Consistent Estimation of the Risk-Return Tradeoff in the Presence of Measurement Error*

Anisha Ghosh† Oliver Linton‡

this version: November 12, 2013

Abstract

The empirical risk-return relation is typically measured by a least squares regression of an estimate of the conditional mean return of the stock market on an estimate of its conditional variance. We offer a solution to the measurement error problem that arises in this setting when high-frequency returns are used to construct a nonparametric proxy for the latent conditional variance. In particular, we propose a bias-correction to the standard GMM estimator derived under a double asymptotic framework wherein the number of high-frequency intra-period returns, \( N \), as well as the number of low frequency time periods, \( T \), simultaneously go to infinity. Simulation exercises show that the bias-correction is particularly relevant for small values of \( N \) which is the case in empirically realistic scenarios. Applying our methodology, we find a statistically significantly negative unconditional correlation between the conditional mean and variance over the last four decades, while the conditional correlation (conditional on the lagged mean and variance) is positive albeit not statistically significant.

Keywords: Bias-Correction, Nonparametric Volatility, Return, Risk.

JEL classification: C14, G12

---

*We thank Jushan Bai, George M. Constantinides, Christian Julliard, Sydney Ludvigson, and Antonio Mele for helpful comments on an earlier draft of this paper. We remain responsible for errors and omissions. This research was supported by the ESRC and the Leverhulme foundation.

†Department of Finance, Tepper School of Business, Carnegie Mellon University; 5000 Forbes Avenue, Pittsburgh, PA 15213, e-mail: anishagh@andrew.cmu.edu

‡Department of Economics, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DD, United Kingdom, e-mail: obl20@cam.ac.uk
1 Introduction

The relation between the expected excess return on the aggregate stock market - the so called "equity risk premium" - and its conditional variance has long been the subject of both theoretical and empirical research in financial economics. The risk-return relation is an important ingredient in optimal portfolio choice, and is central to the development of theoretical asset-pricing models aimed at explaining a host of observed stock market patterns.

On the theoretical side, asset pricing models generally predict a positive relationship between the risk premium on the market portfolio and the variance of its return. Prominent examples include the classic capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965), as well as the more recent external habit model of Campbell and Cochrane (1999) and the Long Run Risks model of Bansal and Yaron (2004). While the CAPM implies that the equity risk premium is proportional to its conditional variance, i.e. a constant price of risk, the external habit model is characterized by a time varying price of risk generated by time varying risk aversion of investors in the economy. While the above models imply a positive risk-return tradeoff, a negative risk-return relation is not inconsistent with equilibrium (see, e.g., Abel (1988), Backus and Gregory (1993), and Whitelaw (2000)).


Some studies have relied on parametric and semi-parametric ARCH or stochastic volatility models that impose a high degree of structure on the return generating process, about which there is little direct empirical evidence. The results have been found to be very sensitive to the particular model specification. Other studies have typically measured the conditional expectations underlying the
conditional mean and conditional variance as projections onto predetermined conditioning variables. Practical constraints, such as choosing among a few conditioning variables, introduce an element of arbitrariness into the econometric modeling of expectations and can lead to omitted information estimation bias. Also, as pointed out by Hansen and Richard (1987), if investors have information not reflected in the chosen conditioning variables used to model market expectations, measures of the conditional mean and conditional variance will be misspecified and possibly highly misleading. Lettau and Ludvigson (2010) distinguish between the unconditional correlation and conditional correlation (conditional on lagged mean and variance) between the expected excess market return and its conditional variance. Using the dynamic factor analysis methodology of Ludvigson and Ng (2007) on a set of 172 financial indicators, they find a strong statistically significant positive conditional correlation while the unconditional correlation is negative and statistically insignificant. They also conclude that the risk-return tradeoff results depend critically on the conditioning information used to construct the conditional expected market return. Most of the literature also ignores the measurement error that arises in this setting, as a result of using proxies for the latent conditional variance is estimating the risk-return relation via a least squares regression of the estimate of the conditional mean on the proxy for the conditional variance.

In this paper, we propose an approach to estimating the risk-return relation that overcomes some of the limitations of existing empirical analyses. First, we focus on a nonparametric measure of the ex post return variability over a finite time interval, namely integrated variance, that is unbiased for the conditional variance and is void of any specific functional form assumptions about the stochastic process generating returns. This property of the integrated variance enables us to express the risk-return relation in terms of a conditional moment restriction involving the realized excess returns, the integrated variance, and other predetermined variables that prior research has shown to be important in uncovering the risk-return relation. Although the integrated variance is latent, it may be consistently estimated using the realized variance that is computed as the sum of squares of high-frequency intra-period returns. This gives feasible moment restrictions and we then estimate the parameters of the risk-return relation using the Generalized Method of Moments (GMM) approach. This approach, while being robust to potential misspecification in the assumed dynamics of the conditional moments, also overcomes the endogeneity problem inherent in a least squares regression of an estimate of the conditional mean on the estimate of the conditional variance.\footnote{For an excellent survey of this extensive literature, see Andersen, Bollerslev, and Diebold (2002). See also Barndorff-Nielsen and Shepherd (2002) and Andersen, Bollerslev, Diebold, and Labys (2003).}

Second, and more importantly, we offer a solution to the measurement error problem that arises \footnote{The concept of proxying the latent integrated variance with realized variance was used earlier in French, Schwert, and Stambaugh (1987) and Bandi and Perron (2008).}
because of the use of realized variance as a proxy for the latent integrated variance. Our asymptotic framework requires $N \to \infty$ and $T \to \infty$, where $N$ denotes the number of high-frequency intra-period returns used to compute the realized variance in every period, and $T$ denotes the number of low-frequency time-periods used in the GMM estimation. We derive the limiting distribution of the estimated coefficients under this double asymptotic framework.\(^3\) We find that under fairly strong conditions on $N$ and $T$, the estimates are $\sqrt{T}$-consistent and have the standard distribution as when there is no measurement-error. However, if the above condition is not satisfied, there is an asymptotic bias that would invalidate this approximation. In that case, we find that under weaker conditions on $N$ and $T$, a bias-corrected estimator has the standard limiting distribution. This improvement is particularly relevant in the empirical case we examine where $N$ is quite modest (e.g., daily returns within a month or quarter).

The above is an important methodological contribution to the extant literature on high-frequency volatility estimation. Most work has currently been about just estimating that quantity itself and using it to compare discrete time models in settings where the noise is small. Our approach is concerned with small sample issues when using estimated realized volatility as regressors in the estimation of parameters associated with the unobserved quadratic variation. This involves a useful extension of the existing asymptotic results for realized volatility\(^4\) concerned with the uniformity of the estimation error. We establish the properties of the parameter estimates and propose a bias correction in the case where the estimation error is large.

In the empirical analysis, we focus on the risk-return relation at the monthly and quarterly frequencies. We use $(N)$ daily returns of the CRSP value-weighted stock market index to obtain monthly and quarterly estimates of the realized variance. We then estimate the parameters of the risk-return relation using the GMM approach with $T$ (monthly and quarterly, respectively) observations on the realized excess market returns and realized variance. We find a negative relation between the mean and the variance that is statistically insignificant over the entire available historical sample 1927 – 2010 but is strongly statistically significant over the latter half of the sample period. Moreover, we find that the bias-correction that we propose is instrumental in delivering the strongly statistically significant results. This finding is robust to the choice of instruments. Upon inclusion of the lagged integrated variance and the lagged market return as additional right hand side variables

\(^3\)Corradi and Distaso (2006) use realized variance estimators to test for the correct specification of the functional form of the volatility process within the class of eigenfunction stochastic volatility models. The procedure is based on the comparison of the moments of realized volatility measures with the corresponding ones of integrated volatility implied by the model under the null hypothesis. They allow for measurement error in the realized variance and consider an asymptotic framework similar to ours.

\(^4\)See Barndorff-Nielsen and Shepherd (2002).
in the specification of the moment restriction, we obtain a positive, albeit statistically insignificant, relation between the conditional mean and variance of the market return.

The remainder of the paper is organized as follows. The econometric framework and estimation methodology are described in Section 2. Section 3 derives the asymptotic properties of the GMM estimator in the presence of measurement error. In Section 4, we perform Monte-Carlo simulations to examine the finite-sample performance of the estimator. Section 5 provides a description of the data along with the empirical results. In the concluding Section 6, we discuss extensions of the approach and work in progress. The Appendix contains the proofs of our main results.

2 Estimation Methodology

2.1 Model and Hypothesis

Our econometric framework focuses on the empirical risk-return relation given by the following reduced-form equation:

\[ E(r_{m,t} - r_{f,t}|\mathcal{F}_{t-1}) = b_0 + b_1(x_{t-1}) \text{var}(r_{m,t}|\mathcal{F}_{t-1}) + b_2z_{t-1}, \]  

where \( r_{m,t} \) and \( r_{f,t} \) are the continuously compounded returns on the stock market and the risk-free rate, respectively, over \([t-1, t]\), and \( \mathcal{F}_{t-1} \) denotes all information observed at time \( t-1 \). Our empirical specification is very general and nests most of the specifications considered in the literature. For example, the risk-return tradeoff exhibits substantial time-variation with the business cycle as well as with several macroeconomic indicators (see, e.g., Harvey (2001), Lettau and Ludvigson (2010)). In order to accommodate this feature of the data, the coefficient \( b_1 \) of the conditional variance is allowed to vary over time. For example, the time variation in \( b_1 \) could be modeled as a linear function of a chosen set of variables, in which case \( b_1(x_{t-1}) = b_1^1 x_{t-1} \). Also, Guo and Whitelaw (2006) and Scruggs (1998) advocate the inclusion of a set of predetermined conditioning variables on the right hand side of the above equation in order to accurately uncover the risk-return relation. In particular, Whitelaw (1994), Brandt and Kang (2004), and Ludvigson and Ng (2007) show that it is important to include lags of the conditional mean and conditional variance as additional right hand side variables. Our empirical specification accommodates these findings by including a set of \( \mathcal{F}_{t-1}\)-measurable variables \( z_{t-1} \) on the right hand side of the above equation.

The risk-return relation in Equation (1) implies the following conditional moment restriction

\[ E[r_{m,t} - r_{f,t} - b_0 - b_1(x_{t-1}) \text{var}(r_{m,t}|\mathcal{F}_{t-1}) - b_2z_{t-1}|\mathcal{F}_{t-1}] = 0. \]

4
The parameters in the above moment restriction can be estimated using a standard GMM approach with a long time series of observations on \( r_{m,t}, r_{f,t}, z_{t-1}, x_{t-1} \), and \( \var(r_{m,t} | F_{t-1}) \). The main difficulty in the estimation process arises because of the unobservability of \( \var(r_{m,t} | F_{t-1}) \). Our strategy is to replace this quantity by a feasible approximately unbiased nonparametric estimator computed from high frequency data. We next describe the framework in which this procedure makes sense.

We suppose that we observe returns \( r_{t,j} ; j = 1, 2, \ldots, N_t \) for each \( t \), where \( t = 1, 2, \ldots, T \), in our empirical application \( t \) denotes days, while \( t \) denotes months or quarters. We shall suppose that these within-period returns are generated by the following discrete-time model

\[
  r_{t,j} = N_t^{-1} \mu_{t,j} + N_t^{-1/2} \sigma_{t,j} \eta_{t,j} ,
\]

where \( \eta_{t,j} \) and \( \eta_{t,j}^2 - 1 \) are martingale difference sequences with respect to \( F_{t-1} \). The stochastic processes \( \{ \mu_{t,j}, \sigma_{t,j} \} \) are measurable with respect to \( F_{t-1} \). The processes \( \eta_{t,j} \) are not assumed to be independent of each other, i.e., we allow for the well documented leverage and volatility feedback effects. In particular, \( \eta_{t,j} \) can affect \( \sigma_{s,t+k} \) for \( s = t, k \geq 1 \) and \( s > t, k \geq 0 \).

This framework is broadly consistent with observed returns being the discretized approximation to the continuously compounded return \( r_{t,j}^* = p_{t,j}^* - p_{t,j-1}^* \), where the true underlying efficient log-price \( p^* \) follows the continuous time diffusion

\[
  dp_t^* = \mu(\cdot)dt + \sigma(\cdot)dW_t ,
\]

for functions \( \mu(\cdot), \sigma(\cdot) \), and Brownian motion \( W \). Clearly, if \( \mu(\cdot) \equiv 0 \) and \( \sigma(\cdot) = \sigma \) (a constant), we have \( p_{t,j}^* = \sigma W_t \), so that \( r_{t,j} \) are independent and normally distributed and \( r_{t,j} = r_{t,j}^* \) so that \( \eta_{t,j} \sim N(0,1) \) and are i.i.d. More generally, one can show (under some conditions) that, with probability one, \( r_{t,j} = r_{t,j}^* + o(N_t^{-\rho}) \) for some \( \rho > 1 \) (Euler, Milstein approximations; see, e.g., Goncalves and Meddahi (2009), Mykland and Zhang (2009)). We do not assume a distribution for \( \eta_{t,j} \) in (3), so they can be heavy tailed, and so our specification is perhaps consistent with intraperiod jumps.

We next define the quadratic variation of the diffusion process, or the integrated variance, \( v_t \)

\[
  v_t \equiv \lim_{N_t \to \infty} \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t,j}^2 = \int_0^1 \sigma^2(t - 1 + s)ds .
\]

\( ^5 \)We will use the continuous time theory to justify some of our methodology. We recognize that the approximation we make here in principle could affect our results, but remark that the more complicated higher order approximations to the discrete time process may not necessarily work better in practice.
The integrated variance is an (ex-ante) unbiased estimator for the conditional variance (see, e.g., Protter (2004)), so that
\[ \text{var} (r_{m,t}|\mathcal{F}_{t-1}) \approx E[v_t|\mathcal{F}_{t-1}], \]
with strict equality if \( \mu(.) \equiv 0 \). The unbiasedness property of the integrated variance gives us the following infeasible moment restriction
\[ E \left[ r_{m,t} - r_{f,t} - b_0 - b_1 (x_{t-1}) v_t - b_2 z_{t-1} | \mathcal{F}_{t-1} \right] = 0. \] (6)
We will use this moment condition as the basis for estimation.

### 2.2 Estimation Procedures

We first estimate the integrated variance by the realized variance computed from the intra period returns
\[ \hat{v}_t = \sum_{j=1}^{N_t} r_{t,j}^2. \] (7)
The theory of quadratic variation implies that the realized variance provides a consistent nonparametric measure of the integrated variance (see, e.g., Andersen, Bollerslev, Diebold, and Labys (2003) and Barndorff-Nielsen and Shephard (2002)): \( p \lim_{N \to \infty} \hat{v}_t = v_t \), where the convergence is uniform in probability (over \( t = 1, \ldots, T \)). Also, Jacod (1994), Jacod and Protter (1998), and Barndorff-Nielsen and Shephard (2002) develop the following asymptotic distribution theory for realized variance as an estimator of the integrated variance
\[ N_t^{1/2} (\hat{v}_t - v_t) \overset{N}{\to} \sqrt{2} \left( \int_0^1 \sigma^2 (t - 1 + s) dB(t - 1 + s) \right), \]
where \( B \) is a Brownian motion independent of \( W \) in Equation (4) and the convergence is in law stable as a process. This result implies that \( N_t^{1/2} (\hat{v}_t - v_t) \Rightarrow MN(0, 2 \int_0^1 \sigma^4 (t - 1 + s) ds) \), where \( MN \) denotes a mixed Gaussian distribution. Barndorff-Nielsen and Shephard (2002) showed that the above result can be used in practice as the integrated quarticity \( IQ \equiv \int_0^1 \sigma^4 (t - 1 + s) ds \) can be consistently estimated using \( (1/3) RQ_t \), where
\[ RQ_t \equiv N_t \sum_{j=1}^{N_t} r_{t,j}^4. \]
It further follows that \( (1.5 RQ_t^{-1} N_t)^{1/2} (\hat{v}_t - v_t) \Rightarrow N(0, 1) \). This is a nonparametric result as it does not require us to specify the form of the drift, \( \mu(.) \), or the diffusion, \( \sigma(.) \), in Equation (4). The integrated quarticity plays an important role in our procedure below.

Plugging the realized variance into the infeasible moment restriction (6), we obtain the feasible moment restriction
Finally, with a set of chosen instruments, \( y_{t-1} \) (that could include, for instance, lagged variances), we have the unconditional moment restrictions:

\[
E \left[ (r_{m,t} - r_{f,t} - b_0 - b_1 (x_{t-1}) \hat{v}_t - b_2^r z_{t-1}) | \mathcal{F}_{t-1} \right] = 0. 
\]  

(8)

Defining \( y_t \equiv (y_t', v_t, v_{t-1}, \ldots, v_{t-p}) \), \( z_t \equiv (z_t', v_t, v_{t-1}, \ldots, v_{t-p}) \), where \( y_t' \) and \( z_t' \) are the observable components of \( y_t \) and \( z_t \), respectively, \( X_t \equiv (r_{m,t}, r_{f,t}, y_t', z_t', x_t) \), \( V_t = (v_t, v_{t-1}, \ldots, v_{t-p})' \), and \( \hat{V}_t = (\hat{v}_t, \hat{v}_{t-1}, \ldots, \hat{v}_{t-p})' \), we can rewrite the feasible moment restriction as:

\[
E \left[ G(X_t, \hat{V}_t; \theta_0) \right] = 0.
\]

where \( \theta = (b_0, b_1', b_2')' \) with true value \( \theta_0 \). The above set of moment restrictions are expressed entirely in terms of observable variables and, therefore, the parameter vector \( \theta \) may be estimated using the GMM approach. Specifically, we define the estimator \( \hat{\theta}_T \in \Theta \) as the minimizer of

\[
\hat{\theta}_T = \arg \min_{\theta \in \Theta} \left\| G_T(\theta) \right\|_W, \quad G_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} G(X_t, \hat{V}_t; \theta),
\]

where \( W \) is a symmetric positive definite weighting matrix, and \( ||A||_W = (\text{tr}(A^T W A))^{1/2} \).

3 Asymptotic Properties

We derive an asymptotic approximation to the properties of our estimators of \( \theta \). Our asymptotic framework has \( T \to \infty \) and \( N_t \to \infty \) for each \( t = 1, 2, \ldots, T \). The number of high-frequency, intra-period returns \( N_t \to \infty \) is required for realized variance, \( \hat{v}_t \), to accurately estimate the integrated variance, \( v_t \), while we need the number of low-frequency time periods \( T \to \infty \) for the asymptotics of the GMM estimator. Empirically, \( N_t \) is really only moderate size and so the quality of the asymptotic approximation is likely to be an issue. We show how to address this issue by providing a bias correction method that improves the approximation error.

We first present a lemma that involves a useful extension of the existing asymptotic results obtained for realized volatility in Barndorff-Nielsen and Shephard (2002). This lemma is concerned with the uniformity of the estimation error. The existing financial econometrics literature on non-parametric volatility estimation has focused on estimating financial market volatility over a finite time horizon, typically daily or monthly. In these applications, it suffices to establish consistency.
of the estimator over the finite time interval. In our present application, however, the number of finite-length time periods tends to infinity, thereby requiring a stronger consistency result. In this paper, we apply the methodology to estimate the empirical risk-return relation. However, the results are considerably general and might be useful in other contexts that require financial market variance estimation over successive time periods.

Our first result establishes the consistency of \( \hat{\nu}_t \) for \( \nu_t \), uniformly in \( t \). To derive the result, we make the following regularity assumptions.

**Assumptions A**

1. There exists a small \( \epsilon > 0 \) such that with probability one, for large enough \( T \) and some constant \( M \),

\[
\max_{1 \leq t \leq T} \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t,j}^4 \leq MT^\epsilon
\]

2. For some \( \gamma > 0 \),

\[
N = \min_{1 \leq t \leq T} N_t = O(T^{\gamma}) \leq \max_{1 \leq t \leq T} N_t = O(T^{\gamma})
\]

3. For some \( \lambda > 0 \),

\[
\max_{1 \leq t \leq T} \left[ \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t,j}^2 - p \lim_{N_t \to \infty} \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t,j}^2 \right] = O_p(\sqrt{N_t}^{-\lambda})
\]

4. The process \( \{ \sigma_{t,j}^2 (\eta_{t,j}^2 - 1) \}_{j=1}^{N_t, T} \) has finite \( k \)th moment, \( k > 3 \), and exponentially decaying \( \alpha \)-mixing coefficient: \( \alpha(k) = \exp\{-ck\} \), for some \( c > 0 \).

**Remarks.** (i) Condition (A1) controls the behaviour of the volatility process over long time spans. One possibility is to require that the process \( \sigma_{t,j}^2 \) is uniformly bounded over all \( t \) and all \( j \) and all sample paths, but this is a little strong. Instead, we shall control the rate of growth of the maximum value this process can achieve over many periods. Let \( m_t = \sum_{j=1}^{N_t} \sigma_{t,j}^4 / N_t \) denote the intraperiod second moment of volatilities. Suppose, for example, that the stochastic process \( m_t \) was stationary and Gaussian, then \( \max_{1 \leq t \leq T} m_t \) would grow to infinity at a logarithmic rate. We shall allow instead this process to grow at an algebraic rate that is much faster than logarithmic. Over the sample period 1927 – 2010, daily excess market returns are highly leptokurtic with the degree of excess kurtosis being 20.9. The evidence of very fat tails in the distribution of returns highlights the importance of this assumption.

(ii) Condition (A3) implies that the process for \( \sigma_{t,j}^2 \) is continuous, but is less strong than it being differentiable, i.e., it can be only Holder Continuous of order less than 1/2.
(iii) Condition (A4) ensures that the random variables \( \sigma_t^2, (\eta_t^2 - 1) \), although not necessarily bounded, satisfy Cramer’s conditions. This enables use of the exponential inequality for strongly-mixing time series processes (Theorem 1.4 of Bosq (1998)). Note that these represent the rescaled (by \( N_t^{-1} \)) returns. For example in the diffusion process (4) with \( \mu(.) \equiv 0 \) and \( \sigma(.) = \sigma \) (a constant), \( \eta_t \) are independent and normally distributed and \( \sigma_t = \sigma \) so this condition is automatically satisfied.

We have the following result, a formal proof of which is contained in Appendix A.1.

**Lemma 1** Suppose that Assumptions A1-A4 hold. Then, for \( \alpha < \frac{1}{2} - \varepsilon \), we have

\[
T^{\alpha} \max_{1 \leq t \leq T} (\hat{\theta}_t - \theta) = o_p(1) \quad (10)
\]

We next turn to the main result of this section - the asymptotic distribution of the parameter estimator \( \hat{\theta}_T \). We define \( G_T(\theta) \equiv \frac{1}{T} \sum_{t=1}^{T} G(X_t, V_t; \theta) \) and the infeasible GMM estimator \( \hat{\theta}_T \) that minimizes \( ||G_T(\theta)||_W \). Let \( \overline{G}(\theta) = E(G(X_t, V_t; \theta)) \) and define

\[
\Gamma \equiv \frac{\partial}{\partial \theta} \overline{G}(\theta_0), \\
\Omega \equiv \text{var} \left[ \sqrt{T}G_T(\theta_0) \right].
\]

Then, under suitable regularity conditions, the infeasible GMM estimator, \( \hat{\theta}_T \), satisfies

\[
\sqrt{T}(\hat{\theta}_T - \theta_0) = -(\Gamma^T W \Gamma)^{-1} \Gamma^T W \sqrt{T} G_T(\theta_0) + o_p(1) \overset{d}{\longrightarrow} N(0, \Sigma), \quad (11)
\]

where \( \Sigma = (\Gamma^T W \Gamma)^{-1} \Gamma^T W \Omega W \Gamma (\Gamma^T W \Gamma)^{-1} \) (see Pakes and Pollard (1989)). It is natural to suppose that the process \( \{X_t, V_t\} \) is stationary and weakly dependent, e.g., strong mixing, which would support the central limit theorem in (11). It is also a reasonable assumption in this context that \( G(X_t, V_t; \theta_0) \) is a martingale difference sequence, in which case \( \Omega \equiv \text{var} [G(X_t, V_t; \theta_0)] \).

In order to derive the asymptotic distribution of the estimator \( \hat{\theta}_T \), we make some additional assumptions. Our theory parallels the work of Pakes and Pollard (1989), so we adopt their regularity conditions:

**Assumptions B**

1. As \( T \to \infty \), \( ||G_T(\hat{\theta}_T)||_W = \inf_{\theta} ||G_T(\theta)||_W + o_p(1) / \sqrt{T} \);

2. The matrix \( \Gamma(\theta) = \frac{\partial}{\partial \theta} \overline{G}(\theta) \) is continuous in \( \theta \) and is of full (column) rank at \( \theta = \theta_0 \).
3. For all sequences of positive numbers \( \delta_T \) such that \( \delta_T \to 0 \),

\[
\sup_{\|\theta - \theta_0\| \leq \delta_T} \| G_T(\theta) - \overline{G}(\theta) \|_W = O_p(1/\sqrt{T});
\]

\[
\sup_{\|\theta - \theta_0\| \leq \delta_T} \| \sqrt{T}[G_T(\theta) - \overline{G}(\theta)] - \sqrt{T}[G_T(\theta_0) - \overline{G}(\theta_0)] \|_W = o_p(1);
\]

4. As \( T \to \infty \),

\[
\sqrt{T} G_T(\theta_0) \Rightarrow N(0, \Omega)
\]

5. The true parameter \( \theta_0 \) is in the interior of \( \Theta \).

6. For some \( \omega > 0 \),

\[
\sup_{T \geq 1} \frac{1}{T} \sum_{t=1}^{T} E |G(X_t; V_t; \theta_0)|^{2+\omega} < \infty
\]

7. The first three partial derivatives of \( G \) with respect to \( \theta \) and \( V_t \) exist and satisfy dominance conditions, namely for all vectors \( \nu \) (pertaining to \( (V_t, \theta) \)) with \( |\nu| \leq 3 \), and for some sequence \( \delta_T \to 0 \),

\[
\sup_{\|x\| \leq \delta_T} \sup_{\theta \in \Theta} \| D^\nu G(X_t; V_t + x; \theta) \| \leq U_t,
\]

where \( EU_t < \infty \).

Remarks. (i) The first condition is quite general and allows the estimator to be only an approximate minimizer of the criterion function. Condition B2 is important for identification. For example, when \( b_{1,t-1} = b_1 \) (a scalar constant) and \( b_2 = 0 \), Condition B2 holds provided the integrated variance process, \( \int_{t-1}^{t} \sigma^2(s)ds \), is not independent of the instruments used in the estimation. For instance, when lagged integrated variance is used as an instrument, this condition requires that the integrated variance process is not independent across non-overlapping time periods. Condition B3 is a technical condition that is satisfied in our case because of the linearity of the moment condition and the assumptions we made on the data in A. The central limit theorem in B4 is satisfied if \( G(X_t, V_t; \theta_0) \) is a martingale difference sequence and Assumption B6 holds (See Pakes and Pollard (1989)). Condition B7 is a smoothness condition on \( G(\cdot) \). Note that the asymptotic derivations in Pakes and Pollard (1989) do not require \( G(X_t, V_t; \theta) \) to be smooth in \( \theta \) or \( (X_t, V_t) \) but does require \( \overline{G}(\theta) \) to be smooth. However, for the purposes of our current application, it is natural to assume the function \( G \) to be smooth.
The following theorem provides an asymptotic expansion for the estimator $\hat{\theta}_T$. Appendix A.2 provides a formal proof of this result. Let $G_{v_t}$ denote the second partial derivative of $G$ with respect to $v_t$, and let $IQ^t$ denote the integrated quarticity

$$IQ^t = p \lim_{N_t \to \infty} \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t_j}^4.$$

**Theorem 1** Suppose that conditions A and B are satisfied. Then,

$$\hat{\theta}_T - \theta_0 = -(\Gamma^T W T)^{-1} \Gamma^T W G_T(\theta_0) - (\Gamma^T W T)^{-1} \Gamma^T W b_T(\theta_0) + o_p(T^{-1/2}),$$

where

$$b_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \sum_{j=0}^{p} \frac{1}{N_t} \mathbb{E} \left[ G_{v_{t-j} v_{t-j}} (X_t, V_t; \theta) IQ^{t-j} \right].$$

The first term on the right hand side of Equation (12) is the standard one that arises in a GMM procedure in the absence of any measurement error. The second term, on the other hand, is an additional bias term that arises because of the measurement error in realized variance as an estimator of the integrated variance. Note that the quantity $b_T(\theta)$ is of order $T^{-1}$ in probability (based on Assumption A1). Its relative magnitude depends on the assumption we make connecting $N$ and $T$. The direction of the bias depends on the partial derivatives of the moment conditions. We obtain the following result.

**Corollary 1** Suppose that $b_T(\theta_0) = o(T^{-1/2})$. Then, we have

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \implies N(0, \Sigma).$$

(13)

Note that this requires $N^x/T \to \infty$, where $x > \frac{1}{\gamma}$ and $\gamma > \varepsilon + \frac{1}{2}$. When (13) holds, standard inference can be applied. Specifically, since $G(X_t, V_t; \theta_0)$ is a martingale difference sequence, $\hat{\Sigma} = (\hat{\Gamma}^T W \hat{\Gamma})^{-1} \hat{\Gamma}^T W \hat{\Omega} W \hat{\Gamma} (\hat{\Gamma}^T W \hat{\Gamma})^{-1}$ is a consistent estimator of $\Sigma$, where

$$\hat{\Gamma} = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \theta^T} G(X_t, \hat{V}_t; \hat{\theta}_T),$$

$$\hat{\Omega} = \frac{1}{T} \sum_{t=1}^{T} G(X_t, \hat{V}_t; \hat{\theta}_T) G(X_t, \hat{V}_t; \hat{\theta}_T)^\top.$$
When the condition in Corollary 1 is not satisfied, Equation (13) does not hold. Define the bias corrected estimator, $\tilde{\theta}_T^{bc}$, as

$$\tilde{\theta}_T^{bc} = \tilde{\theta} + (\hat{\Pi}^\top W \hat{\Pi})^{-1}\hat{\Pi}^\top W \hat{b}_T(\hat{\theta}_T),$$

where

$$\hat{b}_T(\theta) = \frac{1}{T} \sum_{t=1}^T \sum_{j=0}^p \frac{1}{N_t} G_{d_{t-j}t_{t-j}} \left(X_t, \tilde{V}_t; \hat{\theta}_T\right) \hat{I}Q^{t-j},$$

and $\hat{I}Q^t = \frac{N_t}{3} \sum_{j=1}^{N_t/3} r_{t_j}^4$ is a consistent estimator of the integrated quarticity. In this case, we have the following result.

**Corollary 2** Suppose that the condition in Corollary 1 is not satisfied. Then, we have

$$\sqrt{T}(\tilde{\theta}_T - \theta_0) \Rightarrow N(0, \Sigma).$$

This result requires the weaker condition that $N_x^x / T \to \infty$, where $x > \frac{1}{2}$ and $\gamma > \varepsilon + \frac{1}{2} - \alpha$. This result is the basis of the application we conduct in the empirical section. In particular, it provides the basis for confidence intervals and test statistics regarding $\theta$, and provides the methodology to take into account the potential consequences of small intraperiod samples.

4 Simulation Results

We perform Monte Carlo simulations in order to examine the finite-sample performance of the estimators of $\theta$. We assume that the continuously compounded returns on the market portfolio are generated by a diffusion process:

$$dp(t) = \mu(t)dt + \sigma(t)dW_1(t).$$

Note that our nonparametric estimation approach, described in Sections 2 and 3, does not require us to specify the functional forms of the drift, $\mu(t)$, or diffusion, $\sigma(t)$, in Equation (14), i.e., the approach remains valid for any specific functional form specification for these stochastic processes.

We consider two different models for $\sigma(t)$ that have been employed extensively in the literature and shown to provide a good fit to the dynamic properties of returns. The first is the lognormal diffusion (see, e.g., Andersen, Benzoni, and Lund (2002)):

$$d\log \sigma^2(t) = -0.0136 \left( a_2 + \log \sigma^2(t) \right) dt + 0.1148dW_2(t).$$

$$dp(t) = \mu(t)dt + \sigma(t)dW_1(t).$$

Note that our nonparametric estimation approach, described in Sections 2 and 3, does not require us to specify the functional forms of the drift, $\mu(t)$, or diffusion, $\sigma(t)$, in Equation (14), i.e., the approach remains valid for any specific functional form specification for these stochastic processes.

We consider two different models for $\sigma(t)$ that have been employed extensively in the literature and shown to provide a good fit to the dynamic properties of returns. The first is the lognormal diffusion (see, e.g., Andersen, Benzoni, and Lund (2002)):
The second model is the \textit{GARCH}(1,1) diffusion (see, e.g., Andersen and Bollerslev (1998)):

\begin{equation}
\begin{aligned}
\frac{d\sigma^2(t)}{\sigma^2(t)} &= 0.035 \left( a_3 - \sigma^2(t) \right) dt + 0.144 \sigma^2(t) dW_3(t).
\end{aligned}
\end{equation}

The Brownian motions \( W_2 \) and \( W_3 \) are assumed to be independent of \( W_1 \), i.e., there are no leverage and volatility feedback effects.

Our modeling of \( \mu(t) \) is motivated by the empirical specification of the risk-return relation considered in this paper. In particular, we consider a linear time-invariant relation between the conditional mean and the conditional variance of the stock market return, obtained by setting \( b_1 (x_{t-1}) = b_1 \) (a scalar constant), \( b_2 = 0 \), and \( \text{var} \left( r_{m,t} | \mathcal{F}_{t-1} \right) \approx E[v_t | \mathcal{F}_{t-1}] = v_{t-1} \) in Equation (1):

\begin{equation}
\mu(t) = b_0 + b_1 \sigma^2(t - 1).
\end{equation}

Note that time-aggregating (17) and taking conditional expectations of both sides with respect to time \( t - 1 \) information set delivers Equation (1). In the simulations, we set \( b_0 = 0 \) and \( b_1 = 2 \). These choices of parameter values are motivated by the CAPM that implies such a linear time-invariant relation between the first and second conditional moments of returns with \( b_1 \) corresponding to the risk aversion coefficient of the representative investor.\footnote{The simulation results are largely similar for alternative values of \( b_1 \) between 1 and 10 and are available from the authors upon request.}

Finally, we assume that the above specifications of the drift and diffusion processes generate high frequency (daily) returns on the market portfolio. The parameters \( a_2 \) and \( a_3 \) in (15) and (16), respectively, are calibrated to match the second moment of high-frequency (daily) squared returns within the low-frequency (monthly, quarterly, semi-annual, and annual) horizons considered. This yields \( a_2 = 6.03, 4.95, 4.50, \) and 3.50, when the normalized unit time interval corresponds to a month, quarter, semi-annual, and annual time horizon, respectively. The corresponding values for \( a_3 \) are \( 0.002, 0.007, 0.014, \) and 0.028, respectively. The monthly market return is computed as the sum of daily continuously compounded market returns and the realized monthly market variance as the sum of squares of the daily continuously compounded market returns. The quarterly, semi-annual and annual returns and realized market variances are calculated analogously. These are then used in the GMM estimation problem (9), with the lagged realized variance being used as an instrument, to estimate the parameter vector, \( \theta \). This procedure is repeated 2000 times.

Table I reports the simulation results for the \textit{GARCH}(1,1) model for the diffusion process, \( \sigma(t) \). Consider first Panel A. Each row of Panel A reports results for \( N = 22 \) and a different value of \( T \). In the context of our empirical application, this corresponds to estimating the risk-return relation at
the monthly frequency using daily returns to estimate the integrated variance. \( T = 1000 \) corresponds to the length of the historical time series (\( \sim 84 \) years). We also report results for smaller and larger values of \( T \) to show the effect of increasing the number of lower frequency time periods on the performance of the estimators.

Panel A, Row 1 corresponds to \( N = 22 \) high frequency data within each of \( T = 500 \) time periods. The second and third columns report the mean, the standard deviation (in parentheses), and 95% confidence interval (in square brackets) of the estimators of \( b_0 \) and \( b_1 \), respectively, across the 2000 simulations. The fourth and fifth columns report the same statistics as columns two and three but for the bias-corrected estimators of these parameters. Panel A, Row 1 reveals that the bias correction proposed in Section 3 substantially reduces the bias in estimating the risk-return tradeoff coefficient, \( b_1 \), even for small values of \( N \) and \( T \). The mean of the standard GMM estimator \( \hat{b}_1 \) across the simulations is 1.29 while the mean of the bias-corrected estimator, \( \hat{b}_{1}^{bc} \), is 1.70 - much closer to the population value of 2. Rows 2, 3, and 4 of Panel A show that, not surprisingly, increasing \( T \) to 1000, 5000, and 10,000, respectively, reduces the bias of both the standard estimators as well the bias-corrected ones. Note, however, that while the bias of the standard estimator remains substantial at 0.50 even for \( T = 10,000 \), that of the bias-corrected estimator is reduced to 0.13 for this value of \( T \).

Consider next Panel B. Each row of Panel B reports results for \( T = 1000 \) and a different value of \( N \). In the context of our empirical application, this corresponds to estimating the risk-return relation at a particular frequency (monthly, quarterly, semi-annual, or annual based on the value of \( N \)) using daily returns within that period to estimate the integrated variance. In particular, \( N = 22, 66, 132, \) and 262 correspond to monthly, quarterly, semi-annual, and annual frequencies, respectively. This allows us to study the effect of increasing the number of high frequency time periods within a low frequency period on the performance of the estimators. The results in Panel B show that, similar to Panel A, the bias-corrected estimator of \( b_1 \) has a much smaller bias for all the values of \( N \) compared to the standard estimator.

Table II reports simulation results for the lognormal model for the diffusion term, \( \sigma(t) \). The results are largely similar to those obtained in Table 1 for the GARCH(1,1) diffusion model. Panels A and B shows that the bias of the bias-corrected estimator is substantially smaller than that of the standard estimator for all combinations of \( N \) and \( T \).
5 Empirical Results

5.1 Data

In our empirical analysis, we focus on the risk-return relation at the monthly and quarterly frequencies. The data is from the Centre for Research in Security Prices (CRSP) daily returns data file. Our market proxy is the CRSP value-weighted index (all stocks on the NYSE, AMEX, and NASDAQ). The sample extends from January 1927 - December 2010. The monthly market return is obtained as the sum of daily continuously compounded market returns and the realized monthly market variance as the sum of squares of the daily continuously compounded market returns. The quarterly returns and realized market variances are computed analogously.

To set the stage, Table III reports summary statistics for the returns and the corresponding realized variances. Panels A and B report results at the monthly and quarterly frequencies, respectively. The table reports results for the full sample and for two subsamples of equal length. Panel A shows that the monthly market return has mean 0.8% and volatility 5.4% in the full sample. Returns are negatively skewed and highly leptokurtic, with the coefficients of skewness and kurtosis being -0.55 and 9.58, respectively, over the full sample. The first order autocorrelation coefficient of monthly returns is small at 0.11. The sum of the first 12 autocorrelation coefficients remains small at 0.22. The mean market return is similar across the two subsamples while the volatility of monthly returns is higher in the first subsample (6.1% vs. 4.7%). Both subsamples exhibit negative skewness and high kurtosis. The realized variance has a mean of 0.3% in the overall sample, which closely matches the variance of monthly returns over the same period. The realized variance process displays considerable persistence, with an autocorrelation coefficient of 0.61 and the sum of the first 12 autocorrelation coefficients being equal to 4.10 in the entire sample, and has a much smaller volatility compared to monthly returns (.5% vs. 5.4%). As expected, realized variance is highly skewed and leptokurtic. Figure 1 plots the time series of the monthly returns and realized volatility over the full sample period.

Most of these characteristics of returns and realized variance persist at the quarterly horizon, as shown in Panel B. The coefficient of kurtosis in realized variances declines with the horizon (52.5 in monthly data vs. 28.8 in quarterly data). The degree of skewness in the realized variance also declines with the horizon (6.09 in monthly data vs. 4.55 in quarterly data). No such trend is noticed in the coefficients of skewness and kurtosis for returns.
5.2 Empirical Evidence on the Risk-Return Tradeoff

We first provide support for some of the assumptions underlying the theoretical framework in Sections 2 and 3. Over the sample period, the daily continuously compounded market returns are highly leptokurtic with the degree of excess kurtosis being 17.9. The evidence of very fat tails in the distribution of market returns highlights the importance of Assumption A1 in Lemma 1, which allows the volatility process to grow over time rather than restricting it to be uniformly bounded over all \( t = 1, 2, \ldots, T \) and \( j = 1, 2, \ldots, N_t \).

Also, note that the conditional moment restriction in (6) is obtained by arguing that the integrated variance is approximately unbiased for the conditional variance: \( \text{var}(r_{m,t}|\mathcal{F}_{t-1}) \approx E[v_t|\mathcal{F}_{t-1}] \). The approximation is exact with assumed constant mean returns or with mean returns measurable with respect to the previous time period. While both these assumptions are fairly strong, we show that the approximation is good at the monthly and quarterly horizons even in the absence of these assumptions. In particular, it can be shown that the conditional variance of the market return is the sum of three terms. The first term is the conditional mean of the integrated variance. The integrated variance may be consistently estimated using the realized variance and the latter has a monthly (quarterly) mean of 0.003 (0.008) in our sample period. The second term is the conditional variance of the mean process. The squared mean of market returns is 6.4×10^{-5} (5.3×10^{-4}) in monthly (quarterly) data. Finally, the third term is the conditional covariance between the innovation in the mean process and the integrated variance. The covariance between the market return and the realized market variance at the monthly (quarterly) frequency is −8.2×10^{-5} (−4.4×10^{-4}). Thus, the latter two terms are of smaller order of magnitude than the first term lending support to the approximate unbiasedness of the integrated variance for the conditional variance.

Next, we turn to our main empirical results. The analysis in Section 2 shows that the estimation of the risk-return trade-off parameters can be posed as a GMM estimation problem, with the moment specification in Equation (9) that we restate here for convenience:

\[
E \left[ (r_{m,t} - r_{f,t} - b_0 - b_1 (x_{t-1}) \hat{v}_t - b_2^z z_{t-1}) \otimes y_{t-1} \right] = 0,
\]

where \( \theta = (b_0, b_1, b_2^z) \) is the vector of parameters to be estimated, \( z_{t-1} \) is a vector of predetermined variables, and \( y_{t-1} \) is a vector of instruments. We report estimation results for different specifications of \( b_1 (x_{t-1}), z_{t-1}, \) and \( y_{t-1} \).

Our first specification is obtained by setting \( b_1 (x_{t-1}) = b_1 \) (a scalar constant), \( b_2^z = 0^7 \), and using the lagged integrated variance as an instrument. The rationale for using the lagged integrated variance as an instrument is that the integrated variance is a highly persistent process. The first order autocorrelation coefficient of the realized variance process is 0.61 and 0.54, respectively, in
monthly and quarterly data for the full sample. Hence, the lagged variance is useful in predicting
the contemporaneous variance which enters the moment specification. This makes it a good choice
of instrument improving the efficiency of the estimation procedure. This specification produces an
exactly identified system of two moment restrictions in two unknown parameters, $\theta = (b_0, b_1)$, to
be estimated. Note that for this specification of the moment restrictions and choice of instruments,
the bias-correction is identically zero (see Theorem 1). Table IV, Panel A reports the estimation
results at the monthly frequency. The first row presents results over the entire available sample
period while Rows 2 and 3 do the same for two non-overlapping subsamples of equal length. Row 1
shows that, over the full sample, the estimated coefficient of the conditional variance $b_1$ is negative
and not statistically significant. Rows 2 and 3 show that while the coefficient remains statistically
insignificant over the first subsample covering the period 1927-1968, it becomes marginally significant
in the second subperiod 1969-2010.

Our second specification is identical to the first except that the lagged (instead of contemporane-
ous) integrated variance is used in the moment restriction. This is justified under a martingale
assumption on the integrated variance process, $\text{var}(r_{m,t}|\mathcal{F}_{t-1}) \approx E[v_t|\mathcal{F}_{t-1}] = v_{t-1}$, and has been
frequently employed in the literature. Unlike the first specification, for this specification of the mo-
ment restrictions and choice of instruments, Theorem 1 suggests a bias-correction to the standard
GMM estimator to improve its performance. Table IV, Panel B reports the estimation results at the
monthly frequency. Row 1 shows that, over the full sample, both the standard GMM estimator as
well bias-corrected estimator of the coefficient $b_1$ are negative and statistically insignificant. While
the bias-correction does increase the magnitude of the estimate, it does not do so sufficiently to
make it significantly different from zero. Row 2 shows that similar results are obtained for the first
subsample. Row 3, however, shows that, for the second subsample, the standard GMM estimator
is significantly negative at the 5% level of significance. We obtain even stronger results for the
bias-corrected estimator that is statistically significant even at the 1% level.

The finding in Panels A and B of a negative unconditional correlation between the expected stock
market return and its conditional variance is consistent with those in Graham and Harvey (2008)
and Lettau and Ludvigson (2010). However, while the latter paper finds that the unconditional risk-
return relation is negative but not statistically different from zero, we find a strongly statistically
significant negative relation in the second subsample applying our proposed bias-correction to the
standard estimator.

Our final specification is obtained by setting $b_1 (x_{t-1}) = b_1$ (a scalar constant) and including the
lagged integrated variance and the lagged market return in $z_{t-1}$. This specification is motivated
by the findings in Whitelaw (1994), Brandt and Kang (2004), and Lettau and Ludvigson (2010)
that the lagged conditional mean and conditional variance are a statistically important feature of the empirical risk-return relation and that it is important to distinguish between the unconditional correlation and the conditional correlation (conditional on the lagged mean and variance) between the first two moments of the market return to uncover the relation between them. Thus, we have four parameters to estimate. We use two lags of the integrated variance and one lag of the market return as instruments giving an exactly identified system of four moment restrictions in four parameters. Table IV, Panel C reports the estimation results at the monthly frequency. The results show that the estimated coefficient of the conditional variance \( b_1 \) is positive for the full sample (Row 1) and the second subsample (Row 3). This is consistent with the findings in Lettau and Ludvigson (2010) who find a negative unconditional correlation between the expected market return and its conditional variance but a positive conditional correlation (conditional on the lagged mean and variance) between these moments. However, while the latter paper finds the positive conditional correlation to be strongly statistically significant, our estimate of the conditional correlation is not statistically significant. The bias-correction increases the magnitude of the point estimate but not sufficiently relative to its standard error to make it statistically significant.

Panels A, B, and C of Table V report results similar to those in Table IV for an over-identified system that includes the one-month Treasury Bill rate and the default spread as additional instruments. The results are very similar to those obtained in Table IV. Very similar results are obtained at the quarterly frequency and are omitted for the sake of brevity.

6 Conclusion and Extensions

In this paper, we propose an approach to estimating the risk-return relation that overcomes some of the limitations of existing empirical analyses. First, we focus on a nonparametric measure of the ex-post return variability over a finite time interval, namely integrated variance, that is unbiased for the conditional variance and may be consistently estimated using the realized variance that is computed as the sum of squares of high-frequency intra-period returns. This approach allows us to express the risk-return relation as feasible moment restrictions and we then estimate the parameters of the relation using the GMM approach. This approach, while being robust to potential misspecification in the assumed dynamics of the conditional moments, also overcomes the endogeneity problem inherent in a least squares regression of an estimate of the conditional mean on the estimate of the conditional variance.

Second, and more importantly, we offer a solution to the measurement error problem that arises because of the use of realized variance as a proxy for the latent integrated variance. Our asymptotic
framework requires $N \to \infty$ and $T \to \infty$, where $N$ denotes the number of high-frequency intra-
period returns used to compute the realized variance in every period, and $T$ denotes the number of 
low-frequency time-periods used in the GMM estimation. We derive the limiting distribution of 
the estimated coefficients under this double asymptotic framework. We find that under fairly strong 
conditions on $N$ and $T$, the estimates are $\sqrt{T}$-consistent and have the standard distribution as when 
there is no measurement-error. However, if the above condition is not satisfied, there is an asymptotic 
bias that would invalidate this approximation. In that case, we find that under weaker conditions 
on $N$ and $T$, a bias-corrected estimator has the standard limiting distribution. This improvement is 
particularly relevant in the empirical case we examine where $N$ is quite modest.

In the empirical analysis, we focus on the risk-return relation at the monthly and quarterly 
frequencies. We use $(N)$ daily returns of the CRSP value-weighted stock market index to obtain 
monthly and quarterly estimates of the realized variance. We then estimate the parameters of the risk-
return relation using the GMM approach with $T$ (monthly and quarterly, respectively) observations 
on the realized market returns and realized variance. We find a negative relation between the mean 
and the variance that is statistically insignificant over the entire available historical sample 1927–2010 
but is strongly statistically significant over the latter half of the sample period. Moreover, we find that 
the bias-correction that we propose is instrumental in delivering the strongly statistically significant 
results. This finding is robust to the choice of instruments. Upon inclusion of the lagged integrated 
variance and the lagged market return as additional right hand side variables in the specification of 
the moment restriction, we obtain a positive, albeit statistically insignificant, relation between the 
conditional mean and variance of the market return.

The paper makes an important methodological contribution to the extant literature on high-
frequency volatility estimation. Most work has currently been about just estimating that quantity 
itself and using it to compare discrete time models in settings where the noise is small. Our approach 
is concerned with small sample issues when using estimated realized volatility as regressors in the 
estimation of parameters associated with the unobserved quadratic variation. This involves a useful 
extension of the existing asymptotic results for realized volatility concerned with the uniformity of 
the estimation error. We establish the properties of the parameter estimates and propose a bias 
correction in the case where the estimation error is large.

For instance, there is much empirical evidence to support the presence of jump components in 
the diffusion process for asset prices. In such specifications, the realized variance is still a consistent 
estimate of some overall variance measure that includes the contributions from the jump part of the 
process (see Barndorff-Nielsen and Shephard (2004, 2006) ). It is also possible to estimate separately 
the contributions to variance from the jump part and from the continuous part. This may be useful
in some asset pricing contexts where these risk measures are priced differently.

Also, a multivariate extension of the framework may be used to estimate conditional linear factor pricing models like the conditional CAPM, the conditional Fama-French three factor model, and the conditional Carhart four factor model. The approach does not require any specific functional form assumptions either about the factor betas or the factor risk premia.

References


A Appendix

A.1 Proof of Lemma 1

\[ \sum_{j=1}^{N_t} r_j^2 = \frac{1}{N_t^2} \sum_{j=1}^{N_t} \mu_j^2 + \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma^2_{t_j} \eta_{t_j}^2 + \frac{2}{N_t^{3/2}} \sum_{j=1}^{N_t} \mu_j \sigma_{t_j} \eta_{t_j}. \]

Therefore,

\[ \hat{v}_t - v_t = \left( \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma^2_{t_j} \eta_{t_j}^2 - v_t \right) + \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma^2_{t_j} (\eta_{t_j}^2 - 1) + \frac{2}{N_t^{3/2}} \sum_{j=1}^{N_t} \mu_j \sigma_{t_j} \eta_{t_j} \]

\[ = J_1 + J_2 + J_3 + J_4. \] (18)

We have \( J_1 = O_p(N_t^{-\lambda}) \) by Assumption A3. Consider next the term \( J_2 \). We have,

\[ E \left[ \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma^2_{t_j} (\eta_{t_j}^2 - 1) \right] = \frac{1}{N_t} \sum_{j=1}^{N_t} E \left[ \sigma^2_{t_j} (\eta_{t_j}^2 - 1) \right] = 0 \]

\[ \text{var} \left[ \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma^2_{t_j} (\eta_{t_j}^2 - 1) \right] = \frac{1}{N_t^2} \sum_{j=1}^{N_t} E \left[ \sigma^4_{t_j} (\eta_{t_j}^2 - 1)^2 \right] 
\]

\[ + \frac{1}{N_t^2} \sum_{j=1}^{N_t} \sum_{k=1}^{N_t} \sum_{j \neq k} E \left[ \sigma^2_{t_j} \sigma^2_{t_k} (\eta_{t_j}^2 - 1)(\eta_{t_k}^2 - 1) \right], \]

\[ = E \left[ (\eta_{t_j}^2 - 1)^2 \right] \frac{1}{N_t} E \left( \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma^4_{t_j} \right), \]

\[ \leq E \left[ (\eta_{t_j}^2 - 1)^2 \right] \frac{1}{N_t} \frac{1}{N_t} \frac{1}{N_t} \]

\[ = O \left( \frac{1}{N_t} \right) = o(1), \text{ provided } \gamma > \epsilon \]

Note that the second-to-last line derives from Assumption A1. Therefore, \( J_2 = O_p\left( \sqrt{\frac{N_t}{N_t}} \right) \). Finally, we have \( J_3 = O_p(N_t^{-1}) \) and \( J_4 = O_p(N_t^{-1}) \) under our conditions.

Thus, \( J_1 \) is of smaller order than \( J_2 \), and \( J_3 \) and \( J_4 \) are smaller than \( J_1 \) provided \( 1 > \lambda > \frac{1}{2} \left( 1 - \frac{\gamma}{\epsilon} \right) \), and we have,
\[ \hat{v}_t - v_t \approx J_2 = \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t_j}^2 (\eta_{t_j}^2 - 1). \]

We assume that \( \eta_{t_j} \sim N(0, 1) \), \( \eta_{t_j}^2 - 1 \) is a martingale difference sequence with respect to \( \mathcal{F}_{t_{j-1}} \) and \( \sigma_{t_j} \in \mathcal{F}_{t_{j-1}} \). Then provided

\[
\frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t_j}^2 \rightarrow P \int_0^1 \sigma_{t+s}^2 ds, \\
\frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t_j}^4 \rightarrow P \int_0^1 \sigma_{t+s}^4 ds,
\]

we have

\[
\sqrt{\frac{N_t}{\int_0^1 \sigma_{t+s}^4 ds}} J_2 \implies N(0, \kappa),
\]

where \( \kappa = E(\eta_{t_j}^2 - 1)^2 = 2 \).

Now, by the Bonferroni inequality,

\[
\Pr \left[ \max_{1 \leq t \leq T} |\hat{v}_t - v_t| > \delta_T \right] \leq \sum_{t=1}^T \Pr \left[ \left| \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t_j}^2 (\eta_{t_j}^2 - 1) \right| > \delta_T \right].
\]

\[
= \sum_{t=1}^T \Pr \left[ \left| \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t_j}^2 (\eta_{t_j}^2 - 1) \right| > \delta_T, \max_{1 \leq t \leq T} \sigma_{t_j}^2 < N_t \right] + \\
\sum_{t=1}^T \Pr \left[ \left| \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t_j}^2 (\eta_{t_j}^2 - 1) \right| > \delta_T, \max_{1 \leq t \leq T} \sigma_{t_j}^2 \geq N_t \right]. \tag{19}
\]

Consider the first term of Equation (19). On the set \( \Sigma_T = \{ \max_{1 \leq t \leq T} \sigma_{t_j}^2 < N_t \} \), we can apply the exponential inequality for strongly-mixing time series processes (Theorem 1.4 of Bosq (1998)). Therefore,

\[
\Pr \left[ \left| \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t_j}^2 (\eta_{t_j}^2 - 1) \right| > \delta_T \right] = \Pr \left[ \left| \sum_{j=1}^{N_t} \sigma_{t_j}^2 (\eta_{t_j}^2 - 1) \right| > N_t \delta_T \right]
\]

\[
\leq a_1 \exp \left( - \frac{q \delta_T^2}{25n^2 + 5c \delta_T} \right) + a_2(k) \alpha \left( \left\lceil \frac{N_t}{q + 1} \right\rceil \right)^{\frac{1}{2k + 1}},
\]

where:

25
\[ a_1 = 2 \frac{N_t}{q} + 2 \left( 1 + \frac{\delta_T^2}{25m_2^2 + 5c\delta_T} \right), \quad \text{with } m_2 = \max_{1 \leq j \leq N_t} E \left[ \sigma^2_{\operatorname{i,j}}(\eta_{\operatorname{i,j}}^2 - 1)^2 \right] < \infty \]

\[ a_2(k) = 11N_t \left( 1 + \frac{5m_k}{\delta_T} \right), \quad \text{with } m_k = \max_{1 \leq j \leq N_t} \left\| \sigma^2_{\operatorname{i,j}}(\eta_{\operatorname{i,j}}^2 - 1) \right\|_k < \infty \]

for each \( N_t \geq 2 \), each integer \( q \in \left[ 1, \frac{N_t}{4} \right] \), each \( \delta_T > 0 \), and each \( k \geq 3 \). \( c > 0 \) depends on the distribution of the time series and is such that

\[ E \left| \sigma^2_{\operatorname{i,j}}(\eta_{\operatorname{i,j}}^2 - 1) \right|^k \leq c^{k-2}k!E \left[ \sigma^4_{\operatorname{i,j}}(\eta_{\operatorname{i,j}}^2 - 1)^2 \right] < \infty. \]

For simplicity, assuming \( N_t = N \) for all \( t \), we have

\[
\sum_{t=1}^{T} \Pr \left[ \left| \frac{1}{N} \sum_{j=1}^{N} \sigma^2_{\operatorname{i,j}}(\eta_{\operatorname{i,j}}^2 - 1) \right| > \delta_T \right] \\
\leq Ta_1 \exp \left( -\frac{q\delta_T^2}{25m_2^2 + 5c\delta_T} \right) + Ta_2(k) \alpha \left( \left[ \frac{N}{q+1} \right] \right)^{\frac{2k}{2k+1}}.
\]

Putting \( \delta_T = \frac{T^*}{N^{\phi}} \to 0 \) provided \( \gamma > 2\epsilon \), and \( q = N^\phi, \phi < 1 \), we have \( Ta_1 \exp \left( -\frac{q\delta_T^2}{25m_2^2 + 5c\delta_T} \right) + Ta_2(k) \alpha \left( \left[ \frac{N}{q+1} \right] \right)^{\frac{2k}{2k+1}} \to 0 \) and, therefore,

\[
\sum_{t=1}^{T} \Pr \left[ \left| \frac{1}{N} \sum_{j=1}^{N} \sigma^2_{\operatorname{i,j}}(\eta_{\operatorname{i,j}}^2 - 1) \right| > \delta_T \right] \to 0,
\]

provided \( \phi > 1 - \frac{2\epsilon}{\gamma} \).

Consider now the second term of Equation (19),
\[
\sum_{t=1}^{T} \Pr \left[ \left| \frac{1}{N} \sum_{j=1}^{N} \sigma_{tj}^2 (\eta_{tj}^2 - 1) \right| > \delta_T, \max_{1 \leq i \leq T} \sigma_{tj}^2 \geq N \right] \leq \sum_{t=1}^{T} \Pr \left[ \max_{1 \leq i \leq T} \sigma_{tj}^2 \geq N \right] = T \Pr \left[ \max_{1 \leq i \leq T} \sigma_{tj}^2 \geq N \right] \\
\leq T \sum_{t=1}^{T} \sum_{j=1}^{N} \Pr \left[ \sigma_{tj}^2 \geq N \right] \leq T^2 N \max_{1 \leq t \leq T} E \left( \frac{\sigma_{tj}^{2k}}{N^k} \right) \to 0,
\]

provided \( 2 + \gamma - \gamma k < 0 \), i.e., \( \gamma > \frac{2}{k-1} \). Thus, in order to have \( \gamma < 1 \), we require \( k > 3 \). The third line follows from the Bonferroni inequality and the fourth line follows from Assumption A4.

Thus, it follows that

\[
\max_{1 \leq t \leq T} \left| \hat{v}_t - v_t \right| = O_p \left( \frac{T^\gamma}{N^{1/2}} \right),
\]

provided \( \frac{\gamma}{2} - \epsilon < \gamma (k - 1) - 2 \) (this condition ensures that the second term on the right hand side of Equation (19) is smaller than the first). So, provided \( \alpha + \epsilon < \frac{\gamma}{2} \), the result follows.

We summarize below the conditions required for the above result:

1. \( 1 > \lambda > \frac{1}{2} \left( 1 - \frac{\epsilon}{\gamma} \right) \)
2. \( \gamma > 2 \epsilon \)
3. \( \phi > 1 - \frac{2 \epsilon}{\gamma} \)
4. \( \gamma > \frac{2}{k-1} \Rightarrow \text{to have } \gamma < 1, \text{ we require } k > 3 \)
5. \( \frac{\gamma}{2} - \epsilon < \gamma (k - 1) - 2 \)

For example, \( \epsilon = 0.3, \gamma = 0.8, \lambda = 0.35, \phi = 0.3 \) satisfies all of the above conditions and implies \( \alpha < 0.1 \). Note that our result \( \max_{1 \leq t \leq T} \left| \hat{v}_t - v_t \right| = o_p \left( T^{-\alpha} \right) = o_p \left( N^{-\alpha/\gamma} \right) \) implies a slower rate of convergence of \( \max_{1 \leq t \leq T} \left| \hat{v}_t - v_t \right| \) compared to the convergence rate of \( N^{-1/2} \) for \( \left| \hat{v}_t - v_t \right| \) in a single period as derived in Barndorff-Nielsen and Shephard (2002, 2004).
A.2 Proof of Theorem 1

A.2.1 Consistency of $\hat{\theta}_T$

We just verify the Uniform Law of Large Numbers (ULLN) condition. By the triangle inequality

$$\sup_{\theta \in \Theta} \|\hat{G}_T(\theta) - \overline{G}(\theta)\|_W \leq \sup_{\theta \in \Theta} \|\hat{G}_T(\theta) - G_T(\theta)\|_W + \sup_{\theta \in \Theta} \|G_T(\theta) - \overline{G}(\theta)\|_W.$$ 

Let $A_T = \{\max_{1 \leq t \leq T} |\hat{v}_t - v_t| \leq \delta_T\}$, were $\delta_T$ is a sequence such that $\Pr(A_T) = o(1)$. Note that such a sequence is guaranteed by Lemma 1 with $\alpha = 0$, which just requires $\gamma > 2\epsilon$. Then

$$\Pr \left[ \sup_{\theta \in \Theta} \|\hat{G}_T(\theta) - G_T(\theta)\|_W > \eta \right] \leq \Pr \left[ \sup_{\theta \in \Theta} \|\hat{G}_T(\theta) - G_T(\theta)\|_W > \eta, A_T \right] + \Pr [A_T^c] = \Pr \left[ \sup_{\theta \in \Theta} \|\hat{G}_T(\theta) - G_T(\theta)\|_W > \eta, A_T \right] + o(1).$$

By the Mean Value Theorem, for a set of mean values $V_t$

$$\hat{G}_T(\theta) - G_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} G_V \left( X_t, \overline{V}_t; \theta \right)^\top (\hat{V}_t - V_t)$$

where $\overline{V}_t$ is intermediate between $\hat{V}_t$ and $V_t$. Furthermore, on the set $A_T$,

$$\sup_{\theta \in \Theta} \|\hat{G}_T(\theta) - G_T(\theta)\|_W = \sup_{\theta \in \Theta} \left\| \frac{1}{T} \sum_{t=1}^{T} G_V \left( X_t, \overline{V}_t; \theta \right)^\top (\hat{V}_t - V_t) \right\|_W$$

$$\leq \dim (V_t) \delta_T \frac{1}{T} \sum_{t=1}^{T} U_t = o_p(1),$$

where the second line follows from Assumption B7. Consistency then follows from the identification condition (Assumption B2) and the ULLN condition on the infeasible moment conditions $\sup_{\theta \in \Theta} \|G_T(\theta) - \overline{G}(\theta)\|_W = o_p(1)$ (Assumption B3).

A.2.2 Asymptotic Normality of $\hat{\theta}_T$

Under our conditions, the infeasible GMM estimator has the following limiting distribution:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, (\Gamma^\top W \Gamma)^{-1} \Gamma^\top W \Omega W^\top (\Gamma^\top W \Gamma)^{-1}).$$

For the asymptotic expansion, our proof parallels the work of Pakes and Pollard (1989). We expand the estimated moment condition out to third order. Therefore, for the $i$-th moment restriction, we
\[ \hat{G}_{i,T}(\theta_0) - G_{i,T}(\theta_0) = \frac{1}{T} \sum_{t=1}^{T} G_{i,V}(X_t, V_t; \theta_0)^\top (\hat{V}_t - V_t) + \frac{1}{2T} \sum_{t=1}^{T} (\hat{V}_t - V_t)^\top G_{i,VV}(X_t, V_t; \theta_0) (\hat{V}_t - V_t) + \frac{1}{6T} \sum_{t=1}^{T} G_{i,VVV}(X_t, \overline{V}_t; \theta_0) (\hat{V}_t - V_t)^\otimes 3 \]

where \( \overline{V}_t \) is intermediate between \( \hat{V}_t \) and \( V_t \).

Consider the first term

\[ \frac{1}{T} \sum_{t=1}^{T} G_{i,V}(X_t, V_t; \theta_0)^\top (\hat{V}_t - V_t) = \frac{1}{T} \sum_{t=1}^{T} \sum_{k=0}^{P} G_{i,\nu_{i-k}}(X_t, V_t; \theta_0) \frac{1}{N_{t-k}} \sum_{j=1}^{N_{t-k}} \sigma_{t-k_j}^2 (\eta_{t-k_j}^2 - 1). \]

For each of the terms on the right hand side of the above expression, we have

\[ E \left[ \frac{1}{T} \sum_{t=1}^{T} G_{i,\nu_{i-k}}(X_t, V_t; \theta_0) \frac{1}{N_{t-k}} \sum_{j=1}^{N_{t-k}} \sigma_{t-k_j}^2 (\eta_{t-k_j}^2 - 1) \right] = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{N_{t-k}} \sum_{j=1}^{N_{t-k}} E \left[ G_{i,\nu_{i-k}}(X_t, V_t; \theta_0) \right] E \left[ \sigma_{t-k_j}^2 (\eta_{t-k_j}^2 - 1) \right] = 0. \]

Also,

\[ \text{var} \left[ \frac{1}{T} \sum_{t=1}^{T} G_{i,\nu_{i-k}}(X_t, V_t; \theta_0) \frac{1}{N_{t-k}} \sum_{j=1}^{N_{t-k}} \sigma_{t-k_j}^2 (\eta_{t-k_j}^2 - 1) \right] = \frac{1}{T^2} \sum_{t=1}^{T} \frac{1}{N_{t-k}^2} \sum_{j=1}^{N_{t-k}} E \left[ G_{i,\nu_{i-k}}^2(X_t, V_t; \theta_0) \right] E \left[ \sigma_{t-k_j}^4 (\eta_{t-k_j}^2 - 1)^2 \right] = E \left[ (\eta_{t-k_j}^2 - 1)^2 \right] E \left[ G_{i,\nu_{i-k}}^2(X_t, V_t; \theta_0) \right] \frac{1}{TF} \sum_{t=1}^{T} \frac{1}{N_{t-k}} E \left( \frac{1}{N_{t-k}} \sum_{j=1}^{N_{t-k}} \sigma_{t-k_j}^4 \right) \leq \frac{1}{NT} M T^\kappa = O \left( \frac{T^\epsilon}{NT} \right) = o \left( \frac{1}{T} \right), \text{ provided } \gamma > \epsilon. \]
Next, consider the second term,

\[
\frac{1}{2T} \sum_{t=1}^{T} (\hat{V}_t - V_t)^\top G_{VV}(X_t, V_t; \theta_0)(\hat{V}_t - V_t)
\]

\[
= \frac{1}{2T} \sum_{t=1}^{T} \left[ \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma^2_{t,j}(\eta_{i,j}^2 - 1), \ldots, \frac{1}{N_{t-p}} \sum_{j=1}^{N_{t-p}} \sigma^2_{t-p,j}(\eta_{t-p,j}^2 - 1) \right] G_{VV}(X_t, V_t; \theta_0)
\]

\[
\left[ \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma^2_{t,j}(\eta_{i,j}^2 - 1), \ldots, \frac{1}{N_{t-p}} \sum_{j=1}^{N_{t-p}} \sigma^2_{t-p,j}(\eta_{t-p,j}^2 - 1) \right]^\top
\]

\[
= \sum_{k=0}^{p} \sum_{l=0}^{p} \frac{1}{2T} \sum_{t=1}^{T} \frac{1}{N_{t-k}N_{t-l}} \sum_{j=1}^{N_j} G_{V_kV_t}(X_t, V_t; \theta_0) \sigma^2_{t-k,j}(\eta_{t-k,j}^2 - 1) \sigma^2_{t-l,j}(\eta_{t-l,j}^2 - 1).
\]

We have

\[
E \left[ \frac{1}{2T} \sum_{t=1}^{T} (\hat{V}_t - V_t)^\top G_{VV}(X_t, V_t; \theta_0)(\hat{V}_t - V_t) \right]
\]

\[
= \frac{1}{2T} \sum_{t=1}^{T} \text{tr} \left( E \left[ (\hat{V}_t - V_t)(\hat{V}_t - V_t)^\top \right] E \left[ G_{VV}(X_t, V_t; \theta_0) \right] \right)
\]

\[
\simeq \sum_{k=0}^{p} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{N_{t-k}} E \left[ G_{V_kV_t}(X_t, V_t; \theta_0) \right] E \left[ IQ^{t-k} \right],
\]

where the last line follows because

\[
E[(\hat{V}_t - V_t)(\hat{V}_t - V_t)^\top | \sigma] = \text{diag} \left( \frac{1}{N_t^2} \sum_{j=1}^{N_t} \sigma^4_{t,j} E(\eta_{i,j}^2 - 1)^2, \ldots, \frac{1}{N_{t-p}^2} \sum_{j=1}^{N_{t-p}} \sigma^4_{t-p,j} E(\eta_{t-p,j}^2 - 1)^2 \right)
\]

\[
\simeq \text{diag} \left( \frac{1}{N_t} E(\eta_{i,j}^2 - 1)^2 \int_0^1 \sigma^4_{t+s} ds, \ldots, \frac{1}{N_{t-p}} E(\eta_{t-p,j}^2 - 1)^2 \int_0^1 \sigma^4_{t-p+s} ds \right).
\]

Similar calculations show that \( \text{var} \left[ \frac{1}{2T} \sum_{t=1}^{T} (\hat{V}_t - V_t)^\top G_{VV}(X_t, V_t; \theta_0)(\hat{V}_t - V_t) \right] = o \left( \frac{1}{T} \right) \) provided \( \gamma > \epsilon \).

Finally, we consider the third order terms,
\[
\frac{1}{6T} \sum_{t=1}^{T} G_{v_t v_t}(X_t, \tilde{V}_t \theta_0)(\tilde{v}_t - v_t)^3 \\
\leq \left( \max_{1 \leq t \leq T} |\tilde{v}_t - v_t| \right)^3 \frac{1}{6T} \sum_{t=1}^{T} \sup_{|x|, |x'|-|x''| \leq \delta_T} |G_{vvv}(X_t, v_t + x, v_{t-1} + x, v_{t-2} + x, \theta_0)| \\
= O_p(T^{-3\alpha}).
\]

For this term to be \(O_p(T^{-1/2})\), we require \(\alpha \geq 1/6\). This requires \(\gamma > \frac{1}{3}(1 + \epsilon)\).

Hence,

\[
\hat{G}_T(\theta_0) \approx G_T(\theta_0) + \sum_{k=0}^{p} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{N_{t-k}} E \left[ G_{v_t v_t v_t}(X_t, V_t; \theta_0) \right] E \left[ IQ^{t-k} \right] \\
= G_T(\theta_0) + b_T(\theta_0).
\]

Therefore, we have

\[
\hat{\theta}_T - \theta_0 = -(\Gamma^T W\Gamma)^{-1}\Gamma^T W G_T(\theta_0) - (\Gamma^T W\Gamma)^{-1}\Gamma^T W b_T(\theta_0) + o_p(T^{-1/2}).
\]

**Corollary 1:** \(\sqrt{T}b_T(\theta_0) = o_p(1)\)

\[
b_T(\theta_0) = \sum_{k=0}^{p} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{N_{t-k}} E \left[ G_{v_t v_t v_t}(X_t, V_t; \theta_0) \right] E \left[ IQ^{t-k} \right]
\]

Now,

\[
\frac{1}{T} \sum_{t=1}^{T} \frac{1}{N_{t-k}} E \left[ G_{v_t v_t v_t}(X_t, V_t; \theta_0) \right] E \left[ IQ^{t-k} \right] \leq E \left[ G_{v_t v_t v_t}(X_t, V_t; \theta_0) \right] \frac{1}{T} \sum_{t=1}^{T} \frac{1}{N} MT^e \\
= O \left( \frac{T^e}{N} \right) = o \left( T^{-1/2} \right),
\]

provided \(\gamma > \epsilon + 1/2\). Therefore, this requires \(\frac{N^x}{T} \to \infty\) where \(x > \frac{1}{\gamma}\), where \(\gamma > \epsilon + 1/2\). In this case,

\[
\sqrt{T} \left( \hat{\theta}_T - \theta_0 \right) = -(\Gamma^T W\Gamma)^{-1}\Gamma^T W \sqrt{T} G_T(\theta_0) + o_p(1).
\]

Hence,
\[ \sqrt{T} \left( \hat{\theta}_T - \theta_0 \right) \xrightarrow{d} N(0, \Sigma), \quad \text{where } \Sigma = (\Gamma^T W \Gamma)^{-1} \Gamma^T W \Omega W \Gamma (\Gamma^T W \Gamma)^{-1}. \]

**Corollary 2:** When the above condition is not satisfied, we may not have \( T^{1/2} \) consistency because of the asymptotic bias. However, we show that a bias corrected estimator \( \tilde{\theta} + (\Gamma^T W \Gamma)^{-1} \Gamma^T W b_T(\theta_0) \) would be \( T^{1/2} \) consistent. We propose to make a bias correction, which requires that we estimate \( b_T(\theta_0) \). Provided the estimation error is small enough we will achieve the limiting distribution in (13).

Define the estimated bias function

\[
\hat{b}_T(\theta) = \sum_{k=0}^{p} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{N_{t-k}} G_{V_{t-k} V_{t-k}} (X_t, \hat{V}_t; \theta) \hat{I}^t_{Q}^{-k}
\]

where \( \hat{Q}^t = \frac{N_t}{3} \sum_{j=1}^{N_t} r_{t_j}^4 \)

is an estimator of the integrated quarticity. Then for the bias corrected estimator

\[ \hat{\theta}^bc = \hat{\theta}_T + (\Gamma^T W \Gamma)^{-1} \Gamma^T W \hat{b}_T(\hat{\theta}_T). \]

we have that

\[ \sqrt{T} (\hat{\theta}^bc - \theta_0) \Rightarrow N(0, \Sigma). \]

provided

\[ \sqrt{T} b_T(\theta_T) - \sqrt{T} b_T(\theta_0) = o_p(1). \]

This requires \( \frac{N^x}{T} \to \infty \) where \( x > \frac{1}{\gamma} \), where \( \gamma > \epsilon + \frac{1}{2} - \alpha \). 

\[ \blacksquare \]
### B Tables

**Table I: Simulation Results for GARCH(1,1) Diffusion; $b_0 = 0, b_1 = 2$**

<table>
<thead>
<tr>
<th></th>
<th>Estimators</th>
<th>Bias-Corrected Estimators</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b_0$</td>
<td>$b_1$</td>
</tr>
<tr>
<td><strong>Panel A: Effect of Changing $T$ Keeping $N$ Fixed</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N = 22; T = 500$</td>
<td>.001</td>
<td>1.29</td>
</tr>
<tr>
<td></td>
<td>(.004)</td>
<td>(2.18)</td>
</tr>
<tr>
<td>$N = 22; T = 1,000$</td>
<td>.001</td>
<td>1.40</td>
</tr>
<tr>
<td></td>
<td>(.003)</td>
<td>(1.50)</td>
</tr>
<tr>
<td></td>
<td>$[-.004,.007]$</td>
<td>$[-1.72,4.24]$</td>
</tr>
<tr>
<td>$N = 22; T = 5,000$</td>
<td>.001</td>
<td>1.46</td>
</tr>
<tr>
<td></td>
<td>(.001)</td>
<td>(.71)</td>
</tr>
<tr>
<td></td>
<td>$[-.002,.004]$</td>
<td>[.013,2.85]</td>
</tr>
<tr>
<td>$N = 22; T = 10,000$</td>
<td>.001</td>
<td>1.50</td>
</tr>
<tr>
<td></td>
<td>(.001)</td>
<td>(.50)</td>
</tr>
<tr>
<td></td>
<td>$[-.001,.003]$</td>
<td>[.53,2.48]</td>
</tr>
<tr>
<td><strong>Panel B: Effect of Changing $N$ Keeping $T$ Fixed</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N = 22; T = 1,000$</td>
<td>.001</td>
<td>1.40</td>
</tr>
<tr>
<td></td>
<td>(.003)</td>
<td>(1.50)</td>
</tr>
<tr>
<td></td>
<td>$[-.004,.007]$</td>
<td>$[-1.72,4.24]$</td>
</tr>
<tr>
<td>$N = 66; T = 1,000$</td>
<td>.002</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td>(.006)</td>
<td>(.93)</td>
</tr>
<tr>
<td>$N = 132; T = 1,000$</td>
<td>.002</td>
<td>1.88</td>
</tr>
<tr>
<td></td>
<td>(.008)</td>
<td>(.67)</td>
</tr>
<tr>
<td>$N = 264; T = 1,000$</td>
<td>.002</td>
<td>1.93</td>
</tr>
<tr>
<td></td>
<td>(.012)</td>
<td>(.48)</td>
</tr>
<tr>
<td></td>
<td>$[-.022,.026]$</td>
<td>[.97,2.87]</td>
</tr>
</tbody>
</table>

The table reports simulation results for the system $E_{t-1} [r_{m,t} - b_0 - b_1 v_{t-1}] = 0$. The diffusion term $\sigma(t)$ is assumed to follow a GARCH(1,1) diffusion, while the drift term $\mu(t) = b_0 + b_1 \sigma^2(t - 1)$ with $b_0=0$ and $b_1=2$. The table reports the mean, standard deviation (in parentheses), and the 95% confidence interval (in square brackets) of the standard GMM parameter estimates and the bias-corrected estimates across 2000 simulations.
The table reports simulation results for the system $E_{t-1} \left[ r_{m,t} - b_0 - b_1 v_{t-1} \right] = 0$. The diffusion term $\sigma \left( t \right)$ is assumed to follow a lognormal diffusion, while the drift term $\mu \left( t \right) = b_0 + b_1 \sigma^2 \left( t-1 \right)$ with $b_0 = 0$ and $b_1 = 2$. The table reports the mean, standard deviation (in parentheses), and the 95% confidence interval (in square brackets) of the standard GMM parameter estimates and the bias-corrected estimates across 2000 simulations.

### Table II: Simulation Results for Lognormal Diffusion; $b_0 = 0$, $b_1 = 2$

<table>
<thead>
<tr>
<th></th>
<th>Estimators</th>
<th></th>
<th>Estimators</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b_0$</td>
<td>$b_1$</td>
<td>$b_{bc}^0$</td>
</tr>
<tr>
<td><strong>Panel A: Effect of Changing $T$ Keeping $N$ Fixed</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N = 22; T = 500$</td>
<td>(.002)</td>
<td>1.43</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>(1.70)</td>
<td>(.006)</td>
<td>(2.22)</td>
</tr>
<tr>
<td></td>
<td>[.007, .011]</td>
<td>[-2.08, 4.72]</td>
<td>[-.011, .012]</td>
</tr>
<tr>
<td>$N = 22; T = 1,000$</td>
<td>(.003)</td>
<td>1.49</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>(1.13)</td>
<td>(.004)</td>
<td>(1.41)</td>
</tr>
<tr>
<td></td>
<td>[.005, .008]</td>
<td>[-.79, 3.68]</td>
<td>[.007, .008]</td>
</tr>
<tr>
<td>$N = 22; T = 5,000$</td>
<td>(.001)</td>
<td>1.57</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>(.49)</td>
<td>(.002)</td>
<td>(.59)</td>
</tr>
<tr>
<td></td>
<td>[.002, .004]</td>
<td>[.61, 2.51]</td>
<td>[.003, .004]</td>
</tr>
<tr>
<td>$N = 22; T = 10,000$</td>
<td>(.001)</td>
<td>1.61</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>(.36)</td>
<td>(.001)</td>
<td>(.42)</td>
</tr>
<tr>
<td></td>
<td>[.001, .003]</td>
<td>[.92, 2.34]</td>
<td>[.002, .003]</td>
</tr>
</tbody>
</table>

|                  | Estimators |                  | Estimators |
| **Panel B: Effect of Changing $N$ Keeping $T$ Fixed** |
| $N = 22; T = 1,000$ | (.002)     | 1.49             | .000       | 1.84       |
|                  | (1.13)     | (.004)           | (1.41)     |
|                  | [.005, .008] | [-.79, 3.68]    | [.007, .008] | [.99, 4.56] |
| $N = 66; T = 1,000$ | (.006)     | 1.79             | .000       | 1.97       |
|                  | (.73)      | (.006)           | (.81)      |
| $N = 132; T = 1,000$ | (.007)     | 1.90             | .000       | 1.20       |
|                  | (.59)      | (.008)           | (.62)      |
| $N = 264; T = 1,000$ | (.012)     | 1.94             | .000       | 2.00       |
|                  | (.37)      | (.012)           | (.38)      |
|                  | [.022, .026] | [1.20, 2.67]    | [.024, .024] | [1.23, 2.75] |
The table reports summary statistics of the continuously compounded returns on the stock market, $r_{m,t}$, and the associated realized variance, $\hat{\sigma}_t$. Our market proxy is the CRSP value-weighted index (all stocks on the NYSE, AMEX, and NASDAQ). Statistics are reported for the monthly and quarterly frequencies. Monthly returns are calculated by compounding daily returns within calendar months. Monthly realized variances are constructed by cumulating squares of daily returns within each month, and so on. The table shows the mean, standard deviation, skewness, kurtosis, first-order autocorrelation, and the sum of the first 12 autocorrelation coefficients AC(1-12) for each of the variables. The statistics are shown for the full sample 1927-2010 and for two subsamples of equal length.
Table IV: Estimation Results at the Monthly Frequency

<table>
<thead>
<tr>
<th>Sample</th>
<th>Bias-Uncorrected</th>
<th>Bias-Corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimators</td>
<td>Estimators</td>
</tr>
<tr>
<td></td>
<td>$b_0$ $b_1$ $b_2$ $b_3$ $b_0^c$ $b_1^c$ $b_2^c$ $b_3^c$</td>
<td></td>
</tr>
<tr>
<td>1927-2010</td>
<td>.01 (-.95)</td>
<td>.01 (-1.11)</td>
</tr>
<tr>
<td>1927-1968</td>
<td>.01 (-.95)</td>
<td>.01 (-.29)</td>
</tr>
<tr>
<td>1969-2010</td>
<td>.01 (-2.00)</td>
<td>.01 (-1.40)</td>
</tr>
</tbody>
</table>

Panel A: $E_{t-1} [r_{m,t} - b_0 - b_1 v_t] = 0$

<table>
<thead>
<tr>
<th>Sample</th>
<th>Bias-Uncorrected</th>
<th>Bias-Corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimators</td>
<td>Estimators</td>
</tr>
<tr>
<td></td>
<td>$b_0$ $b_1$ $b_2$ $b_3$ $b_0^c$ $b_1^c$ $b_2^c$ $b_3^c$</td>
<td></td>
</tr>
<tr>
<td>1927-2010</td>
<td>(5.15) $- .63$</td>
<td>(5.32) $- .72$</td>
</tr>
<tr>
<td>1927-1968</td>
<td>(3.07) $- .25$</td>
<td>(3.11) $- .29$</td>
</tr>
<tr>
<td>1969-2010</td>
<td>(4.75) $- 1.20$</td>
<td>(4.98) $- 1.40$</td>
</tr>
</tbody>
</table>

Panel B: $E_{t-1} [r_{m,t} - b_0 - b_1 v_{t-1}] = 0$

<table>
<thead>
<tr>
<th>Sample</th>
<th>Bias-Uncorrected</th>
<th>Bias-Corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimators</td>
<td>Estimators</td>
</tr>
<tr>
<td></td>
<td>$b_0$ $b_1$ $b_2$ $b_3$ $b_0^c$ $b_1^c$ $b_2^c$ $b_3^c$</td>
<td></td>
</tr>
<tr>
<td>1927-2010</td>
<td>.04 (-.39)</td>
<td>.01 (-.73)</td>
</tr>
<tr>
<td>1927-1968</td>
<td>.16 (-.33)</td>
<td>.04 (-.46)</td>
</tr>
<tr>
<td>1969-2010</td>
<td>-.01 (-.73)</td>
<td>.01 (-.78)</td>
</tr>
</tbody>
</table>

Panel C: $E_{t-1} [r_{m,t} - b_0 - b_1 v_t - b_2 v_{t-1} - b_3 r_{m,t-1}] = 0$

The table reports estimation results at the monthly frequency for exactly identified systems.
Table V: Estimation Results at the Monthly Frequency for Over-Identified Systems

<table>
<thead>
<tr>
<th></th>
<th>Estimators</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( b_0 )</td>
<td>( b_1 )</td>
<td>( b_2 )</td>
<td>( b_3 )</td>
<td>Jstat ( b_0^c )</td>
<td>Jstat ( b_1^c )</td>
<td>Jstat ( b_2^c )</td>
<td>Jstat ( b_3^c )</td>
</tr>
<tr>
<td>Panel A: ( E_{t-1} [r_{m,t} - b_0 - b_1 v_t] = 0 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1927-2010</td>
<td>.01</td>
<td>-.99</td>
<td></td>
<td></td>
<td>( 1.2 \times 10^{-8} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.17)</td>
<td>(-.90)</td>
<td></td>
<td></td>
<td>(&gt; .10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1927-1968</td>
<td>.01</td>
<td>-.68</td>
<td></td>
<td></td>
<td>( 7.6 \times 10^{-8} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.92)</td>
<td>(-.46)</td>
<td></td>
<td></td>
<td>(&gt; .10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1969-2010</td>
<td>.01</td>
<td>-1.99</td>
<td></td>
<td></td>
<td>( 7.2 \times 10^{-8} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.12)</td>
<td>(-1.71)</td>
<td></td>
<td></td>
<td>(&gt; .10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel B: ( E_{t-1} [r_{m,t} - b_0 - b_1 v_{t-1}] = 0 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1927-2010</td>
<td>.009</td>
<td>-.64</td>
<td></td>
<td></td>
<td>( 1.4 \times 10^{-8} )</td>
<td>( .01 )</td>
<td>-.73</td>
<td>( 1.9 \times 10^{-8} )</td>
</tr>
<tr>
<td></td>
<td>(5.22)</td>
<td>(-.92)</td>
<td></td>
<td></td>
<td>(&gt; .10)</td>
<td>(5.37)</td>
<td>(-1.06)</td>
<td>(&gt; .10)</td>
</tr>
<tr>
<td>1927-1968</td>
<td>.009</td>
<td>-.43</td>
<td></td>
<td></td>
<td>( 1.1 \times 10^{-7} )</td>
<td>( .009 )</td>
<td>-.48</td>
<td>( 1.1 \times 10^{-7} )</td>
</tr>
<tr>
<td></td>
<td>(3.42)</td>
<td>(-.41)</td>
<td></td>
<td></td>
<td>(&gt; .10)</td>
<td>(3.48)</td>
<td>(-.46)</td>
<td>(&gt; .10)</td>
</tr>
<tr>
<td>1969-2010</td>
<td>.01</td>
<td>-1.11</td>
<td></td>
<td></td>
<td>( 4.7 \times 10^{-8} )</td>
<td>( .01 )</td>
<td>-1.29</td>
<td>( 5.5 \times 10^{-8} )</td>
</tr>
<tr>
<td></td>
<td>(4.60)</td>
<td>(-1.88)</td>
<td></td>
<td></td>
<td>(&gt; .10)</td>
<td>(4.82)</td>
<td>(-2.25)</td>
<td>(&gt; .05)</td>
</tr>
<tr>
<td>Panel C: ( E_{t-1} [r_{m,t} - b_0 - b_1 v_t - b_2 v_{t-1} - b_3 r_{m,t-1}] = 0 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1927-2010</td>
<td>.008</td>
<td>-.34</td>
<td>-.02</td>
<td>.10</td>
<td>( 2.7 \times 10^{-8} )</td>
<td>( .008 )</td>
<td>-.33</td>
<td>-.03</td>
</tr>
<tr>
<td></td>
<td>(2.02)</td>
<td>(-1.11)</td>
<td>(-.01)</td>
<td>(1.54)</td>
<td>(&gt; .10)</td>
<td>(2.02)</td>
<td>(-1.11)</td>
<td>(-.02)</td>
</tr>
<tr>
<td>1927-1968</td>
<td>.01</td>
<td>-3.53</td>
<td>2.29</td>
<td>.09</td>
<td>( 7.0 \times 10^{-9} )</td>
<td>( .01 )</td>
<td>-4.42</td>
<td>3.12</td>
</tr>
<tr>
<td></td>
<td>(2.35)</td>
<td>(-.97)</td>
<td>(1.01)</td>
<td>(.95)</td>
<td>(&gt; .10)</td>
<td>(2.39)</td>
<td>(-1.21)</td>
<td>(1.33)</td>
</tr>
<tr>
<td>1969-2010</td>
<td>-.003</td>
<td>7.56</td>
<td>-4.54</td>
<td>.15</td>
<td>( 1.5 \times 10^{-9} )</td>
<td>-.003</td>
<td>11.6</td>
<td>-7.33</td>
</tr>
<tr>
<td></td>
<td>(-.03)</td>
<td>(1.09)</td>
<td>(-1.04)</td>
<td>(1.34)</td>
<td>(&gt; .10)</td>
<td>(-.28)</td>
<td>(1.11)</td>
<td>(-1.07)</td>
</tr>
</tbody>
</table>

The table reports estimation results at the monthly frequency for over identified systems. The short term risk free rate and the default spread are used as instruments, in addition to the lagged integrated variance. The identity matrix is used as the weighting matrix. Therefore, the Jstat in each panel has a nonstandard distribution, the critical values of which are obtained by simulation.
The figure plots the time series of the monthly stock market returns and realized volatility over the full sample period 1927-2010. The realized variance for a given month is computed as the sum of squares of the daily returns within the month.