the estimating equations follow from the continuous risk neutral pricing model for contingent claims but where sampling of market derivative prices (and underlying asset prices) occurs at discrete, perhaps random, times.

Econometric issues connected to the use of discrete data for continuous-time derivative pricing models in other estimation frameworks (e.g., maximum likelihood estimation [MLE]) have been considered more recently by Chernov and Ghysels (1998), Duffie and Glynn (1998), Pedersen (1995), and others. This paper differs from the direction taken in this work both in focus of estimation and the estimation methodology. The existing literature has not dealt with estimation of excess returns from derivative prices $V(X)$. To construct this estimator, the proposed methodology first identifies a conditional martingale difference equation (CMDE) by constructing an Itô expansion of the discounted derivative process between two given sampling intervals under the risk neutral measure, then applies the Feynman–Kac result to reduce terms, and last introduces the parameter of interest (excess returns) by switching to the empirical measure. Once the CMDE is constructed, the optimal orthogonality restriction on the CMDE is obtained from quasi-likelihood theory. A discrete "feasible" estimator is next developed from this procedure in which all quantities are measurable with respect to information available at the beginning of each sampling period.

Fourth, the proposed method for testing the risk neutral hypothesis does not require estimation of the risk neutral probability measure from observed prices (Banz and Miller, 1978; Breedan and Litzenberger, 1978). The estimation regimes of Longstaff (1991) and Ait-Sahalia and Lo (1998) for call options estimate a nonparametric risk neutral probability density (histogram) from a sequence of calls with the same maturity but different strikes. Maximum likelihood estimation and testing are considered by Lo (1988). The approach of inverting market prices of options to estimate parameters of the risk neutral measure under parametric density models is pursued by Sherrick, Irwin, and Forster (1990). Bekaert, Hodrick, and Marshall (1997) discuss biases in tests of the expectations hypothesis of the term structure of interest rates.

Fifth, the arbitrage bond pricing models of Vasicek (1977), Brennan and Schwartz (1979), and Artzner and Delbaen (1987) require an "inversion of the term structure" to remove the market price of risk when valuing contingent claims as the initial step (the models of Ho and Lee [1986] and Health, Jarrow, and Morton [1992] take the bond price process and forward rate process, respectively, as exogenous and avoid the inversion). There are computational difficulties in this inversion because bond pricing formulae are highly nonlinear and the spot rate and bond price processes are not independent of the market price of risk. The econometric approach of this paper offers linear estimation of excess returns, averting the nonlinearity problem, and may offer an advantage in these models over calibration-based estimation.

The estimation methodology and its empirical properties are tested and verified using an extensive Monte-Carlo diagnostic study. The empirical study also enables resolution of important sample design issues. The S&P500 index and call options defined on it are simulated using historical trend and volatility. Excess returns and market prices of risk are estimated separately from both the index and call option prices under various scenarios to investigate the impact of the strike level, length of maturity cycle, and strike replication. Differences in the estimated excess returns from calls and the index quantify extra premiums not explained by the risk neutral pricing model.

The results of the diagnostic study verify the ability of the econometric model and estimators to estimate the excess returns correctly and test the hypothesis of risk neutral pricing. The call data generated in the empirical study are based on the risk neutral pricing model (Black–Scholes formula for calls), and estimates of the market price of risk from both the index and its derivative calls are very close for any given sample size. Therefore, the empirical study reveals that no premia are found when none should exist. Estimation is unaffected by the strike level of the call. It is also found that the addition of replicates based on different strikes does not improve the standard errors of the estimator due to dependence among strike replicates.

In the market setting, the vast majority of traded calls are of maturities less than 1 year. Therefore, it is not feasible to increase the sample size by extending the time to maturity. An alternative sampling design that cycles over calls of smaller (nonoverlapping) maturities is considered as a way to reduce the variance of estimators. It is found that sampling from cycles of shorter maturities with the same effective sample size (number of cycles times maturity length) yields similar and stable estimation as a single maturity sample of larger but equivalent duration. This result gives confidence in the applicability of the estimation methodology to market derivative prices where larger samples derived from sampling over multiple (overlapping) maturities can be used to reduce variance.

The remainder of the paper is organized as follows. Section 2 sets out the probability model and stochastic processes for the arbitrary derivative process and introduces the main features of the quasi-likelihood (EF) estimation framework. The estimator of excess returns from an arbitrary derivative price process and its variance are derived in Section 3. Section 4 establishes the strong consistency and asymptotic normality of the feasible EF estimator. The estimation of excess returns from the underlying asset process is considered in Section 5. Section 6 presents the results from the Monte-Carlo study in which an index and calls are simulated using the historical volatility and trend of the S&P500 index to verify the estimation and evaluate the impact of sample size, strike level, strike replication, and maturity length on the estimation. Conclusions follow in Section 7.

## 2. PRELIMINARIES: STOCHASTIC PROCESSES FOR DERIVATIVES AND QUASI-LIKELIHOOD ESTIMATION

This section defines the probability model and stochastics for an arbitrary derivative process and introduces the Feynman–Kac result required to develop the estimator of excess returns. The essential features of quasi-likelihood estimation (or estimating function theory) are also presented.

### 2.1. The Probability Model and Stochastic Process for Derivative Claims

Fix the probability space $(\Omega, \mathcal{F}_T, (\mathcal{F}_t)_{0\le t\le T}, P)$ where $(\mathcal{F}_t)_{0\le t\le T} = \{\mathcal{F}_t;\ 0 \le t \le T\}$ is the filtration defined on the event space $\Omega$ satisfying the usual conditions (i.e., filtration is right continuous and $\mathcal{F}_0$ contains all null sets of $\mathcal{F}_T$). The probability space is assumed large enough to support an $\mathbb{R}^d$-valued stochastic processes $X = \{X_t, \mathcal{F}_T;\ 0 \le t \le T\}$ that is right continuous with left limits (RCLL) whose elements generate the $\sigma$-fields $\mathcal{F}_t = \sigma\{X_s;\ 0 \le s \le t\}$. The process $X$ will represent the state variable of the pricing model (e.g., stock, bond, or index). We view $X$ as a diffusion process following the general stochastic differential equation $P$-a.s.:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \tag{2.1}$$

where $b(t, X_t)\colon [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ is the drift vector, $\sigma(t, X_t)\colon [0, T] \times \mathbb{R}^d \to \mathbb{R}^d X \mathbb{R}^d$ is the dispersion matrix (of rank $d$), and $W_t$ is a $d$-dimensional $\mathcal{F}_t$-measurable standard Brownian motion with respect to the probability measure $P$. Moreover, $b(t, X_t)$ and $\sigma(t, X_t)$ are taken to satisfy the global Lipschitz and linear growth conditions (see Karatzas and Shreve, 1991, p. 289). This ensures that there exists a strong-form solution to (2.1) relative to $W = \{W_t, \mathcal{F}_t;\ 0 \le t \le T\}$ and the process $X$ is square integrable over $[0, T]$. Last, define $a(t, X_t) = \sigma(t, X_t)\sigma^T(t, X_t)$ to be the diffusion matrix.

The preceding quantities are defined with respect to the empirical measure $P$. Let $Q$ be the unique equivalent risk neutral measure under which expectations of the $X$ process discounted at the risk-free spot interest rate process $r = \{r_t;\ 0 \le t \le T\}$ are $Q$-martingales where $r$ is the growth process of the money market discount factor $B(t,T) = \exp(-\int_{s=t}^{T} r_s\, ds)$. Risk neutral valuation theory (Harrison and Kreps, 1979; Harrison and Pliska, 1981) asserts that an attainable contingent claim can then be valued as a discounted expectation under the measure $Q$.

The process of making the discounted asset a martingale requires the transformation

$$dW_t = d\widetilde{W}_t - \gamma_t(t, X_t)dt, \tag{2.2}$$

where the market price of risk $\gamma(t, X_t)$: $[0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\widetilde{W}_t$ is a Brownian motion with respect to the risk neutral measure $Q$. The relationship between the equivalent measures $P$ and $Q$ is readily obtained from Girsanov's change of measure formula. Note that the existence of $\gamma_t$ follows from the nonsingularity of $\sigma(t, X_t)$ (see Harrison and Pliska, 1981). Substituting (2.2) into (2.1) leads to the differential equation

$$dX_t = r_t X_t\, dt + \sigma(t, X_t)\, d\widetilde{W}_t. \tag{2.3}$$

Let $f(t, X_t)$:$[0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ be a function in the class $C^2([0,T] \times \mathbb{R}^d)$ defined on the state variable process $X$ with second order differential operator

$$\lim_{s \to 0} \frac{E^Q(f(t+s, X_{t+s}) - f(t, X_t) | \mathcal{F}_t)}{s} = (\mathrm{A}_t f)(x),$$

where $(\mathrm{A}_t f)(x) \equiv \frac{1}{2}\sum_{j=1}^{d} \sum_{k=1}^{d} a_{j,k}(t,x)[\partial f(x)/\partial x_j\, \partial x_k] + \sum_{j=1}^{d} r_t X_t [\partial f(x)/\partial x_j]$.

We are now ready to define the value process for an arbitrary derivative security. Let $V(X) = \{V(t, X_t), \mathcal{F}_t;\ 0 \le t \le T\}$ be the generic value process of the derivative claim based on the state variable process $X$ where $V(t, X_t)$: $[0,T] \times \mathbb{R}^d \to \mathbb{R}$ is in the class $C^2([0,T] \times \mathbb{R}^d)$. It is known from the Feynman–Kac theorem that when $X$ is a diffusion, $V(X)$ has a heat equation representation and a corresponding probabilistic representation as a discounted $Q$-martingale. This result is stated here for reference.

Feynman–Kac Result (Karatzas and Shreve, 1991, p. 366). Let $V(t, X_t) \in C^2([0,T] \times \mathbb{R}^d)$ and $B(t,T)$ be as defined earlier and consider the continuous functions $g(t, X_t)$: $[0,T] \times \mathbb{R}^d \to \mathbb{R}$ and $r(t, X_t)$: $[0,T] \times \mathbb{R}^d \to \mathbb{R}$ satisfying certain bounded conditions. If $V(t, X_t)$ satisfies the heat equation

$$\begin{aligned} &\frac{\partial V}{\partial t} + A_t V - rV = -g \quad \text{in } [0,T] \times \mathbb{R}^d \\ &\qquad V(T,x) = f(x), \qquad x \in \mathbb{R}^d, \end{aligned} \tag{2.4}$$

then $V(t, X_t)$ admits the unique stochastic representation

$$V(t,x) = E_x^Q\left[ f(X_T)\exp\left\{-\int_{s=0}^{t} r(s,X_s)ds\right\} + \int_{s=t}^{T} g(s,X_s)\exp\left\{-\int_{t}^{s} r(u,X_u)du\right\} ds|\mathcal{F}_t\right] \quad \textbf{(2.5)}$$

on $[0,T] \times \mathbb{R}^d$.

An analytical expression for the quasi-likelihood estimator of excess returns $\lambda = b - r$ is possible for any arbitrary diffusion price process when the volatility $\sigma(t, X_t)$ and drift $b(t, X_t)$ are "time-state separable": $b(t, X_t) = b_t\beta(X_t)$ and $\sigma(t, X_t) = \sigma_t\phi(X_t)$. We will retain this assumption for the econometrics of this paper.

## 2.2. The Quasi-Likelihood Estimation Framework

Estimating function theory (a.k.a. quasi-likelihood) provides a general framework for parameter estimation that includes maximum likelihood estimation (MLE) as a special case when an exact distribution is specified for the data generating process and incorporates least squares (LS) estimation for linear models with no distributional assumptions. It borrows the strengths of both approaches while eliminating their weaknesses. For example, LS estimation becomes biased when the variance of the dependent process depends on parameters appearing in the mean (see Godambe and Kale, 1991). For a further overview of EF theory, see Heyde (1989) and Godambe and Heyde (1987).

Assuming a discrete setting, the general approach to identifying the optimal estimating equation for the parameter $\theta \in \mathbb{R}^d$ is to first form estimating functions $H = (h_j, \mathcal{F}_j)$ of the data $Y$ and the parameter $\theta$ from a particular class of functions (e.g., linear) such that $E(h_j|\mathcal{F}_j) = 0$, $j = 1, \ldots, n$, with $\mathcal{F}_{j-1} \subset \mathcal{F}_j$. The optimality criterion of Godambe (1960) (or its sufficient versions) can then be applied to determine the optimal estimating equations $H^* = (h_j^*, \mathcal{F}_j)$. In relation to generalized method of moments (GMM) estimation (Hansen, 1982), $H^*$ may be viewed as the optimal orthogonality system. The EF framework, therefore, gives a systematic framework for identifying the optimal estimating function starting with a primitive "error" or martingale difference restriction $E(h_j|\mathcal{F}_j) = 0$, $j = 1, \ldots, n$.

The stress on the estimating equation, as opposed to the parameter estimator, of this framework is justified by the following observations: (i) Fischer's information and the Cramer–Rao inequality are both an estimating equation property rather than that of the MLE; (ii) asymptotic properties of an estimator are almost invariably obtained, as in the case of the MLE, via asymptotics of the estimating equation; (iii) estimating equations have the property of invariance under one-to-one transformation of the estimator; and (iv) separate estimating functions can be combined more easily than the estimators implicitly defined by them.

We will be interested in finding the optimal estimating equation in the class of linear $\mathcal{F}_j \equiv \mathcal{F}_{t_{j-1}}$ measurable estimating equations such as

$$H = \left\{ H\colon H = \sum_{j=1}^{n} \alpha_j(\theta) h_j(\theta) \right\}, \tag{2.6}$$

where $\alpha_j(\theta)$ is a predictable $\mathcal{F}_{t_{j-1}}$-measurable process and $E(h_j(\theta) | \mathcal{F}_{t_{j-1}}) = 0$, $j = 1, \ldots, n$.

The optimal choice of $\alpha_j(\theta)$ is given by

$$\alpha_j^* = \left( E \frac{\partial h_j}{\partial \theta} \middle| \mathcal{F}_{t_{j-1}} \right)' (E h_j h_j' | \mathcal{F}_{t_{j-1}})^{-1}, \qquad j = 1, \ldots, n. \tag{2.7}$$

which was shown by Godambe (1960) to minimize the (conditional) variance of the "standardized estimating equation"

$$H^s = \left( E \frac{\partial H}{\partial \theta} \right)^{-1} H \tag{2.8}$$

with

$$\mathrm{Var}(H^s) = \left( E \frac{\partial H}{\partial \theta} \right)^{-1} E(HH') \left( E \frac{\partial H}{\partial \theta} \right)^{-1}. \tag{2.9}$$

The criterion of minimizing $\mathrm{Var}(H^s)$ is justified by the dual objective of (i) minimizing $E(HH')$ and (ii) maximizing the sensitivity of the estimating function to departures from the true parameter value $(\partial H / \partial \theta)$.

## 3. ESTIMATION OF EXCESS RETURNS FROM DERIVATIVE PRICES

We begin by discussing the structure of the derivative market data to be used in the EF estimation of excess risk returns from derivative prices. Let the observed prices for the derivative security be sampled at the points in the sequence $\{t_0, t_1, \ldots, t_n\} \in [0, T]$ with $t_0 = 0$ indexing the start of the sampling period and $t_n = T$ representing the time to maturity. Then, $\Delta_j = t_j - t_{j-1}, j = 1, \ldots, n$ is the length of the period between points in the term structure. At each sampling point $t_j$, a cross section of replicate prices may exist indexed by $k = 1, \ldots, m$ (e.g., calls of different strikes $K_k$). Further, prices for nonoverlapping cycles of maturity times are available given by the sequence $\{T_1, \ldots, T_g, \ldots, T_p\}$. The price data consist of a sequence of market prices on the derivative claim given by $\{V_k(t_j, T_g) \equiv V(t_j, X_{t_j}; K_k, T_g),\ g = 1, \ldots, p,\ j = 1, \ldots, n,\ k = 1, \ldots, m\}$. The exact structure of the price sequence $V_k(t_j, T_g)$ will depend on the sample design (single/multiple maturities, strike replicates, and length of maturity cycles).

In the context of obtaining the estimating function for the excess return parameter $\lambda$ from derivative prices, we will consider the class of linear estimating functions given by

$$H = \left\{ H\colon H = \sum_{g=1}^{p} \sum_{j=1}^{n} \sum_{k=1}^{m} \alpha_{jkg}(\lambda) h_{jkg}(\lambda) \right\}. \tag{3.1}$$

We described the main features of the estimating function theory in Section 2.2. The key remaining issues are the specification of the martingale difference functions $h_{jkg}(\lambda)$ and the weighting factors $\alpha_{jkg}(\lambda)$, $g = 1, \dots, p$, $j = 1, \dots, n$, $k = 1, \dots, m$ in (3.1). The optimal estimating function and implied quasi-likelihood estimator for $\lambda$ can then be identified by choosing $\alpha^*_{jkh}(\lambda)$ optimally and will depend on the sample design used. Three cases are considered: (i) single strike and maturity $(T)$, (ii) multiple strike replicates on a single maturity, and (iii) replicates on multiple nonoverlapping maturity cycles. The key results relating to the EF estimator of $\lambda$ and its variance are derived in Propositions 1–8. The strong asymptotic consistency and normality of the estimator are established in Section 4. Without loss of generality, we start by obtaining the estimating function for $\lambda$ under the first case $(p = 1, m = 1)$.

PROPOSITION 1 (The Estimating Function $h_j(\lambda)$ for Excess Returns $\lambda$ [Single Strike and Maturity]. *Let* $g(t, X_t)\colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$ *and* $r(t, X_t)\colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$ *be a continuous function where the value process* $V(t, X_t) \in C^2([0, T] \times \mathbb{R}^d)$ *satisfies the partial differential equation*

$$\begin{aligned} &\frac{\partial V}{\partial t} + A_t V - rV = -g \quad \text{in } [0, T] \times \mathbb{R}^d \\ &\qquad V(T, X_T) = f(X_T), \qquad X_T \in \mathbb{R}^d. \end{aligned} \tag{3.2}$$

*Then, given the market derivative prices* $\{V(t_j, T) \equiv V(t_j, X_{t_j}; K, T)$, $j = 1, \dots, n\}$, *the estimating functions* $h_j(\lambda)$, $j = 1, \dots, n$ *in the linear class*

$$H = \left\{ H\colon H = \sum_{j=1}^{n} \alpha_j(\lambda) h_j(\lambda) \right\} \tag{2.6}$$

*are given by* $(d = 1)$

$$\begin{aligned} h_j(\lambda) = Y_j - \lambda \tilde{Z}_j = V(t_j, T) B(t_{j-1}, t_j) - V(t_{j-1}, T) + \int_{u=t}^{s} g(u, X_u) B(t_{j-1}, u) du \\ - \lambda \int_{t_{j-1}}^{t_j} \left( \frac{\partial V(u, T)}{\partial X} \right) X_u B(t_{j-1}, u) du, \end{aligned} \tag{3.3}$$

*where* $\lambda = b - r$ *is the excess risk return.*

Proof. See Appendix A. The financial interpretation of (3.3) is as follows. The first three terms of $h_j(\lambda)$ represent the dependent "$Y$" observation in the regression sense, whereas the last term represents the corresponding independent "$X$" variable. The first two terms give the change in the discounted value of the contingent claim observed over the sampling interval. The third term adds back the discounted dividends paid out over this period. In the case of European call options, this term is zero. The fourth integral term involving the "delta" of the derivative claim represents the cumulative discounted value of the underlying asset held to replicate the change in the claim's value over the interval (plus dividends). Therefore, net change in the value of the claim minus its hedge replication should be approximately zero.

Some of the basic properties of $h_j(\lambda)$ required in determining the estimator are summarized in Proposition 2.

PROPOSITION 2 (Properties of the Estimating Function $h_j(\lambda)$). *The first and second moments of $h_j(\lambda)$, $j = 1, \ldots, n$ determined in Proposition 1 satisfy*

*(i)* $E(h_j(\lambda)|\mathcal{F}_{t_{j-1}}) = 0.$

*(ii)* $E(h_j^2(\lambda)|\mathcal{F}_{t_{j-1}}) = \sum_{j=1}^{d} \int_{u=t_{j-1}}^{t_j} E\left(\left(\frac{\partial V(u,T)}{\partial X_j}\right)^2 (X_u\sigma)^2 \middle| \mathcal{F}_{t_{j-1}}\right) B_u^2\, du.$

*(iii)* $E(h_j(\lambda)h_k(\lambda)|\mathcal{F}_{t_{j-1}}) = 0,\ k > j.$

*(iv)* $\dfrac{\partial h_j(\lambda)}{\partial \lambda} = \int_{u=t_{j-1}}^{t_j} \left(\frac{\partial V(u,T)}{\partial X}\right) X_u B_u\, du \ (d = 1 \text{ case}).$

Proof. Property (i) is immediate from the definition of $h_j(\lambda)$, which is a stochastic integral w.r.t. Brownian motion with probability measure $P$. The second property follows from the isometry property of the squared stochastic integral (Karatzas and Shreve, 1991, p. 137) which in this case delivers

$$E(h_j^2(\lambda)|\mathcal{F}_{t_{j-1}}) = \sum_{j=1}^{d} \int_{u=t_{j-1}}^{t_j} E\left(\left(\frac{\partial V(u,T)}{\partial X_j}\right)^2 (X_u\sigma)^2 \middle| \mathcal{F}_{t_{j-1}}\right) B_u^2\, d[W^j]_u, \qquad \textbf{(3.4)}$$

where $d[W^j]_u = du$ is the quadratic variation of the Brownian motion $W_u^j$. The third result follows from the disjointness of the stochastic integrals, and the fourth property is immediate from (3.3). ■

PROPOSITION 3 (Optimal Estimating Equation, Estimator for $\lambda$ and Variance: Single Strike and Maturity). *Let* $\{V(u,T) \equiv V(u, X_u; T),\ u = t_j,\ j = 1, \ldots, n\}$ *define the sequence of derivative prices with maturity T. Then the following hold:*

*(i) The optimal estimating function for $\lambda$ in the linear class of estimating equations*

$$H = \left\{H\colon H = \sum_{j=1}^{n} \alpha_j(\lambda) h_j(\lambda)\right\} \qquad \textbf{(2.6)}$$

*is given by*

$$
\begin{aligned}
H^*(\lambda) &= \sum_{j=1}^{n} \frac{E\widetilde{Z}_j}{E\widetilde{W}_j} (Y_j - \lambda \widetilde{Z}_j) \\
&= \sum_{j=1}^{n} \left( \frac{\int_{u=t_{j-1}}^{t_j} E\left( \frac{\partial V(u,T)}{\partial X} X_u \middle| \mathcal{F}_{t_{j-1}} \right) B(t_{j-1},u)\,du}{\int_{u=t_{j-1}}^{t_j} E\left( \left( \frac{\partial V(u,T)}{\partial X} \right)^2 (X_u \sigma)^2 \middle| \mathcal{F}_{t_{j-1}} \right) B^2(t_{j-1},u)\,du} \right) \\
&\quad \times \left( V(t_j,T)B(t_{j-1},t_j) - V(t_{j-1},T) + \int_{u=t_{j-1}}^{t_j} g(u,X_u)B(t_{j-1},u)\,du \right. \\
&\qquad \left. - \lambda \int_{t_{j-1}}^{t_j} \frac{\partial V(u,T)}{\partial X} X_u B(t_{j-1},u)\,du \right),
\end{aligned}
\tag{3.5}
$$

*yielding*

$$
\hat{\lambda}_a^* = \frac{\sum_{j=1}^{n} E\widetilde{Z}_j (E\widetilde{W})_j^{-1} Y_j}{\sum_{j=1}^{n} E\widetilde{Z}_j (E\widetilde{W})_j^{-1} \widetilde{Z}_j}
$$

*and*

$$
\mathrm{Var}(\hat{\lambda}_a^*) = E_Z \left( \frac{1}{\sum_{j=1}^{n} E\widetilde{Z}_j (E\widetilde{W})_j^{-1} \widetilde{Z}_j} \right),
$$

*where the conditioning is done over the "stochastic regressors"* $\widetilde{Z} = (\widetilde{Z}_1, \ldots, \widetilde{Z}_n)$. *The conditional finite sample distribution of* $\hat{\lambda}_{a|\widetilde{Z}}^*$ *is given by*

$$
\hat{\lambda}_{a|\widetilde{Z}}^* \sim N\left( \lambda, \frac{1}{\sum_{j=1}^{n} E\widetilde{Z}_j (E\widetilde{W})_j^{-1} \widetilde{Z}_j} \right).
$$

(ii) *Define* $Y_j = V(t_j,T)B(t_{j-1},t_j) - V(t_{j-1},T) + \int_{u=t_{j-1}}^{t_j} g(u,X_u)B(t_{j-1},u)\,du$, $Z_j = \int_{u=t_{j-1}}^{t_j} (\partial V(t_{j-1},T)/\partial X) X_{t_{j-1}} B(t_{j-1},u)\,du$, *and* $W_j = \int_{u=t_{j-1}}^{t_j} (\partial V(t_{j-1},T)/\partial X)^2 \times (X_{t_{j-1}}\sigma)^2 B^2(t_{j-1},u)\,du$. *Then, the feasible* EF *estimator for* $\hat{\lambda}_a^*$ *and its variance* $V(\hat{\lambda}_a^*)$ *implied by the optimal estimating function are*

$$
\hat{\lambda}_a = \frac{\sum_{j=1}^{n} Z_j W_j^{-1} Y_j}{\sum_{j=1}^{n} Z_j W_j^{-1} Z_j}
\tag{3.6}
$$

*and*

$$\widehat{\mathrm{Var}}(\hat{\lambda}_a) = \frac{1}{\sum_{j=1}^{n} Z_j W_j^{-1} Z_j}. \qquad \textbf{(3.7)}$$

Some remarks are in order before discussing the proof of Proposition 3.

(i) The direct EF estimator $\hat{\lambda}_a^*$ is not computable because it requires information in the interval $[t_{j-1}, t_j]$ that is not available between sampling points.
(ii) Also, $\tilde{Z}_j$ is a random variable with respect to the information set $\mathcal{F}_{t_{j-1}}$. The feasible EF estimator is developed by replacing unknown quantities $\hat{\lambda}_a^*$ with their $\mathcal{F}_{t_{j-1}}$-measurable surrogates defined in Proposition 3 to obtain $\hat{\lambda}_a$ (3.6).
(iii) The volatility parameter $\sigma$ is assumed constant only over the sampling period (e.g., day) and does not influence the feasible EF estimator $\hat{\lambda}_a$ as it cancels out in the numerator and denominator; however, it is required in the computation of the variance estimator (3.7) and is embedded in the weights $W_j$. It can be estimated consistently from the state price process $X$ using standard methods (Campbell, Lo, and McKinlay, 1997, p. 36):

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{j=1}^{n} (\ln(X_k) - \ln(X_{k-1}) - \hat{\alpha}\Delta_j)^2,$$

where $\hat{\alpha} = 1/N \sum_{j=1}^{n}(\ln(X_k) - \ln(X_{k-1}))$ and $N = \sum_{j=1}^{n} \Delta_j$.

The proof of Proposition 3 follows.

Proof. For the linear class of orthogonal estimating functions defined by $H$ in (2.6), the optimal estimating equation (3.5) follows from choosing $\alpha_j(\lambda)$ according to (2.7) and making use of the terms defined in Proposition 2:

$$\alpha_j^* = \left( E \frac{\partial h_j}{\partial \lambda} \middle| \mathcal{F}_{t_{j-1}} \right)' (E h_j h_j' | \mathcal{F}_{t_{j-1}})^{-1}$$

$$= \left( \frac{\int_{u=t_{j-1}}^{t_j} E\left( \frac{\partial V(u,T)}{\partial X_j} X_u \middle| \mathcal{F}_{t_{j-1}} \right) B(t_{j-1}, u)\, du}{\int_{u=t_{j-1}}^{t_j} E\left( \left( \frac{\partial V(u,T)}{\partial X_j} \right)^2 (X_u \sigma)^2 | \mathcal{F}_{t_{j-1}} \right) B^2(t_{j-1}, u)\, du} \right) = \frac{E\tilde{Z}_j}{E\tilde{W}_j},$$

$$j = 1, \ldots, n. \qquad \textbf{(3.8)}$$

Next note that $\alpha_j^*$ must be $\mathcal{F}_{t_{j-1}}$-measurable and the conditional expectations in (3.8) is not known inside the sampling interval. Therefore, to obtain the feasible estimator, we replace the conditional expectations in $\alpha_j^*$ by their best available $\mathcal{F}_{t_{j-1}}$-measurable surrogate to obtain

$$\hat{\alpha}_j^* = \frac{Z_j}{W_j}, \qquad j = 1, \ldots, n.$$

The estimator $\hat{\lambda}_a^*$ follows from solving $H^*(\lambda) = 0$, and $V(\hat{\lambda}_a^*)$ follows from applying the variance operator in two steps, first conditioning on the "stochastic regressors" $\tilde{Z} = (\tilde{Z}_1, \dots, \tilde{Z}_n)$ ($g = 0$ w.l.g.):

$$V(\hat{\lambda}_a^*) = E_Z[V(\hat{\lambda}_a^*|\tilde{Z})] + V_Z[E(\hat{\lambda}_a^*|\tilde{Z})]$$
$$= E_Z\left(\frac{1}{\sum_{j=1}^{n} E\tilde{Z}_j(E\widetilde{W})_j^{-1}E\tilde{Z}_j}\right) + V_Z[\lambda].$$

Conditional on $Z = (\tilde{Z}_1, \dots, \tilde{Z}_n)$, the finite sample normality of $\hat{\lambda}_{a|\tilde{Z}}^*$ follows directly from the "error-side" representation of $h_j$ given by the stochastic integral $h_j = \int_{u=t_{j-1}}^{t_j} V^X(u,T)X_u\sigma B(t_{j-1},u)\,dW_u$, where from Proposition 2, $E(h_j^2|\mathcal{F}_{t_{j-1}}) = \int_{u=t_{j-1}}^{t_j} E(V_X^2(u,T)X_u^2|\mathcal{F}_{t_{j-1}})\sigma^2B_u^2\,du \equiv E\widetilde{W}_j$ and $E(h_j|\mathcal{F}_{t_{j-1}}) = 0$. Using this in $\hat{\lambda}_{a|\tilde{Z}}^*$ directly along with the fact that a stochastic integral w.r.t. Brownian motion (with predictable integrands) is Gaussian yields the normality of the conditional EF estimator $\hat{\lambda}_{a|\tilde{Z}}^*$.

The feasible estimators $\hat{\lambda}_a$ and $V(\hat{\lambda}_a)$ are finally obtained by replacing the integrals and expectations in $\hat{\lambda}_{a|\tilde{Z}}^*$ and $V(\hat{\lambda}_{a|\tilde{Z}}^*)$ with their best available $\mathcal{F}_{t_{j-1}}$-measurable surrogates yielding (3.6) and (3.7). ■

It will be shown in Section 4 (Propositions 6–9) that the feasible EF estimator is strongly consistent and asymptotically normal. The next proposition gives the estimators for excess returns in the case of using derivative replicates (e.g., calls of different strikes with the same maturity). The introduction of replicates on the same underlying asset process introduces dependence among the martingale difference functions $h_{jk}(\lambda)$. Aside from this complication the development of the estimator follows Proposition 3, and the details are omitted for brevity. The final result is stated in Proposition 4.

PROPOSITION 4 (Optimal Estimating Equation, Estimator for $\lambda$ and Variance: Strike Replicates of Single Maturity). *Let* $\{V_k(u,T) \equiv V(u, X_u; K_k, T),\ u = t_j,\ j = 1, \dots, n,\ k = 1, \dots, m\}$ *define the sequence of derivative prices with replicates (e.g., strikes)* $k = 1, \dots, m$. *Also, let* $V_k^X(u,T) \equiv \partial V(u, X_u; K_k, T)/\partial X$ *and define* $Y_{jk} = V_k(t_j,T)B(t_{j-1},t_j) - V_k(t_{j-1},T) + \int_{u=t_{j-1}}^{t_j} g(u,X_u)B(t_{j-1},u)\,du$, $Z_{jk} = \int_{u=t_{j-1}}^{t_j} V_k^X(t_{j-1},T)X_{t_{j-1}}B(t_{j-1},u)\,du$, $W_{jk} = \int_{u=t_{j-1}}^{t_j} (V_k^X(t_{j-1},T)^2(X_{t_{j-1}}\sigma)^2B^2(t_{j-1},u)\,du$, $\mathrm{Cov}(Y_{jk},Y_{jl}) = \int_{u=t_{j-1}}^{t_j} E(V_k^X(u,T) \times V_l^X(u,T)(X_u\sigma)^2|\mathcal{F}_{t_{j-1}})B^2(t_{j-1},u)\,du$, *and* $\widehat{\mathrm{Cov}}(Y_{jk},Y_{jl}) = \int_{u=t_{j-1}}^{t_j} V_k^X(t_{j-1},T) \times V_l^X(t_{j-1},T)(X_{t_{j-1}}\sigma)^2B^2(t_{j-1},u)\,du$. *Then, the feasible* EF *estimator for* $\lambda$ *and its variance implied by the optimal estimating equation are*

$$\hat{\lambda}_b = \frac{\sum_{j=1}^{n}\sum_{k=1}^{m} Z_{jk}W_{jk}^{-1}Y_{jk}}{\sum_{j=1}^{n}\sum_{k=1}^{m} Z_{jk}W_{jk}^{-1}Z_{jk}} \tag{3.9}$$

*and*

$$\widehat{\mathrm{Var}}(\hat{\lambda}_b) = \frac{\sum_{j=1}^{n}\left[\sum_{k=1}^{m}\sum_{l=1}^{m} Z_{jk} W_{jk}^{-1}\, \widehat{\mathrm{Cov}}(Y_{jk}, Y_{jl}) W_{jk}^{-1} Z_{jk}\right]}{\left(\sum_{j=1}^{n}\sum_{k=1}^{m} Z_{jk} W_{jk}^{-1} Z_{jk}\right)^2}. \tag{3.10}$$

Proposition 5, which follows, gives the optimal estimating function, estimator, and its variance for excess risk returns if replicates of derivative prices (e.g., calls of different strikes) are used over multiple cycles of nonoverlapping maturities. The proof follows easily from Propositions 3 and 4 upon noting that the sequence of maturity cycles is nonoverlapping. Hence, summations over $g = 1, \ldots, p$ are analogous to summations over nonoverlapping intervals indexed by $j = 1, \ldots, n$, and so the maturity cycles do not affect the correlation structure induced by the underlying Brownian motion in the stochastic integrals.

PROPOSITION 5 (Optimal Estimating Equation, Estimator for $\lambda$ and Variance: Strike Replicates with Multiple Nonoverlapping Maturities). *Let* $\{V_k(u, T_g) \equiv V(u, X_u; K_k, T_g),\ u = t_j,\ g = 1, \ldots, p,\ j = 1, \ldots, n,\ k = 1, \ldots, m\}$ *define the sequence of derivative market prices with nonoverlapping sequence of maturities* $\{T_1, \ldots, T_p\}$ *and replicates (e.g., strikes)* $k = 1, \ldots, m$. *Also, let* $V_k^X(u, T_g) \equiv \partial V(u, X_u; K_k, T_g)/\partial X$ *and define* $Y_{jkg} = V_k(t_j, T_g) B(t_{j-1}, t_j) - V_k(t_{j-1}, T_g) + \int_{u=t_{j-1}}^{t_j} g(u, X_u) B(t_{j-1}, u)\,du$, $Z_{jkg} = \int_{u=t_{j-1}}^{t_j} V_k^X(t_{j-1}, T_g) \times X_{t_{j-1}} B(t_{j-1}, u)\,du$, $W_{jkg} = \int_{u=t_{j-1}}^{t_j} (V_k^X(t_{j-1}, T_g))^2 (X_{t_{j-1}}\sigma)^2 B^2(t_{j-1}, u)\,du$, *and* $\widehat{\mathrm{Cov}}(Y_{jkg}, Y_{jlg}) = \int_{u=t_{j-1}}^{t_j} V_k^X(t_{j-1}, T_g) V_l^X(t_{j-1}, T_g)(X_{t_{j-1}}\sigma)^2 B^2(t_{j-1}, u)\,du$. *Then, the feasible* EF *estimator for $\lambda$ and its variance implied by the optimal estimating function are*

$$\hat{\lambda}_c = \frac{\sum_{g=1}^{p}\sum_{j=1}^{n}\sum_{k=1}^{m} Z_{jkg} W_{jkg}^{-1} Y_{jkg}}{\sum_{g=1}^{p}\sum_{j=1}^{n}\sum_{k=1}^{m} Z_{jkg} W_{jkg}^{-1} Z_{jkg}} \tag{3.11}$$

*and*

$$\widehat{\mathrm{Var}}(\hat{\lambda}_c) = \frac{\sum_{g=1}^{p}\sum_{j=1}^{n}\left[\sum_{k=1}^{m}\sum_{l=1}^{m} Z_{jkg} W_{jkg}^{-1}\, \widehat{\mathrm{Cov}}(Y_{jkg}, Y_{jlg}) W_{jlg}^{-1} Z_{jlg}\right]}{\left(\sum_{g=1}^{p}\sum_{j=1}^{n}\sum_{k=1}^{m} Z_{jkg} W_{jk}^{-1} Z_{jkg}\right)^2}. \tag{3.12}$$

## 4. ASYMPTOTIC CONSISTENCY AND NORMALITY OF FEASIBLE EF EXCESS RETURNS ESTIMATOR

This section establishes the strong consistency and asymptotic normality of the feasible EF estimator of excess returns. In the diffusion context, the randomness is driven by stochastic integrals with respect to Brownian motion. Although this leads to an exact finite sample Gaussian distribution for the exact conditional EF estimator $\hat{\lambda}^*_{a|\tilde{Z}}$, the asymptotic results are important to situations where the underlying $X$ process is only approximately a Brownian diffusion (e.g., other stochastic processes such as Poisson jumps are mixed with the underlying diffusion). Moreover, the feasible EF estimator of Section 3 is a discretized approximation to the exact estimator implied by the continuous risk neutral pricing framework. It is important to establish its consistency and asymptotic normality.

Results on strong consistency and asymptotic normality imply that the feasible EF estimators developed in Section 3 are robust and continue to hold when the exact distribution of the underlying $X$ process is not completely known. The finite sample distributional assumption of a Brownian motion driving the diffusion process is replaced by a milder conditional second moment assumption.

Without loss of generality, the results on strong consistency and asymptotic distribution are obtained in the case of a single strike with multiple maturity cycles (see Propositions 1 and 3) and extend easily to the case of strike replicates. With a fixed time to maturity, the sample size can only be increased by cycling over nonoverlapping maturity cycles. Therefore, the effective sample size is $np$, and the estimators involve a double index over $j = 1, \dots, n$ and $g = 1, \dots, p$. To keep the notation simple, the asymptotics that follow will view $n$ as the effective sample size without any loss of generality. We keep in mind that the effective sample size becomes large only when the maturity cycles $p$ are increased while the number of sample points in each maturity cycle remains fixed.

### 4.1. Strong Consistency of $\hat{\lambda}_n$

The first result gives a bound in $L_2$ norm $(E(|\hat{\lambda}_n|^2)^{1/2})$ for the sequence $\{\hat{\lambda}_n\}$ that will be useful to establish consistency.

PROPOSITION 6 (A Bound[1] for the Sequence $\{\hat{\lambda}_n\}$ in $L_2$ Norm). *The sequence of* EF *estimators for* $\lambda$ *obtained in Proposition 3 given by*

$$\hat{\lambda}_n = \frac{\sum_{j=1}^{n} Z_j W_j^{-1} Y_j}{\sum_{j=1}^{n} Z_j W_j^{-1} Z_j} \text{ is bounded in } L_2 \text{ norm by } E\left( \sum_{j=1}^{n} \frac{E(X_{t_j}^2 | \mathcal{F}_{t_{j-1}})}{n^2 X_{t_{j-1}}^2} \right). \tag{4.1}$$

Proof. See Appendix B.

The next proposition establishes a sufficient condition for the strong consistency of the EF estimator $\hat{\lambda}_n$. Proposition 8 shows this condition is met, thereby establishing its strong consistency.

PROPOSITION 7 (Sufficient Condition for Strong Consistency of EF Estimator of $\lambda$). *The* EF *estimator of Proposition 3 is strongly convergent:*

$$\hat{\lambda}_n = \frac{\sum_{j=1}^{n} Z_j W_j^{-1} Y_j}{\sum_{j=1}^{n} Z_j W_j^{-1} Z_j} \to \lambda, \quad a.s., \text{ on the set } \left\{ \sum_{j=1}^{n} \frac{E(X_{t_j}^2 | \mathcal{F}_{t_{j-1}})}{n^2 X_{t_{j-1}}^2} \to 0 \right\}. \tag{4.2}$$

Proof. Define $S_n = \hat{\lambda}_n - \lambda$ and $E_n = \{\omega: |S_n| > \epsilon\}$, $\epsilon > 0$. We also have the following set relationships: $\lim\sup_{n\to\infty} E_n = \downarrow \lim_m \{\sup_{n>m}[|S_n| > \epsilon]\} \subseteq \{\sup_{n>1}[|S_n| > \epsilon]\}$ because $\{\sup_{n>m}[|S_n| > \epsilon]\}$ is a decreasing sequence of sets in $m$. Therefore, we have

$$\begin{aligned}
P(\limsup_{n\to\infty} |S_n| > \epsilon) &\le P\left(\sup_{n>1} |S_n| > \epsilon\right) \\
&= \lim_{n\to\infty} P\left(\max_{1<j<n} |S_n| > \epsilon\right) \\
&\le \lim_{n\to\infty} \frac{1}{\epsilon^2} E(|S_n|^2) \\
&= \lim_{n\to\infty} \frac{1}{\epsilon^2} (\lambda E(A_n^2) + E(B_n^2)) \\
&\le \frac{(\lambda^2 k_1 + k_2)}{n^2 \epsilon^2} E\left( \sum_{j=1}^{n} \frac{E(X_{t_j}^2 | \mathcal{F}_{t_{j-1}})}{X_{t_{j-1}}^2} \right) \\
&\to 0
\end{aligned} \tag{4.3}$$

as $n \to \infty$ by the hypothesis of Proposition 7 ($k_1$ and $k_2$ are positive constants). The third inequality of (4.3) follows from Kolgomorov's inequality (Hall and Heyde, 1980), the fourth inequality is obtained in the proof of Proposition 6 (see Appendix B, equation (B.7)), and the fifth inequality follows from Proposition 7. This establishes the result $\hat{\lambda}_n \to \lambda$, *a.s.*, on the set $\{\sum_{j=1}^{n} E(X_{t_j}^2 | \mathcal{F}_{t_{j-1}})/n^2 X_{t_{j-1}}^2 \to 0\}$. ∎

It now remains to show that the sufficient condition for the strong consistency of the EF estimator $\hat{\lambda}_n$ derived in Propositions 6 and 7 holds for stochastic differential equations satisfying conditions for strong solutions. This is verified in Proposition 8, thereby establishing the strong consistency of the EF estimator.

PROPOSITION 8 (The Sufficient Consistency Condition of Proposition 7 Is Satisfied). *For the diffusion process X defined in (2.1) satisfying the Lipschitz and growth conditions*

$$\sum_{j=1}^{n} \frac{E(X_{t_j}^2 | \mathcal{F}_{t_{j-1}})}{n^2 X_{t_{j-1}}^2} \to 0. \tag{4.4}$$

Proof. We assume the state variable $X$ follows the stochastic differential equation

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \tag{2.1}$$

where $b(t, X_t)$: $[0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ is the drift vector, $\sigma(t, X_t)$: $[0, T] \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ is the dispersion matrix (of rank $d$), and $dW_t$ is a $d$-dimensional $\mathcal{F}_t$-measurable standard Brownian motion with respect to the probability measure $P$.

If $b(t, X_t)$ and $\sigma(t, X_t)$ satisfy the global Lipschitz and linear growth conditions (see Karatzas and Shreve, 1991, p. 289), then an $L_2$ bound on $X$ given in Duffie (1992, p. 292) can be written as

$$E(|X_{t_j}|^2 | \mathcal{F}_{t_{j-1}}) \le Ce^{Ct}(1 + |X_{t_{j-1}}|^2) = O(|X_{t_{j-1}}|^2) \tag{4.5}$$

for some constant $P$. Because $X^2_{t_{j-1}}$ is $\mathcal{F}_{t_{j-1}}$ measurable, this leads to

$$\sum_{j=1}^{n} \frac{E(|X_{t_j}|^2 | \mathcal{F}_{t_{j-1}})}{n^2 |X_{t_{j-1}}|^2} = O\left(\frac{1}{n}\right) \to 0 \quad \text{as } n \to \infty. \qquad \blacksquare$$

### 4.2. Asymptotic Normality of $\hat{\lambda}_n$

By Proposition 3, the finite sample distribution of the conditional EF estimator is Gaussian because the underlying state process $X$ is driven by a Brownian motion and the terms in the estimator involve stochastic integrals w.r.t. this Brownian motion. The asymptotic distribution is relevant when this distributional assumption is relaxed and replaced by a weaker conditional second moment restriction. The asymptotic consistency and large sample normality allow inference for the feasible EF estimator even when the exact distribution of the underlying $X$ process is not completely known. The asymptotic normality of $\hat{\lambda}_n$ is established in Proposition 9, which follows.

PROPOSITION 9 (Asymptotic Normality). *If* $\hat{\lambda}_n \to \lambda$, *p*, *and* $E(|h_j|^2 \times |\mathcal{F}_{t_{j-1}}) \propto E(|X_{t_j}|^2 | \mathcal{F}_{t_{j-1}})$ *then*

$$\left(\sum_{j=1}^{n} Z_j W_j^{-1} Z_j\right)^{1/2} (\hat{\lambda}_n - \lambda) \Rightarrow N(0,1). \tag{4.6}$$

Proof. Define the conditional variance $I_n(\lambda) \equiv \sum_{j=1}^{n} Z_j W_j^{-1} \times E(h_j^2 | \mathcal{F}_{t_{j-1}}) W_j^{-1} Z_j$ and let $\dot{H}_n(\hat{\lambda}_n) \equiv \partial H_n(\lambda_n^*)/\partial\lambda = -\sum_{j=1}^{n} Z_j W_j^{-1} Z_j$. Then the first-order Taylor expansion of the feasible optimal estimation function $H_n(\hat{\lambda}_n) = \sum_{j=1}^{n} \hat{\alpha}_j^* h_j = \sum_{j=1}^{n} Z_j W_j^{-1} h_j$ yields

$$\begin{aligned} 0 = H_n(\hat{\lambda}_n) &= H_n(\lambda) + \left[\frac{\partial H_n(\hat{\lambda}_n^*)}{\partial\lambda}\right](\hat{\lambda}_n - \lambda) \\ &= H_n(\lambda_n) + \dot{H}_n(\lambda)(\hat{\lambda}_n - \lambda) + [-\dot{H}_n(\lambda) + \dot{H}_n(\hat{\lambda}_n^*)](\hat{\lambda}_n - \lambda), \end{aligned} \tag{4.7}$$

where $\hat{\lambda}_n^* = \gamma\hat{\lambda}_n + (1-\gamma)\lambda$. Suppose $a_n$ is an increasing sequence $a_n \to \infty$ chosen such that $a_n^{-1}\dot{H}_n(\lambda) \Rightarrow N(0,1)$ and rewrite (4.7) as

$$-a_n^{-1}\dot{H}_n(\lambda)(\hat{\lambda}_n - \lambda) = a_n^{-1}H_n(\lambda) + a_n^{-1}[-\dot{H}_n(\lambda) + \dot{H}_n(\hat{\lambda}_n^*)](\hat{\lambda}_n^* - \lambda), \qquad \textbf{(4.8)}$$

where in the case at hand $-\dot{H}_n(\lambda) + \dot{H}_n(\hat{\lambda}_n^*) = 0$ by the linearity of the estimating function. The left hand side converges in distribution to a standard normal variate if $a_n^{-1}H_n(\lambda) \Rightarrow N(0,1)$. It remains now to find this sequence $a_n$ and prove the convergence.

Consider the normalized conditional variance

$$\mathrm{Var}(a_n^{-1}H_n(\lambda)) = a_n^{-2}I_n(\lambda) = a_n^{-2}\sum_{j=1}^{n} Z_j W_j^{-1} E(h_j^2 \mid \mathcal{F}_{t_{j-1}}) W_j^{-1} Z_j. \qquad \textbf{(4.9)}$$

It is easy to check that if $E(|h_j|^2 \mid \mathcal{F}_{t_{j-1}}) \propto E(|X_{t_j}|^2 \mid \mathcal{F}_{t_{j-1}})$, then this conditional variance is bounded by the $\mathcal{F}_{t_{j-1}}$-measurable term $|X_{t_{j-1}}|^2$:

$$E(|h_j|^2 \mid \mathcal{F}_{t_{j-1}}) \propto E(|X_{t_j}|^2 \mid \mathcal{F}_{t_{j-1}}) \le Ce^{Ct}(1 + |X_{t_{j-1}}|^2) = O(|X_{t_{j-1}}|^2), \qquad \textbf{(4.10)}$$

where the bound follows from (4.5) and derives from the global Lipschitz and linear growth conditions on the drift and volatility parameters of the diffusion $X$.

It is easy to verify that (4.10) allows the normalized conditional variance (4.9) to be written as

$$\mathrm{Var}(a_n^{-1}H_n(\lambda)) = a_n^{-2}O\left(\sum_{j=1}^{n} Z_j W_j^{-1} Z_j\right), \qquad \textbf{(4.11)}$$

which implies choosing $a_n = (\sum_{j=1}^{n} Z_j W_j^{-1} Z_j)^{1/2}$ so that $\mathrm{Var}(a_n^{-1}H_n(\lambda)) = O(1)$. Because the conditional variance $I_n(\lambda)$ is bounded under the assumption of Proposition 10 (see 4.10), the result

$$a_n^{-1}H_n(\lambda) = \left(\sum_{j=1}^{n} Z_j W_j^{-1} Z_j\right)^{(-1)/2} H_n(\lambda) \Rightarrow N(0,1) \qquad \textbf{(4.12)}$$

follows from application of the central limit theorem for martingales with bounded conditional second moments $E(|h_j|^2 \mid \mathcal{F}_{t_{j-1}})$ (see Billingsley, 1986, Theorem 35.9, p. 498). This shows that the right hand side of (4.8) converges in distribution to a normal standard variate and gives the desired result for the left hand side:

$$-a_n^{-1}\dot{H}_n(\lambda)(\hat{\lambda}_n - \lambda) = \left(\sum_{j=1}^{n} Z_j W_j^{-1} Z_j\right)^{1/2} (\hat{\lambda}_n - \lambda) \Rightarrow N(0,1). \qquad \textbf{(4.13)}$$

■

## 5. ESTIMATING EXCESS RETURNS FROM THE UNDERLYING ASSET PROCESS

Estimation of excess returns from observed market derivative prices using EF theory was developed in Section 4. This section discusses the estimation of $\lambda$ from the underlying asset process (stock, bond, or index) on which the derivative securities are defined. The paper's methodology for estimating premiums and testing for risk neutral pricing rests on comparing these two estimates for equality.

The relevant class of linear estimating functions for the excess returns parameter $\lambda$ is

$$H = \left\{ H\colon H = \sum_{j=1}^{n} \alpha_j(\lambda) h_j(\lambda) \right\}. \tag{5.1}$$

The component functions $h_j(\lambda)$ for the stock process, the optimal weights $\alpha_j^*(\lambda)$, and the estimators for $\lambda$ and its variance are derived in the propositions that follow. Note that the issues of sample design pertaining to strike-based replicates and maturity cycles do not arise in this situation.

PROPOSITION 10 (The Estimating Function for $\lambda$ Based on the Stock Process $X$). *Let $X$ be a diffusion process following the stochastic differential equation*

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \quad P\text{-}a.s., \tag{5.2}$$

*where $b(t, X_t)\colon [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ is the drift vector, $\sigma(t, X_t)\colon [0, T] \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ is the dispersion matrix (of rank d), and $dW_t$ is a d-dimensional $\mathcal{F}_t$-measurable standard Brownian motion with respect to the empirical probability measure P. Then, under the further assumption of "time-state separability" (see Section 2.1), the conditional martingale difference functions $h_j(\lambda)$, $j = 1, \ldots, n$, in the linear class*

$$H = \left\{ H\colon H = \sum_{j=1}^{n} \alpha_j(\lambda) h_j(\lambda) \right\} \tag{5.3}$$

*are given by $(d = 1)$*

$$h_j(\lambda) = X(t_j)B(t_{j-1}, t_j) - X(t_{j-1}) - \lambda \int_{t_{j-1}}^{t_j} X(u)B(t_{j-1}, u)du, \tag{5.4}$$

*where $\lambda = b - r$ is the excess return.*

Proof. The proof is identical to the development of Proposition 1 with the discounted price process $f(X_s, B_s) = X_s B(t, s)$ forming the basis of the Itô expansion.

The optimal estimating equation and the EF estimator for and its variance estimator are stated in Proposition 11. The proof parallels Proposition 3.

PROPOSITION 11 (Optimal Estimating Equation, Estimator for $\lambda$ and Variance from Stock Process). *Let $\{X(t_j), j = 1, \ldots, n\}$ define the sequence of market stock (index) prices. The optimal estimating function for $\lambda$ is given by*

$$H^*(\lambda) = \sum_{j=1}^{n} \left( \frac{\displaystyle\int_{u=t_{j-1}}^{t_j} E(X_u | \mathcal{F}_{t_{j-1}}) B(t_{j-1}, u)\, du}{\displaystyle\int_{u=t_{j-1}}^{t_j} E((X_u \sigma)^2 | \mathcal{F}_{t_{j-1}}) B^2(t_{j-1}, u)\, du} \right) \times \left( X(t_j) B(t_{j-1}, t_j) - X(t_{j-1}) - \lambda \int_{t_{j-1}}^{t_j} X_u B(t_{j-1}, u)\, du \right). \tag{5.5}$$

*Define*

$$Y_j = X(t_j) B(t_{j-1}, t_j) - X(t_{j-1}) B(t_{j-1}), \qquad Z_j = \int_{u=t_{j-1}}^{t_j} X_{t_{j-1}} B(t_{j-1}, u)\, du$$

*and* $$W_j = \int_{u=t_{j-1}}^{t_j} (X_{t_{j-1}} \sigma)^2 B^2(t_{j-1}, u)\, du.$$

*Then, the feasible estimator for $\lambda$ and its variance implied by the optimal estimating function are*

$$\hat{\lambda}_d = \frac{\sum_{j=1}^{n} Z_j W_j^{-1} Y_j}{\sum_{j=1}^{n} Z_j W_j^{-1} Z_j} \tag{5.6}$$

*and*

$$\widehat{\mathrm{Var}}(\hat{\lambda}_d) = \frac{1}{\sum_{j=1}^{n} Z_j W_j^{-1} Z_j}. \tag{5.7}$$

## 6. DIAGNOSTIC STUDY

The empirical properties of the estimation framework developed in the previous sections are tested and evaluated using an extensive Monte-Carlo study. The empirical study also enables resolution of important sample design issues for derivative prices (impact of strike level and maturity length), verifies certain theoretical implications (e.g., strike replicates do not reduce variance—

compare (3.7) and (3.10)), and gives confidence in implementation to market derivative data. In the study, the S&P500 index and call options defined on it are simulated using historical trend and volatility, and excess returns are estimated from both the underlying asset and call option prices under various scenarios with differing strike levels and strike replication and shorter maturity lengths. Differences in the estimated excess returns quantify extra premiums not explained by the risk neutral pricing model. Major conclusions of the study are summarized first before detailing the analysis and results.

### 6.1. Major Findings

The results obtained from the diagnostic study verify the ability of the feasible EF estimator to correctly estimate excess returns from derivative prices and test the hypothesis of risk neutral pricing. The simulations are based on assumptions satisfying the risk neutral null hypothesis (Black–Scholes valuation of calls), and estimates of excess returns (and equivalently market price of risk) from both the index and its derivative calls are very close for any given sample size. Therefore, the empirical study reveals that no pricing errors and premiums are found when none should exist. Standard errors tend to be large but go down with larger sample sizes. Second, the estimation is found to be invariant to the strike level ("moneyness") of the calls. Third, it is found that addition of replicates based on different strikes does not improve the standard errors of the estimator because of high dependence among replicates.

In the market setting, it is rare to find calls of large maturity. Therefore, it is not feasible to increase the sample size by extending the time to maturity. An alternative sampling design that cycles over calls of smaller (nonoverlapping) maturities is considered as a way to improve precision. It is found that estimation remains stable when sampling from cycles of shorter maturities with the same effective sample size (cycles times maturity length). This result of the study gives confidence in the applicability of the estimation methodology to derivative market prices where large samples based on cycling over shorter maturities (3 months best) may be used to reduce variance.

### 6.2. Generation of S&P500 Index and Call Options

The S&P500 index used in the diagnostic study is generated using trend and historical volatility estimates obtained from the Bloomberg online system. Estimates of the annualized historical volatility over a 260 day (1 trading year) trading period over the last 6 months of 1998 fluctuated in the range 17–23%. For the empirical study, the annualized volatility was set at $\sigma = 20\%$. From a similar examination of the S&P500 index, an annualized price appreciation of $b = 9\%$ was chosen, and the starting value of the index was set at $X_0 = 1{,}000$ (the estimation holds uniformly for data generated under other parameter values for drift and volatility). The S&P index process was simulated as a discrete geometric Brownian motion using the recursive formula

$$X_j \equiv X_{t_j} = X_{j-1}\exp\left\{\left(b - \frac{\sigma^2}{2}\right)\Delta_j + \sigma\sqrt{\Delta_j Z}\right\}, \qquad j = 1,\ldots,n, \tag{6.1}$$

where $\Delta_j = 1/260$ and $Z \sim N(0,1)$ is a standard normal random variate (generated by a random number generator). Following a suggestion of an anonymous referee, the initial 1,000 recursions of (6.1) were discarded to remove "start-up" problems in the series, and the simulation size was expanded to 2,000 (from 100). This improved the performance of the excess return estimator even at the lowest sample size of 100 (see Table 1) and when call options are sampled from shorter maturity cycles (Table 3). The empirical study was carried out in the Gauss programming language.

Call options $\{V_k(t_j, T_g) \equiv V(t_j, X_{t_j}; K_k, T_g), g = 1,\ldots,p,\ j = 1,\ldots,n_g,\ k = 1,\ldots,m_g\}$ on the S&P500 index were generated by using (6.1) in the Black–Scholes formula:

$$V_k(t_j, T_g) \equiv V_k(t_j, X_{t_j}; K_k, T_g) = X_{t_j} N(d_1) - K\exp(-r\tau_j)N(d_2),$$

$$d_1 = \frac{\ln\left(\dfrac{X_{t_j}}{K}\right) + \left(r + \dfrac{1}{2}\sigma^2\right)\tau_j}{\sigma\sqrt{\tau_j}},$$

$$d_2 = d_1 - \sigma\sqrt{\tau_j}, \tag{6.2}$$

where $\tau_j = \sum_{i=j-1}^{n_g} \Delta_j, j = 1,\ldots,n_g$ is the time to maturity for each nonoverlapping maturity cycle with lengths $\{T_1,\ldots,T_g,\ldots,T_p\}$. A random draw of the index and calls (at the money) over 500 trading days is plotted in Figure 1. In the plot, the index value is added to the call prices.

### 6.3. Performance of EF Estimator and Effect of Sample Size

The performance of the proposed estimation framework in correctly estimating the excess return $\lambda = b - r = 9\%$, and equivalently the market price of risk $\gamma = \lambda/\sigma = .45$, as a function of sample size is first examined. It is important to note that these are "theoretical true values" under the assumption of a continuous geometric Brownian motion. As a result of the discretization involved in generating the index recursively with formula (6.2), some distortion is introduced, and the "actual true values" for the excess return and market price of risk will differ from 9% and .45, respectively. This should be kept in mind when comparing the performance of the estimator with the "true value" at different sample sizes.

Sample sizes (trading days; with 260 trading days per year) varying from 100 to 10,000 with single maturity and strike (at the money) were used in the

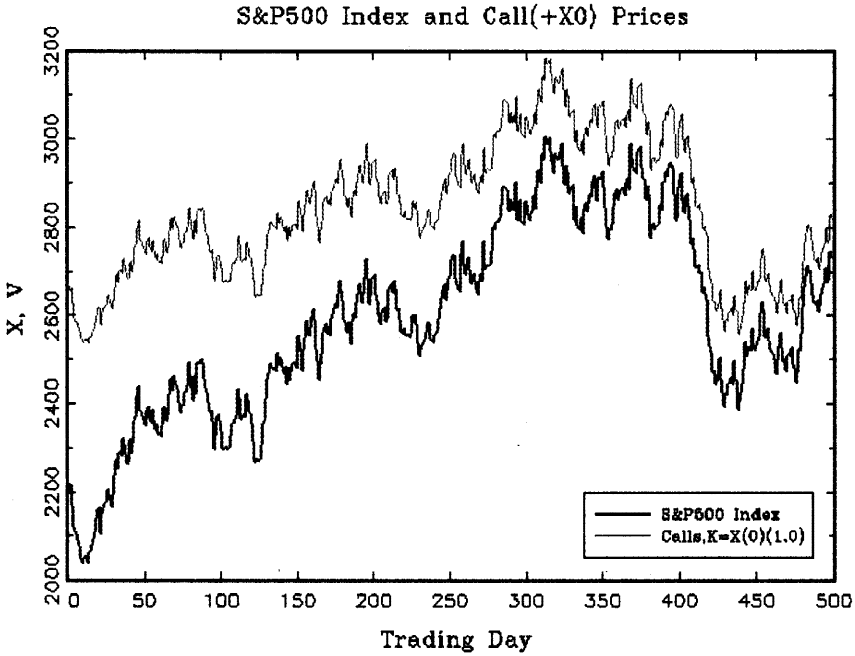


**FIGURE 1.** Random draw of S&P500 index and calls.

first experiment. The results averaged over 2,000 simulations are reported in Table 1. EF estimates of $\gamma(=.45)$ and $\lambda(=9\%)$ from both the call and S&P index are very close at each sample size. Even at a small sample size of 100, the mean of estimates is within 1.9% of the "theoretical true value."

**TABLE 1.** Estimating function estimates of $\lambda$ and $\gamma$ by sample size

| | | Sample Size (Trading Days) | | | | | |
|---|---|---|---|---|---|---|---|
| Estimator | Average | 100 | 500 | 1,000 | 2,000 | 5,000 | 10,000 |
| EF-call | $\hat{\gamma}_{c,EF}$ | .4584 | .4639 | .4585 | .4578 | .4517 | .4487 |
| | $\mathrm{Var}(\hat{\gamma}_{c,EF})$ | 2.600 | .5200 | .2600 | .1300 | .0520 | .0260 |
| | $\mathrm{SE}(\hat{\gamma}_{c,EF})$ | 1.612 | .7211 | .5099 | .3606 | .2280 | .1612 |
| | $\hat{\lambda}_{c,EF}$ | 9.17% | 9.28% | 9.17% | 9.16% | 9.03% | 8.97% |
| EF-index | $\hat{\gamma}_{x,EF}$ | .4596 | .4637 | .4585 | .4577 | .4517 | .4487 |
| | $\mathrm{Var}(\hat{\gamma}_{x,EF})$ | 2.600 | .5200 | .2600 | .1300 | .0520 | .02600 |
| | $\mathrm{SE}(\hat{\gamma}_{x,EF})$ | 1.612 | .7211 | .5099 | .3606 | .2280 | .1612 |
| | $\hat{\lambda}_{x,EF}$ | 9.19% | 9.27% | 9.17% | 9.15% | 9.03% | 8.97% |

### 6.4. Testing for Risk Neutral Pricing and Additional Premiums

The hypothesis that the market call prices are derived from risk neutral pricing (and no additional premiums) can be tested by comparing the excess returns from call prices with the same from the S&P500 index. More formally, our null hypothesis is

$$H_0 : \lambda_{c,EF} - \lambda_{x,EF} = 0.$$

From Section 4, we know that under $H_0$, $\hat{\lambda}_{c,EF} - \hat{\lambda}_{x,EF} \overset{asy}{\sim} N(0, \mathrm{Var}(\hat{\lambda}_{c,EF} - \hat{\lambda}_{x,EF}))$, implying the test statistic

$$\frac{\hat{\lambda}_{c,EF} - \hat{\lambda}_{x,EF}}{\mathrm{SE}(\hat{\lambda}_{c,EF} - \hat{\lambda}_{x,EF})}.$$

PROPOSITION 12 (Variance of $\hat{\lambda}_{c,EF} - \hat{\lambda}_{x,EF}$). *Let $Y_{c,j}$ and $Y_{c,j}$ be defined as in Propositions 3 and 11, respectively. Similarly for $Z_{c,j}$, $W_{c,j}$, $Z_{x,j}$, and $W_{x,j}$. Further, define* $\mathrm{Cov}(Y_{c,j}, Y_{x,j}) = \int_{u=t_{j-1}}^{t_j} (V^X(t_{j-1}, T))(X_{t_{j-1}}\sigma)^2 \times B^2(t_{j-1}, u)du$. *Then, the variance of $\hat{\lambda}_{c,EF} - \hat{\lambda}_{x,EF}$ is given by*

$$\mathrm{Var}(\hat{\lambda}_{c,EF} - \hat{\lambda}_{x,EF}) = \frac{1}{\sum_{j=1}^{n} Z_{c,j} W_{c,j}^{-1} Z_{c,j}} + \frac{1}{\sum_{j=1}^{n} Z_{x,j} W_{x,j}^{-1} Z_{x,j}} - 2\,\frac{\sum_{j=1}^{n} Z_{v,j} W_{c,j}^{-1} \mathrm{Cov}(Y_{c,j}, Y_{x,j}) W_{x,j}^{-1} Y_{x,j}}{\left(\sum_{j=1}^{n} Z_{c,j} W_{c,j}^{-1} Z_{c,j}\right)\left(\sum_{j=1}^{n} Z_{x,j} W_{x,j}^{-1} Z_{x,j}\right)}.$$

Proof. The expression follows directly by considering the variance and covariance terms of $\hat{\lambda}^*_{c,EF} - \hat{\lambda}^*_{x,EF}$ and replacing expectations by the appropriate $\mathcal{F}_{t_{j-1}}$-measurable surrogates as shown in the proof of Proposition 3. ■

The null hypothesis of risk neutral pricing was tested over different sample sizes as given in Table 2. Because the call prices are generated under assumptions satisfying the null hypothesis (Black–Scholes valuation), the null should not be rejected by the data. Indeed, as Table 2 shows, differences in the excess returns obtained from the S&P500 calls and the index, $\hat{\gamma}_{c,EF} - \hat{\gamma}_{x,EF}$, are insignificant, and the standard errors will not reject the null hypothesis. The variance of the difference is dramatically reduced by the covariance term as the sample size increases. These results show that EF estimation reveals no pricing errors and premiums in call prices when none should exist.

### 6.5. Effect of Strike Level or “Moneyness”

The results so far were obtained using strikes “at the money.” The impact on the estimation of changing the strike level (“moneyness”) was also investi-

**Table 2.** Differences in EF estimates of $\lambda$ and $\gamma$ from calls and index

| | Sample Size (Trading Days) | | | | | |
|---|---|---|---|---|---|---|
| Estimate | 100 | 500 | 1,000 | 2,000 | 5,000 | 10,000 |
| $\hat{\gamma}_{c,EF} - \hat{\gamma}_{x,EF}$ | $-.120 \times 10^{-2}$ | $.146 \times 10^{-3}$ | $-.731 \times 10^{-5}$ | $.301 \times 10^{-6}$ | $.266 \times 10^{-7}$ | $.209 \times 10^{-8}$ |
| $\mathrm{Var}(\hat{\gamma}_{c,EF} - \hat{\gamma}_{x,EF})$ | 1.893 | .2308 | $.720 \times 10^{-1}$ | $.168 \times 10^{-1}$ | $.938 \times 10^{-3}$ | $.319 \times 10^{-6}$ |
| $\mathrm{SE}(\hat{\gamma}_{c,EF} - \hat{\gamma}_{x,EF})$ | 1.323 | .4510 | .2478 | .1164 | .02705 | .004945 |
| $\hat{\lambda}_{c,EF} - \hat{\lambda}_{x,EF}$ | $-.240 \times 10^{-1}\%$ | $.292 \times 10^{-4}\%$ | $.146 \times 10^{-4}\%$ | $.601 \times 10^{-5}\%$ | $.531 \times 10^{-6}\%$ | $.418 \times 10^{-6}\%$ |

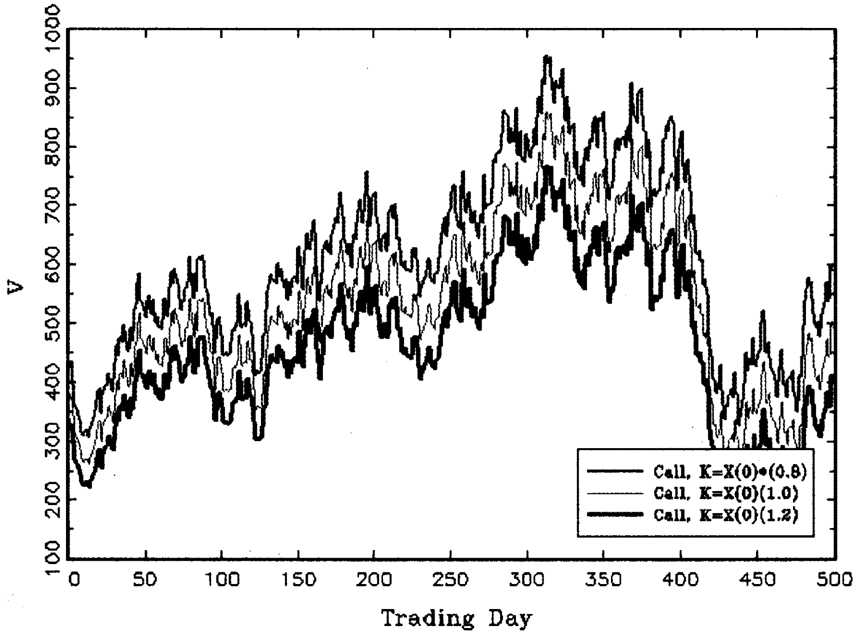


**FIGURE 2.** Calls of different strikes.

gated. Various values of strikes were considered ranging from 70% of the initial index value to 130%. The sample size was set at 5,000, and the results are consistent at other sample sizes. The results demonstrate that the strikes have absolutely no impact on the estimation of the market price of risk: $\hat{\gamma}_{c,EF} = \hat{\gamma}_{x,EF} = .4517$ for the strikes $K = X_o k$, $k = .7, .8, .9, 1, 1.1, 1.2$, and $1.3$. The reason for this is that movements in call value and "delta," $V_k^X(t_{j-1}, T_g)$, move in a parallel fashion across different strikes (see Figure 2). This also shows graphically the strong dependence between strike replicates evidenced in the excess returns estimator of Proposition 4.

### 6.6. Effect of Smaller Maturity Sampling Cycles

In the market setting, the vast majority of traded calls are of maturities less than 1 year. Therefore, it is not feasible to increase the sample size by extending the time to maturity. An alternative sampling design that cycles over calls of smaller (nonoverlapping) maturities is considered in this section as a way to reduce the variances of estimators. Performance of the EF estimators at an effective sample size of 500 and 2,000 trading days is compared by varying the size of the maturity cycles. At an effective sample size $(pn)$ of 500, maturity cycles $(p)$ 1, 10, and 31 are paired with maturity lengths $(n)$ 500, 50, and 15, respectively. And at the effective sample size $(pn)$ of 2,000, maturity cycles $(p)$ 1, 40, and 125 are paired with maturity lengths $(n)$ 2,000, 50, and 16, respectively. The maturity of $n = 16$ corresponds to a call expiring in 1 month (20 trading days) with the last 4 observations (20%) ignored for reasons of numerical stability in the estimation. As time to maturity approaches, the value and delta of the call become very sensitive to time and change very sharply—causing rapidly changing weighting factors in the EF estimator for these observations. For example, at a sample size of 100, estimates corresponding to dropping the last 1%, 10%, 20%, and 30% of observations are .563, .461, .458, and .476 for $\hat{\gamma}_{c,EF}$ and .462, .451, .455, and .473 for $\hat{\gamma}_{x,EF}$, respectively. The differences $\hat{\gamma}_{c,EF} - \hat{\gamma}_{x,EF}$ are .102, .010, .003, and .003 at these drop rates.

The 20% drop rate is chosen because we wish to consider estimation also at the 1 month maturity. This gives around 20 trading days with the last 4 dropped (20%) for the reasons cited previously. Dropping only the last 2 or 3 observations still caused numerical problems for this shorter maturity. Therefore, to be consistent, the drop rate was fixed at a constant 20% for all maturities. Similarly, a maturity of $n = 50$ corresponds to a call expiring in 3 months (62 trading days) with the last 12 observations (20%) left out. The results of this analysis are reported in Table 3.

It is found that sampling from cycles of shorter maturities with the same effective sample size (cycles times maturity length) yields similar results for estimates of excess returns (and market risk) and their standard errors as a single maturity sample of large duration. There is no observed deterioration in moving to the smaller cycles. This result gives confidence in the applicability of the estimation methodology to market derivative prices where larger samples derived from sampling over multiple nonoverlapping shorter maturities can be used for variance reduction.

## 7. CONCLUSIONS AND FURTHER WORK

This paper develops an econometric framework for (i) estimating the underlying security's excess returns from derivative prices, (ii) testing for risk neutral pricing, and (iii) measuring premiums outside the no-arbitrage pricing model. The estimator is derived by applying quasi-likelihood and Feynman–Kac theory to the risk neutral contingent claims pricing model to generate the optimal orthogonality restriction for the parameter of interest (excess returns). A diagnostic study is undertaken to resolve sample design issues such as impact of the strike level, strike replication, and shorter maturity cycles on estimation of excess returns.

Quasi-likelihood estimators for excess returns (and their variance) are obtained for both the derivative price process and the underlying asset price process (e.g., stock, bond, or index). The strong consistency and asymptotic

**TABLE 3.** Effect of smaller maturity cycles on estimation

| Estimator | Average | Sample Size (Trading Days) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 500 | | | | | | 2,000 | | | | | |
| | | $n$ | $p$ | $n$ | $p$ | $n$ | $p$ | $n$ | $p$ | $n$ | $p$ | $n$ | $p$ |
| | | 500 | 1 | 50 | 10 | 16 | 31 | 2,000 | 1 | 50 | 40 | 16 | 125 |
| EF-call | $\hat{\lambda}_{c,EF}$ | .4639 | | .4633 | | .4434 | | .4578 | | .4519 | | .4522 | |
| | $\mathrm{Var}(\hat{\gamma}_{c,EF})$ | .5200 | | .5200 | | .5242 | | .1300 | | .1300 | | .1300 | |
| | $\mathrm{SE}(\hat{\gamma}_{c,EF})$ | .7211 | | .7211 | | .7240 | | .3606 | | .3606 | | .3606 | |
| | $\hat{\lambda}_{c,EF}$ | 9.28% | | 9.27% | | 8.87% | | 9.16% | | 9.04% | | 9.04% | |
| EF-index | $\hat{\gamma}_{x,EF}$ | .4637 | | .4640 | | .4450 | | .4577 | | .4523 | | .4510 | |
| | $\mathrm{Var}(\hat{\gamma}_{x,EF})$ | .5200 | | .5200 | | .5242 | | .1300 | | .1300 | | .1300 | |
| | $\mathrm{SE}(\hat{\gamma}_{x,EF})$ | .7211 | | .7211 | | .7240 | | .3606 | | .3606 | | .3606 | |
| | $\hat{\lambda}_{x,EF}$ | 9.27% | | 9.28% | | 8.40% | | 9.15% | | 9.05% | | 9.02% | |

normality of the estimator are established in the context of a nonstationary underlying state process. These asymptotic properties are derived under a milder conditional second moment assumption that is satisfied by a large class of predictable stochastic processes with finite conditional second moments (or finite variation). Therefore, the proposed estimator is robust to distributional assumptions of risk neutral martingale theory where the stochastics are driven by Brownian motion. However, when the underlying state process is close to being a Brownian motion, the quasi-likelihood estimator offers optimal and efficient estimation. Moreover, the final feasible estimator developed is a discretized version of the estimator implied by the continuous risk neutral pricing framework. The convergence results show that in addition to possessing the robustness property, the feasible estimator offers consistent and efficient estimation.

Much of the estimation literature on stochastic processes in finance has focused on estimation of parameters (e.g., drift, volatility) from the state process $X$ (e.g., index, stock, bond). This paper considers the estimation of the excess return parameter from the derivative process $V(X)$ overlying the state process (and also $X$). The estimator is obtained by applying quasi-likelihood and Feynman–Kac theory to the risk neutral contingent claims pricing model to generate the optimal orthogonality restriction. Nonequivalence between excess returns estimated from derivative and underlying asset prices implies departures from the risk neutral pricing model and presence of additional premiums.

This paper also considers a mixed estimation framework where the estimating equations follow from the continuous risk neutral pricing model for contingent claims but where sampling of market derivative prices (and underlying asset prices) occurs at discrete, perhaps random, times. This paper differs from the direction taken in other work both in focus of estimation and the estimation methodology used. The existing literature has not dealt with the estimation of excess returns from derivative prices $V(X)$. To construct an efficient estimator, the proposed methodology first identifies a conditional martingale difference equation (CMDE) by constructing an Itô expansion of the discounted derivative process between two given sampling intervals under the risk neutral measure, then applies the Feynman–Kac result to reduce terms, and last introduces the parameter of interest (excess returns) by switching to the empirical measure. Once the CMDE is constructed, the optimal orthogonality restriction on the CMDE is obtained from quasi-likelihood theory. A discrete "feasible" estimator is next developed from this procedure in which all quantities are measurable with respect to information available at the beginning of each sampling period.

The estimation also has interesting empirical derivative pricing applications that are being explored. Market prices of risk can be readily constructed from estimates of volatility, and excess returns in the derivative market and contingent claims can then be priced using the risk neutral density from Girsanov's change of measure formula.

There are a number of further implications of this econometric framework for derivatives. The proposed approach is applicable to any arbitrary derivative security, does not require estimation of the risk neutral probability measure, inherits the optimal efficiency properties of EF theory, and has application to spot rate bond pricing models where it offers linear estimation of parameters in highly nonlinear bond formulae.

A diagnostic study based on generating the S&P500 index and calls verifies the ability of the proposed method to correctly estimate excess returns from derivative prices and test for risk neutral pricing—even at sample sizes as low as 100 observations. Sample design issues are also resolved: (i) the estimator is invariant to call strikes; (ii) strike replicates do not reduce variance because of dependence; and (iii) larger samples constructed by cycling over shorter maturity options can be used to reduce its variance. On the last point, it is found that the 3 month maturity is best, yielding more stable estimation of excess returns than 1 month calls. These results give confidence in applicability to market derivative prices. Current ongoing work seeks to apply the proposed estimation framework to S&P500 call options data from the Chicago Board of Options Exchange, and results from this application will be reported in the sequel to this paper. The empirical pricing implications of this estimation for derivatives are also being investigated.

## *NOTE*

1. An anonymous referee made the keen observation that the bound of Proposition 6 is infinity if $X_{t_j}$ approaches 0. In this case, $E(X_{t_j}^2|\mathcal{F}_{t_{j-1}})$ in the numerator is roughly $t_j - t_{j-1}$ (a constant $C$), and the bound depends on the expectation $E[\sum_{j=1}^{n} 1/X_{t_{j-1}}^2] = \infty$. This situation can be excluded by considering the economic rationale implied by this event. In a financial market context, $X$ is typically the price of a stock or market index, and $V(X)$ is the value of the derivative security (e.g., call, put) defined on the underlying asset. There are two arguments that ensure that this "pathological case" cannot really occur in financial markets: (i) the share price of a firm never hits zero—even in the case of bankruptcy; and (ii) the derivative value is zero if the underlying asset approaches zero: $\lim_{X\to 0} V(X) = 0$.

## *REFERENCES*

Ait-Sahalia, Y. & A.W. Lo (1998) Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance* 53, 499–547.

Artzner, P. & F. Delbaen (1987) Term structure of interest rates: The martingale approach. *Advances in Applied Mathematics* 11, 270–302.

Banz, R. & M. Miller (1978) Prices for state-contingent claims: Some estimates and applications. *Journal of Business* 51, 653–672.

Bekaert, G., R.J. Hodrick, & D.A. Marshall (1997) On biases in tests of the expectations hypothesis of the term structure of interest rates. *Journal of Financial Economics* 44, 309–348.

Billingsley, P. (1986) *Probability and Measure*. New York: Wiley and Sons.

Breedan, D. & R.H. Litzenberger (1978) Prices of state-contingent claims implicit in option prices. *Journal of Business* 51, 621–651.

Brennan, M.J. & E.S. Schwartz (1979) A continuous-time approach to the pricing of bonds. *Journal of Banking and Finance* 3, 135–155.

Broze, L., D. Scaillet, & J. Zakoian (1997) Testing for continuous-time models of short-term interest rates. *Journal of Empirical Finance* 2, 199–223.

Campbell, J., A.W. Lo, & A.C. MacKinlay (1997) *The Econometrics of Financial Markets*. Princeton, New Jersey: Princeton University Press.

Chernov, M. & E. Ghysels (1998) What Data Should Be Used to Price Options. Working paper, Department of Finance, Pennsylvania State University.

Dohal, G. (1987) On estimating the diffusion coefficient. *Journal of Applied Probability* 24, 105–114.

Duffie, D. (1992) *Dynamic Asset Pricing Theory*. Princeton, New Jersey: Princeton University Press.

Duffie, D. & P. Glynn (1998) Estimation of Continuous-Time Markov Processes Sampled at Random Time Intervals. Working paper, Graduate School of Business, Stanford University.

Florens-Zmirou, D. (1993) On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability* 30, 790–804.

Godambe, V.P. (1960) An optimal property of regular maximum likelihood estimation. *Annals of Mathematical Statistics* 31, 1208–1211.

Godambe, V.P. & C.C. Heyde (1987) Quasi-likelihood and optimal estimation. *International Statistical Review* 55, 231–244.

Godambe, V.P. & B.K. Kale (1991) Estimating functions: An overview. In V.P. Godambe (ed.), *Estimating Functions*, pp. 3–20. Oxford: Oxford Science Publications.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Application*. New York: Academic Press.

Hansen, L.P. (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.

Hansen, L.P. & J.A. Scheinkman (1995) Back to the future: Moment generating implications for continuous-time Markov processes. *Econometrica* 63, 767–804.

Harrison, J.M. & D.M. Kreps (1979) Martingales and arbitrage in multiperiod security markets. *Journal of Economic Theory* 20, 381–408.

Harrison, J.M. & S.R. Pliska (1981) Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications* 11, 215–260.

Heath, D., R.A. Jarrow, & A. Morton (1992) Bond pricing and the term-structure of interest rates: A new methodology for contingent claims valuation. *Econometrica* 60 (1), 77–105.

Heyde, C.C. (1989) Quasi-likelihood and optimality of estimating functions: Some current unifying themes. *Bulletin of International Statistics* 1, 19–29.

Ho, T.S. & S. Lee (1986) Term structure movements and pricing interest rate contingent claims. *Journal of Finance* 41, 1011–1028.

Karatzas, I. & S.E. Shreve (1991) *Brownian Motion and Stochastic Calculus*. New York: Springer-Verlag.

Lo, A.W. (1988) Maximum likelihood estimation of generalized Itô processes with discretely sampled data. *Econometric Theory* 4, 231–247.

Longstaff, F. (1991) An Empirical Examination of the Risk-Neutral Valuation Model. Working paper, Ohio State University.

Pandher, G.S. (2000) Volatility Estimation from High Frequency Derivative Prices. Working paper, Kellstadt Graduate School of Management, DePaul University.

Pederson, A. (1995) A new approach to MLE for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics* 22, 55–71.

Protter, P. (1990) *Stochastic Integration and Differential Equations*. New York: Springer-Verlag.

Thavaneswaran, A. & M.E. Thompson (1986) Optimal estimation for semimartingales. *Journal of Applied Probability* 23, 409–417.

Sherrick, B.J., S.H. Irwin, & D.L. Forster (1990) An Examination of Option-Implied S&P 500 Futures Price Distributions. Working paper, Ohio State University.

Vasicek, O. (1977) An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177–188.

# APPENDIX A: PROOF OF PROPOSITION 1 (ESTIMATING FUNCTION $h_j(\lambda)$ [SINGLE STRIKE AND MATURITY])

The proof is pursued in the general setting ($d > 1$). Applying a "chain-rule" version of Itô's formula to the functional $f(V_s, B_s) = V(s, X_s)B(t,s)$ with $t, s \in [0, T]$, and $s > t$ yields (Protter, 1990)

$$V(s, X_s)B(t,s) - V(t, X_t) = \int_{u=t}^{s} \frac{\partial f}{\partial B}(X_u, B_u)dB_u + \int_{u=t}^{s} \frac{\partial f}{\partial V}(X_u, B_u)dV_u. \tag{A.1}$$

In the preceding expansion all second derivatives of the functional $f(X_s, B_s)$ are zero, and hence the corresponding quadratic variation terms drop out. Next substitute the following expressions into the right hand side of (A.1): $dV_s = (\partial V/\partial s)ds + A_s V ds + (\partial V(s,T)/\partial X)'\sigma(s, X_s)d\widetilde{W}_s$ and $dB_s = -B_s r_s ds$. This leads to

$$\begin{aligned}
V(s, X_s)&B(t,s) - V(t, X_t) \\
&= \int_{u=t}^{s} V_u B_u r_u \, du \\
&\quad + \int_{u=t}^{s} \left[ \frac{\partial V}{\partial u} + A_u V \right] B_u \, du + \int_{u=t}^{s} B_u \left( \frac{\partial V(u,T)}{\partial X} \right)' \sigma(u, X_u) d\widetilde{W}_u \\
&= \int_{u=t}^{s} \left[ \frac{\partial V}{\partial u} + A_u V - r_u V \right] B_u \, du + \widetilde{M}(t,s) \\
&= -\int_{u=t}^{s} g_u B_u \, du + \widetilde{M}(t,s),
\end{aligned} \tag{A.2}$$

where $\widetilde{M}(t,s) \equiv \int_{u=t}^{s} B_u(\partial V(u,T)/\partial X)'\sigma(u, X_u)d\widetilde{W}_u$ is a $Q$ stochastic integral.

To introduce the excess return parameter (and market price of risk), we reverse the transformation in Brownian motion defined in (2.3) to obtain

$$\begin{aligned}
\widetilde{M}(t,s) &= \int_{u=t}^{s} B_u \left( \frac{\partial V(u,T)}{\partial X} \right)' [b(u, X_u) - r_u X_u] du \\
&\quad + \int_{u=t}^{s} B_u \left( \frac{\partial V(u,T)}{\partial X} \right)' \sigma(u, X_u) dW_u,
\end{aligned} \tag{A.3}$$

where $W_t$ is a $d$-dimensional Brownian motion with respect to the empirical probability measure $P$.

Defining $M(t,s) \equiv \int_{u=t}^{s} B_u(\partial V(u,T)/\partial X)'\sigma(u, X_u)dW_u$ and combining (A.2) and (A.3) yields

$$M(t,s) = V(s,X_s)B(t,s) - V(t,X_t) + \int_{u=t}^{s} g_u B_u \, du$$
$$- \int_{u=t}^{s} B_u \left( \frac{\partial V(u,T)}{\partial X} \right)' [b(u,X_u) - r_u X_u] du. \qquad \textbf{(A.4)}$$

Note that $E(M(t,s)|\mathcal{F}_t) = 0$; hence it is a martingale difference function. Therefore, its "data side" defines

$$h_j(\lambda) \equiv M(t_{j-1},t_j) = V(t_j,T)B(t_{j-1},t_j) - V(t_{j-1},T) + \int_{u=t}^{s} g_u B_u \, du$$
$$- \int_{u=t}^{s} B_u \left( \frac{\partial V(u,T)}{\partial X} \right)' [b(u,X_u) - r_u X_u] du, \qquad j = 1,\ldots,n. \qquad \textbf{(A.5)}$$

Under the assumption of geometric Brownian motion, $b(u,X_u) = (b_1 X_u^1,\ldots,b_d X_u^d)'$ and $\sigma(u,X_u) = \sigma X_u$ where $\sigma$ is a $d \times d$ diagonal matrix of volatilities. Then, $h_j(\lambda)$ and $M(t_{j-1},t_j)$ can be written as

$$h_j(\lambda) = V(t_j,T)B(t_{j-1},t_j) - V(t_{j-1},T) + \int_{u=t}^{s} g_u B_u \, du$$
$$- \int_{u=t}^{s} B_u \left( \frac{\partial V(u,T)}{\partial X} \right)' X_u^* [b_u - r_u] du, \qquad j = 1,\ldots,n, \qquad \textbf{(A.6)}$$

with the "error side"

$$M(t_{j-1},t_j) = \int_{u=t_{j-1}}^{t_j} B_u \left( \frac{\partial V(u,T)}{\partial X} \right)' \sigma X_u^* \, dW_u, \qquad j = 1,\ldots,n, \qquad \textbf{(A.7)}$$

where $X_u^*$ is a diagonal matrix formed from the vector $X_u$. Further setting $d = 1$ and using the definition $\lambda = b - r$ in equation (A.6) leads to the final equation stated in Proposition 1. ■

# APPENDIX B: PROOF OF PROPOSITION 6 (A BOUND FOR THE SEQUENCE $\{\hat{\lambda}_n\}$ IN $L_2$ NORM)

Note that $Y_j = h_j + \lambda \tilde{Z}_j = h_j + \lambda Z_j + \lambda(\tilde{Z}_j - Z_j)$. This allows $\hat{\lambda}_n$ to be written as

$$\hat{\lambda}_n - \lambda = \lambda \frac{\sum_{j=1}^{n} Z_j W_j^{-1}(\tilde{Z}_j - Z_j)}{\sum_{j=1}^{n} Z_j W_j^{-1} Z_j} + \frac{\sum_{j=1}^{n} Z_j W_j^{-1} h_j}{\sum_{j=1}^{n} Z_j W_j^{-1} Z_j} \qquad \textbf{(B.1)}$$
$$= \lambda A_n + B_n.$$

Here, $\tilde{Z}_j - Z_j = \int_{u=t_{j-1}}^{t_j} (V^X(u,T)X_u - V^X(t_{j-1},T)X_{t_{j-1}})B(t_{j-1},u)du$ where for $u > t_{j-1}$ the following bound holds:

$$\begin{aligned}
V^X(u,T)X_u - V^X(t_{j-1},T)X_{t_{j-1}} \\
&= (V^X(u,T)X_u - V^X(u,T)X_{t_{j-1}}) + (V^X(u,T)X_{t_{j-1}} - V^X(t_{j-1},T)X_{t_{j-1}}) \\
&= V^X(u,T)(X_u - X_{t_{j-1}}) + (V^X(u,T) - V^X(t_{j-1},T))X_{t_{j-1}} \\
&\le X_u, \qquad \textbf{(B.2)}
\end{aligned}$$

because $V^X(u,T) \le 1$. This follows directly from the fact that derivative option prices are bounded above by the value of the underlying asset: $V(u, X_u; K, T) \le X_u$. Therefore, using the fact that $X_u^2 (t_{j-1} < u \le t_j)$ is an increasing process, we can write

$$\begin{aligned}
E((\tilde{Z}_j - Z_j)^2 \,|\, \mathcal{F}_{t_{j-1}}) &\le E\left[\left(\int_{u=t_{j-1}}^{t_j} X_u B(t_{j-1},u)du\right)^2 \middle| \mathcal{F}_{t_{j-1}}\right] \\
&\le E\left[X_{t_j}^2 \left(\int_{u=t_{j-1}}^{t_j} B(t_{j-1},u)du\right)^2 \middle| \mathcal{F}_{t_{j-1}}\right] \\
&= W_j \, \frac{E(X_{t_j}^2 | \mathcal{F}_{t_{j-1}})}{(X_{t_{j-1}} \sigma V^X(t_{j-1},T))^2} \, \frac{\left[\int_{u=t_{j-1}}^{t_j} B_u \, du\right]^2}{\left[\int_{u=t_{j-1}}^{t_j} B_u^2 \, du\right]}. \qquad \textbf{(B.3)}
\end{aligned}$$

The conditional second moment bound (B.3) further gives

$$\begin{aligned}
E(A_n^2) &\le E\left(\frac{\sum_{j=1}^n Z_j W_j^{-1} E((\tilde{Z}_j - Z_j)^2 \,|\, \mathcal{F}_{t_{j-1}}) W_j^{-1} Z_j}{\left(\sum_{j=1}^n Z_j W_j^{-1} Z_j\right)^2}\right) \\
&\le E\left(\frac{\sum_{j=1}^n Z_j W_j^{-1} Z_j}{\left(\sum_{j=1}^n Z_j W_j^{-1} Z_j\right)^2} \, \frac{\left[\int_{u=t_{j-1}}^{t_j} B_u \, du\right]^2}{\left[\int_{u=t_{j-1}}^{t_j} B_u^2 \, du\right]} \, \frac{E(X_{t_j}^2 | \mathcal{F}_{t_{j-1}})}{(X_{t_{j-1}} \sigma V^X(t_{j-1},T))^2}\right) \\
&= \left(\frac{r(1 - \exp\{-2r\Delta\})}{2(1 - \exp\{-r\Delta\})^2}\right) \frac{\left[\int_{u=t_{j-1}}^{t_j} B_u \, du\right]^2}{\left[\int_{u=t_{j-1}}^{t_j} B_u^2 \, du\right]} \, \frac{1}{n^2} \, E\left(\sum_{j=1}^n \frac{E(X_{t_j}^2 | \mathcal{F}_{t_{j-1}})}{(X_{t_{j-1}} \sigma V^X(t_{j-1},T))^2}\right) \\
&\le \frac{k_1}{n^2} \, E\left(\sum_{j=1}^n \frac{E(X_{t_j}^2 | \mathcal{F}_{t_{j-1}})}{X_{t_{j-1}}^2}\right) \qquad \textbf{(B.4)}
\end{aligned}$$

for some positive constant $k_1$ where the third equality follows from $Z_j W_j^{-1} Z_j = 2(1-\exp\{-r\Delta\})^2/r(1-\exp\{-2r\Delta\})$.

Next consider the second term of (B.1). From the proof of Proposition 1, the "error side" of $h_j$ is given by the stochastic integral $h_j = \int_{u=t_{j-1}}^{t_j} V^X(u,T) X_u \sigma B(t_{j-1},u)\, dW_u$. Using Proposition 2, it can be bounded in $L_2$ norm as follows:

$$
\begin{aligned}
E(h_j^2|\mathcal{F}_{t_{j-1}}) &= \int_{u=t_{j-1}}^{t_j} E(V_X^2(u,X_u)X_u^2\sigma^2|\mathcal{F}_{t_{j-1}})B_u^2\, du \\
&\le \int_{u=t_{j-1}}^{t_j} E(X_{t_j}^2|\mathcal{F}_{t_{j-1}})\sigma^2 B^2(t_{j-1},u)\, du \\
&= W_j \frac{E(X_{t_j}^2|\mathcal{F}_{t_{j-1}})}{(X_{t_{j-1}}V^X(t_{j-1},T))^2}. \qquad \textbf{(B.5)}
\end{aligned}
$$

Using (B.5) and repeating the steps of (B.4) yields

$$
\begin{aligned}
E(B_n^2) &\le E\left(\frac{\sum_{j=1}^n Z_j W_j^{-1} E(h_j^2|\mathcal{F}_{t_{j-1}}) W_j^{-1} Z_j}{\left(\sum_{j=1}^n Z_j W_j^{-1} Z_j\right)^2}\right) \\
&\le \frac{k_2}{n^2} E\left(\sum_{j=1}^n \frac{E(X_{t_j}^2|\mathcal{F}_{t_{j-1}})}{X_{t_{j-1}}^2}\right) \qquad \textbf{(B.6)}
\end{aligned}
$$

for some positive constant $k_2$.

Note that $E(h_j(\tilde{Z}_j - Z_j)|\mathcal{F}_{t_{j-1}}) = 0$ as a result of the Brownian motion in $h_j$. Therefore, the cross-moment of (B.1) is zero. Finally, we have the desired bound for $\hat{\lambda}_n$ in $L_2$ norm:

$$
\begin{aligned}
E(\lambda_n^2) &\le [\lambda^2 E(A_n^2) + E(B_n^2)] \\
&\le \frac{(\lambda^2 k_1 + k_2)}{n^2} E\left(\sum_{j=1}^n \frac{E(X_{t_j}^2|\mathcal{F}_{t_{j-1}})}{X_{t_{j-1}}^2}\right). \qquad \textbf{(B.7)}
\end{aligned}
$$

■