Working Paper 09-49
Economic Series (28)
July 2009

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 916249875

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

Anisha Ghosh† Oliver Linton‡

this version: July 19, 2009

Abstract

Prominent asset pricing models imply a linear, time-invariant relation between the equity premium and its conditional variance. We propose an approach to estimating this relation that overcomes some of the limitations of the existing literature. First, we do not require any functional form assumptions about the conditional moments. Second, the GMM approach is used to overcome the endogeneity problem inherent in the regression. Third, we correct for the measurement error arising because of using a proxy for the latent variance. The empirical findings reveal significant time-variation in the relation that coincide with structural break dates in the market-wide price-dividend ratio.

Keywords: Bias-Correction, Measurement Error, Nonparametric Volatility, Return, Risk.
JEL classification: C14, G12

* We thank Jushan Bai, Federico Bandi, George M. Constantinides, Valentina Corradi, Ernst Eberlein, Christian Julliard, Sydney Ludvigson, and Enrique Sentana for helpful comments. We remain responsible for errors and omissions. This research was supported by the ESRC and the Leverhulme foundation.

† Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom, e-mail: a.ghosh@lse.ac.uk

‡ Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom, e-mail: o.linton@lse.ac.uk. This paper was partly written while I was a Universidad Carlos III de Madrid-Banco Santander Chair of Excellence, and I thank them for financial support.

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.

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 external habit model of Campbell and Cochrane (1999), and the Long Run Risks model of Bansal and Yaron (2004). However, a negative risk-return relation is not inconsistent with equilibrium.[1]

Unfortunately, the empirical evidence on the risk-return relation is mixed and inconclusive. Ghysels, Santa-Clara, and Valkanov (2005), Lundblad (2005), Pastor, Sinha, and Swaminathan (2008), and Ludvigson and Ng (2007) find a positive risk-return relation, while Campbell (1987), Glosten, Jagannathan, and Runkle (1993), Harvey (2001), and Lettau and Ludvigson (2003) find a negative relation. Still others find mixed and inconclusive evidence like French, Schwert, and Stambaugh (1987), Nelson (1991), Campbell and Hentschel (1992), Linton and Perron (2003), and Whitelaw (1994). Scruggs (1998) and Guo and Whitelaw (2006) document a positive trade-off within specifications that facilitate hedging demands. However, Scruggs and Glabadanidis (2003) find that this partial relationship is not robust across alternative volatility specifications.

The main difficulty in estimating the risk-return relation is that neither the conditional expected return nor the conditional variance of the market is directly observable. The conflicting findings of the above studies are mostly the result of differences in the approaches to modeling the conditional mean and variance.

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.

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

[1] see, Abel (1988), Backus and Gregory (1993), and Whitelaw (2000).

mean and conditional variance will be misspecified and possibly highly misleading.[2]

In addition, the latter studies typically estimate the risk-return relation using a least squares regression of the estimate of the conditional mean on the estimate of the conditional variance. This approach suffers from a couple of shortcomings. First, the conditional mean and the conditional variance are simultaneously determined within the context of a general equilibrium asset pricing model. Hence, the least squares regression suffers from an endogeneity problem. Second, most of the literature ignores the measurement error that arises in this setting, as a result of using proxies for the latent conditional moments in estimating the risk-return relation.

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 expost 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.[3] Hence, under the maintained hypothesis of a linear, time-invariant relation between the conditional mean and the conditional variance, the above 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 and the integrated variance. 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.

Second, 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 if $N^\zeta/T \to \infty$, where $\zeta > 1.5$, 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 the weaker condition $N^\zeta/T \to \infty$, where $\zeta > 3$,

[2] See also, Campbell (1987) and Harvey (2001).

[3] For an excellent survey of this extensive literature, see Andersen, Bollerslev, and Diebold (2002); see also Barndorff-Nielsen and Shephard (2002), and Andersen, Bollerslev, Diebold, and Labys (2003).

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. Our work here is related to some recent work of Corradi, Distaso, and Swanson (forthcoming), Corradi and Distaso (2006).

In the empirical analysis, we focus on the risk-return relation at the monthly, quarterly, semiannual, and annual frequencies. We use (N) daily returns, within the corresponding period, on the CRSP value-weighted index to obtain monthly, quarterly, semiannual, and annual estimates of the realized variance. We then estimate the parameters of the risk-return relation using the GMM approach with T (monthly, quarterly, semiannual, and annual, respectively) observations on the realized excess market returns and realized variance. We find a statistically insignificant relation between the mean and the variance at all the frequencies considered over the entire available sample period 1928-2005. This finding is robust to the choice of instruments and across two subperiods of equal size.

To interpret the results, we turn to a closely related literature on return predictability that has reported evidence in favour of structural breaks in the OLS coefficient in the forecasting regression of returns on the lagged price-dividend ratio (e.g., Viceira (1996) and Paye and Timmermann (2006)). This renders the forecasting relationship unstable if such shifts are not taken into account. Lettau and Nieuwerburgh (2008) find evidence for two breaks in the mean of the log dividend-price ratio around 1954 and 1994. They demonstrate that if these breaks are ignored, the estimated OLS coefficient appears statistically insignificant over the full sample. However, when the sample is split into subsamples corresponding to the break dates, significant coefficient estimates are obtained in each subsample. These results suggest that if the relationship between expected returns and the conditional variance exhibits significant time variation, this could potentially render the estimated coefficient statistically insignificant when estimated over the entire sample.

To explore this possibility, we split the sample into three subsamples based on the break dates identified in Lettau and Nieuwerburgh (2008). The results reveal significant time-variation in the relation. In particular, the relation appears quite unstable in the first subsample that includes periods of great economic uncertainty like the Great Depression and World War I, significantly positive in the second subsample, and mostly significantly negative in the third subsample for the horizons considered.

Finally, 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[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.

The remainder of the paper is organized as follows. The theoretical underpinnings of the nonparametric variance estimator are discussed in Section 2. Section 3 describes the estimation procedure, while section 4 gives the asymptotic distribution of the estimated parameters. Section 5 provides the empirical results. In Section 6, we perform Monte-Carlo simulations to examine the finite-sample performance of the estimators. In the concluding Section 7, we discuss extensions of the approach and work in progress. The Appendix contains the proofs of our main results.

Notation. For matrix A, let $||A||_W = (\mathrm{tr}(A^\top W A))^{1/2}$ for symmetric positive definite W. For a function $f : \mathbb{R}^d \to \mathbb{R}$ and vector $v = (v_1, \ldots, v_d)$ denote $D^v f(x) = \partial^{v_1} \cdots \partial^{v_d} f(x)/\partial x_1^{v_1} \cdots \partial x_d^{v_d}$.

2 Nonparametric Variance Estimator

In this section we describe the conceptual framework behind the estimation of ex-post volatility. Under the assumptions that the return process does not allow for arbitrage and has a finite instantaneous mean, the asset price process, as well as smooth transformations thereof, belong to the class of special semi-martingales (see Back (1991)). If, in addition, it is assumed that the sample paths are continuous, the Martingale Representation Theorem (see Judd (1998) and Protter (2004)), implies the following representation for the return process over a finite interval of length, h:

Proposition 1 *For any square integrable, arbitrage free, logarithmic price process, $p(t)$, with continuous sample path, there exists a representation such that for all $0 \le t \le T$, a.s.(P)*

$$r(t,h) \equiv p(t) - p(t-h) = \mu(t,h) + M(t,h) = \int_0^h \mu(t-h+s)ds + \int_0^h \sigma(t-h+s)dW(s), \qquad (1)$$

where $\mu(s)$ denotes an integrable, predictable and finite variation drift, $\sigma(s)$ is a strictly positive *caglad volatility process satisfying*

$$\Pr\left[\int_0^h \sigma^2(t-h+s)ds < \infty\right] = 1,$$

and $W(s)$ is a standard Brownian motion.

[4] See Barndorff-Nielsen and Shephard (2002)

Here, $a.s.(P)$ denotes almost surely, under the objective probability measure, P. The integral representation (1) is equivalent to the stochastic differential equation specification for the logarithmic price process

$$dp(t) = \mu(t)dt + \sigma(t)dW(t). \tag{2}$$

Crucial to semimartingales, and to the economics of financial risk, is the *quadratic variation (QV)* process associated with it, denoted $[r, r]_t$:

Proposition 2 *Let a sequence of possibly random partitions of* $[0, T]$, (τ_m), *be given such that* $(\tau_m) \equiv \{\tau_{m,j}\}_{j\geq 0}$, $m = 1, 2, \ldots$, *where* $\tau_{m,0} \leq \tau_{m,1} \leq \tau_{m,2} \leq \ldots$ *satisfy, with probability one, for* $m \to \infty$,

$$\tau_{m,0} \to 0; \qquad \sup_{j\geq 1} \tau_{m,j} \to T; \qquad \sup_{j\geq 0} (\tau_{m,j+1} - \tau_{m,j}) \to 0.$$

Then, for $t\epsilon[0, T]$,

$$\lim_{m\to\infty} \left\{ \sum_{j\geq 1} [p(t \wedge \tau_{m,j}) - p(t \wedge \tau_{m,j-1})]^2 \right\} \to [r, r]_t, \tag{3}$$

where $t \wedge \tau \equiv \min\ (t, \tau)$, *and the convergence is uniform in probability.*

A natural theoretical notion of expost return variability in this setting is *notional volatility* (see, Andersen, Bollerslev, and Diebold (2002)). Under the maintained assumption of continuous sample path, the *notional volatility* equals the so-called *integrated volatility*:

Definition 1 *The Notional Volatility over* $[t-h, t]$ *is*

$$\upsilon^2(t, h) \equiv [r, r]_t - [r, r]_{t-h} = [M, M]_t - [M, M]_{t-h} = \int_0^h \sigma^2(t - h + s)ds. \tag{4}$$

It follows, from the properties of the quadratic variation process, that

$$E_{t-h}[\upsilon^2(t, h)] = E_{t-h}[M^2(t, h)]. \tag{5}$$

Now, the conditional variance over $[t-h, t]$, is defined by

$$\begin{aligned}
\mathrm{var}_{t-h}(r(t,h)) &\equiv E_{t-h}\left[\{r(t,h) - E_{t-h}(r(t,h))\}^2\right], \\
&= E_{t-h}\left[\{r(t,h) - E_{t-h}(\mu(t,h))\}^2\right], \\
&= E_{t-h}\left[\{\mu(t,h) - E_{t-h}(\mu(t,h)) + M(t,h)\}^2\right], \\
&= E_{t-h}[M^2(t,h)] + E_{t-h}\left[\{\mu(t,h) - E_{t-h}(\mu(t,h))\}^2\right] \\
&\quad +2E_{t-h}\left[\{\mu(t,h) - E_{t-h}(\mu(t,h))\} M(t,h)\right], \\
&= O_p(h) + O_p(h^2) + O_p(h^{3/2}),
\end{aligned} \tag{6}$$

where the second equality follows as M is a local martingale, and the third equality follows from (1). For a discussion of the orders of magnitude of the three terms as stated in the last equality, see Andersen, Bollerslev, and Diebold (2002). Thus, for small h, the first term dominates the other two, i.e., from equations (5) and (6), we have

$$\mathrm{var}_{t-h}\left(r(t,h)\right) \approx E_{t-h}[M^2(t,h)] = E_{t-h}[\upsilon^2(t,h)]. \tag{7}$$

In other words, the conditional variance is well approximated by expected notional (or integrated) volatility. The above approximation is exact if the mean process, $\mu(t) \equiv 0$, or if $\mu(t,h)$ is measurable with respect to I_{t-h}. However, the result remains approximately valid for a stochastically evolving mean return process over relevant horizons, as long as the returns are sampled at sufficiently high frequencies. We provide empirical evidence in Section 1.6.2 to justify this approximation for the horizons, h, considered in this paper.

Now, notional (or integrated) volatility is latent. However, it can be estimated consistently using the so-called *realized variance*:

Definition 2 *The Realized Variance over* $[t-h,t]$, *for* $0 < h \leq t \leq T$, *is defined by*

$$\widehat{\upsilon}^2(t,h;N) \equiv \sum_{i=1}^{N} r(t-h+(i/N)h, h/N)^2. \tag{8}$$

Thus, the realized variance is simply the second (uncentered) sample moment of the return process over a fixed interval of length h, scaled by the number of observations N (corresponding to the sampling frequency h/N), so that it provides a variance measure calibrated to the h-period measurement interval.

Protter (2004) shows that realized variance is (ex-ante) unbiased for the conditional variance:

Proposition 3 *Realized Variance as an unbiased variance estimator*

$$\mathrm{var}_{t-h}\left(r(t,h)\right) \approx E_{t-h}[\upsilon^2(t,h)] = E_{t-h}[\widehat{\upsilon}^2(t,h;N)], \tag{9}$$

for any $N \geqslant 1$ *and* $h > 0$, *with strict equality if* $\mu(t) \equiv 0$.

The theory of quadratic variation implies the following result (see, e.g., Andersen, Bollerslev, Diebold, and Labys (2003) and Barndorff-Nielsen and Shephard (2002)).

Proposition 4 *The Realized Variance provides a consistent nonparametric measure of the Notional Volatility*

$$p\lim_{N\to\infty} \widehat{\upsilon}(t,h;N) = \upsilon^2(t,h), \quad 0 < h \leq t \leq T, \tag{10}$$

where the convergence is uniform in probability.

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:

Proposition 5 *Suppose that* $p \in BSM$ *is one* $-$ *dimensional and that, for all* $t < \infty$, $\int_0^t \mu(s)ds < \infty$; *then as* $N \to \infty$

$$N^{1/2}\left(\widehat{v}^2(t,h;N) - \int_0^h \sigma^2(t-h+s)ds\right) \to \sqrt{2}\left(\int_0^h \sigma^2(t-h+s)dB(t-h+s)\right), \tag{11}$$

where B *is a Brownian motion independent of* W *in* (23) *and the convergence is in law stable as a process.*

Here, BSM denotes Brownian Semi-Martingale, i.e., of the form (1). The above theorem implies that

$$N^{1/2}\left(\widehat{v}^2(t,h;N) - \int_0^h \sigma^2(t-h+s)ds\right) \Longrightarrow MN(0, 2\int_0^h \sigma^4(t-h+s)ds), \tag{12}$$

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 $\int_0^h \sigma^4(t-h+s)ds$ can be consistently estimated using $(1/3)RQ_t$ where

$$RQ_t \equiv N\sum_{i=1}^{n} r(t-h+(i/N)h, h/N)^4. \tag{13}$$

In particular,

$$\frac{N^{1/2}\left(\widehat{v}^2(t,h;N) - \int_0^h \sigma^2(t-h+s)ds\right)}{\sqrt{\frac{2}{3}RQ_t}} \Longrightarrow N(0,1). \tag{14}$$

This is a nonparametric result as it does not require us to specify the form of the drift, $\mu(t)$, or the diffusion term, $\sigma(t)$, in (1) or (2). Without loss of generality, we set $h = 1$ in what follows.

Finally, there has been much recent work on estimating volatility in jump diffusion processes. There is much empirical evidence to support such specifications. In such cases, the realized volatility is still a consistent estimate of some overall volatility measure that includes the contributions from the jump part of the process (see Barndorff-Nielsen and Shephard (2004)). It is also possible to estimate separately the contributions to volatility 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.

3 Estimation

In this section we describe how we estimate the risk-return relationship. In our empirical work we will focus on a linear relation between the expected excess returns and the conditional variance of the aggregate stock market

$$E_{t-1}\left(r_{m,t} - r_{f,t}\right) = \alpha + \beta \mathrm{var}_{t-1}\left(r_{m,t}\right), \tag{15}$$

where $r_{m,t}$ and $r_{f,t}$ are the continuously compounded returns on the stock market and the risk free rate, respectively, over $[t-1,t]$. Note that such a linear, time-invariant relation is implied by the Intertemporal Capital Asset Pricing Model of Merton (1973) as well as the more recent Long Run Risks model of Bansal and Yaron (2004) model and its extension in Constantinides and Ghosh (2008) that allows for potential cointegration of the consumption and dividend levels.

Given the unbiasedness property of the integrated variance for the conditional variance in (9), the above risk-return relation implies the following conditional moment restriction

$$E_{t-1}\left(r_{m,t} - r_{f,t} - \alpha - \beta \upsilon_t\right) = 0, \tag{16}$$

where $\upsilon_t \equiv \upsilon(t,1) = \int_0^1 \sigma^2(t-1+s)ds$. The conditioning set includes all variables observed at time $t-1$.

Equation (16) is an infeasible moment restriction as the integrated variance, υ_t, is not observable. To obtain a proxy for it, we rely on the nonparametric variance estimator discussed in the previous section. We consider a discretized version of the continuous-time diffusion in (2). Let $\{r_{t_j}\}_{j=1}^{N_t}$ be intra-period continuously compounded returns on the market portfolio for each period $t = 1,\ldots,T$. In our empirical application, $j = 1,2,\ldots,N_t$ denotes days within time period t, while $t = 1,2,\ldots,T$ denotes months, quarters, semiannual, or annual time periods. Suppose that

$$r_{t_j} = N_t^{-1}\mu_{t_j} + N_t^{-1/2}\sigma_{t_j}\eta_{t_j}, \tag{17}$$

where $\eta_{t_j} \sim i.i.d.\ N(0,1)$ and η_{t_j} is independent of $\mathcal{F}_{t_{j-1}}$, where $\mathcal{F}_{t_{j-1}}$ contains all information upto time t_{j-1}. Also, suppose that $\{\mu_{t_j}, \sigma_{t_j}\}$ is measurable with respect to time t_{j-1} information set. The stochastic processes $\{\mu_{t_j}, \sigma_{t_j}\}_{j=1,t=1}^{N_t,T}$ are not assumed to be independent of the process $\{\eta_{t_j}\}_{j=1,t=1}^{N_t,T}$, i.e., we allow for the well documented leverage and volatility feedback effects. In particular, η_{t_j} can affect $\sigma_{s_{j+k}}$ for $s = t$, $k \geqslant 1$ and $s > t$, $k \geqslant 0$. The quantity $\sigma_{t_j}^2$ is the integral of the volatility function over a small interval, (see, e.g., Gonçalves and Meddahi (2009)). Define

$$\upsilon_t \equiv \operatorname*{plim}_{N_t \to \infty} \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t_j}^2 = \int_0^1 \sigma^2(t-1+s)ds. \tag{18}$$

Thus, υ_t is the quadratic variation of the underlying diffusion process (or the integrated variance).

In this setting, the integrated variance can be consistently estimated by the realized variance (see equations (3) and (8)),

$$\widehat{\upsilon}_t = \sum_{j=1}^{N_t} r_{t_j}^2. \tag{19}$$

Plugging the realized variance into the infeasible moment restriction (16), we obtain the feasible moment restriction,

$$E_{t-1}\left(r_{m,t} - r_{f,t} - \alpha - \beta\widehat{\upsilon}_t\right) = 0. \tag{20}$$

Finally, with a set of chosen conditioning variables, z_{t-1} (that could include, for instance, lagged variance), we have the unconditional moment restriction

$$E\left[G\left(r_{m,t} - r_{f,t}, \widehat{\upsilon}_t, z_{t-1}; \theta\right)\right] = 0, \tag{21}$$

where $G\left(r_{m,t} - r_{f,t}, \widehat{\upsilon}_t, z_{t-1}; \theta\right) = \left(r_{m,t} - r_{f,t} - \alpha - \beta\widehat{\upsilon}_t\right) \otimes z_{t-1}$, and $\theta = (\alpha, \beta)^\intercal$.

Given the above set of moment restrictions, the parameter vector θ may be estimated using the GMM approach. Specifically, we define the estimator $\widehat{\theta} \in \Theta \subseteq \mathbb{R}^p$ as the minimizer of

$$\widehat{\theta} = \arg\min_{\theta\in\Theta} \left\|\widehat{G}_T(\theta)\right\|_W, \quad \widehat{G}_T(\theta) \equiv \frac{1}{T}\sum_{t=1}^{T} G\left(r_{m,t} - r_{f,t}, \widehat{\upsilon}_t, z_{t-1}; \theta\right),$$

where W is a symmetric positive definite weighting matrix. The computation of $\widehat{\theta}$ is straightforward in our application as the moment conditions are linear in the parameters.

Although we focus on linear models, our estimation methodology is defined for general nonlinear moment conditions and we present our theoretical results for this case. This estimator can be viewed as a semiparametric estimator with preliminarily estimated nuisance function $\{\widehat{\upsilon}_t\}$, see Newey and McFadden (1994), except that in our case the estimated nuisance function is square root (intraperiod) sample size consistent and asymptotically mixed normal with zero bias. The other complication is that the true value of the nuisance function $\{\upsilon_t\}$ is a stochastic process.

4 Asymptotic Properties

In this section we derive an asymptotic approximation to the properties of our estimators of θ. 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 to accurately estimate the integrated variance, 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 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. Such consistency results are common in the semiparametric literature, Newey and McFadden (1994).

Our first result establishes the consistency of $\widehat{v}_t$ for v_t, uniformly in t. We need some regularity conditions.

Assumptions

(A1) There exists a small $\epsilon > 0$ such that with probability one for large enough T and some $M < \infty$ such that

$$\max_{1 \le t \le T} \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t_j}^4 \le M T^{\epsilon}$$

(A2) $N = \min_{1 \le t \le T} N_t = T^{\gamma}$ for some $\gamma > 0$

(A3) For some $\delta > 0$, $\max\limits_{1 \le t \le T} \left| \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t_j}^2 - p \lim_{N_t \to \infty} \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t_j}^2 \right| = O_p(N^{-\delta})$

(A4) The stochastic process $\{\sigma_{t_j}^2(\eta_{t_j}^2 - 1)\}_{j=1,t=1}^{N_t,T}$ is a strictly stationary with finite k^{th} moment, for some $k > 3$, 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 $\max_{1 \le t \le T} m_t$ 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 $1928-2005$, daily excess market returns are highly leptokurtic with the degree of excess kurtosis being 22.88. 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 Hölder Continuous of order less than 1/2.

(iii) Condition (A4) ensures that the random variables $\{\sigma_{t_j}^2(\eta_{t_j}^2-1)\}$, 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)) to obtain the result.

LEMMA 1. *Suppose that Assumptions A1-A4 hold. Then for some* $\alpha>0$,

$$T^{\alpha}\max_{1\leqslant t\leqslant T}|\widehat{\upsilon}_t-\upsilon_t|=o_p(1). \tag{22}$$

We now turn to the main result of this section, the asymptotic distribution of the parameter estimator $\widehat{\theta}$. We define $G_T(\theta)\equiv T^{-1}\sum_{t=1}^{T}G\left(r_{m,t}-r_{f,t},\upsilon_t,z_{t-1};\theta\right)$ and the infeasible GMM estimator $\widetilde{\theta}$ that minimizes $||G_T(\theta)||_W$. For the asymptotic distributional results, let $\overline{G}(\theta)=E[G\left(r_{m,t}-r_{f,t},\upsilon_t,z_{t-1};\theta\right)]$ and define

$$\begin{aligned}\Gamma &\equiv \frac{\partial}{\partial\theta}\overline{G}(\theta_0)\\ \Omega &\equiv \operatorname{var}\left[\sqrt{T}G_T(\theta_0)\right].\end{aligned}$$

We now let θ_0 be the true value to distinguish from a generic value θ. Then, under suitable conditions, the infeasible GMM estimator, $\widetilde{\theta}$, satisfies

$$\sqrt{T}(\widetilde{\theta}-\theta_0)=-(\Gamma^{\top}W\Gamma)^{-1}\Gamma^{\top}W\sqrt{T}G_T(\theta_0)+o_p(1)\Longrightarrow N(0,\Sigma), \tag{23}$$

where $\Sigma=(\Gamma^{\top}W\Gamma)^{-1}\Gamma^{\top}W\Omega W\Gamma(\Gamma^{\top}W\Gamma)^{-1}$, (see Pakes and Pollard (1989)). This theory does not require $G\left(r_{m,t}-r_{f,t},\upsilon_t,z_{t-1};\theta\right)$ to be smooth in θ or $(r_{m,t}-r_{f,t},\,\upsilon_t,\,z_{t-1})$ 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. Furthermore, it is natural to suppose that the process $\{r_{m,t}-r_{f,t},\upsilon_t,z_{t-1}\}$ is stationary and weakly dependent, e.g., strong mixing, which would support the central limit theorem in (23).

We make some additional assumptions. Our theory parallels the work of Pakes and Pollard (1989) so we adopt their regularity conditions.

ASSUMPTIONS B

(B1) $\|G_T(\widehat{\theta}_T)\|_W=\inf_{\theta\in\Theta}\|G_T(\theta)\|_W+o_p(1/\sqrt{T})$;

(B2) For all $\delta > 0$, there exists $\epsilon(\delta)$ such that

$$\inf_{\|\theta-\theta_0\|>\delta} \|\overline{G}(\theta)\| \geq \epsilon(\delta) > 0.$$

The matrix

$$\Gamma(\theta) = \frac{\partial}{\partial\theta}\overline{G}(\theta)$$

is continuous in θ and is of full (column) rank at $\theta = \theta_0$.

(B3) The infeasible sample moment satisfies

$$\sup_{\theta\in\Theta} \|G_T(\theta) - \overline{G}(\theta)\|_W = o_p(1);$$

For all sequences of positive numbers δ_T such that $\delta_T \to 0$,

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

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

(B4) $\sqrt{T}G_T(\theta_0) \Longrightarrow \quad N(0,\Omega)$

(B5) θ_0 is in the interior of Θ.

(B6) $\sup\limits_{T\geq 1}\frac{1}{T}\sum_{t=1}^{T} E\left\|G(r_{m,t} - r_{f,t}, \upsilon_t, z_{t-1}, \theta_0)\right\|^{2+\lambda} < \infty$ for some $\lambda > 0$.

(B7) The first three partial derivatives of G with respect to θ and υ_t exist and satisfy dominance conditions, namely for all vectors ν (pertaining to $(\upsilon_t, \upsilon_{t-1}, \theta)$) with $|\nu| \leq 3$, and for some sequence $\delta_T \to 0$,

$$\sup_{|x|,|x'|\leq\delta_T} \sup_{\theta\in\Theta} \|D^\nu G(r_{m,t} - r_{f,t}, \upsilon_t + x, \upsilon_{t-1} + x', \theta)\| \leq U_t,$$

where $EU_t \leq M$ for some $M < \infty$.

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 and holds in our case 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 3 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 because $G(r_{m,t}-r_{f,t}, v_t, z_{t-1}, \theta)$ is a martingale difference sequence and Assumption B6 (See Hansen and Hodrick (1980)). Condition 7 is a smoothness condition on $G(.)$. Note that the asymptotic derivations in Pakes and Pollard (1989) do not require $G\left(r_{m,t}-r_{f,t}, v_t, z_{t-1}; \theta\right)$ to be smooth in θ or $(r_{m,t}-r_{f,t},\ v_t,\ z_{t-1})$ 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 $\widehat{\theta}$. We define

$$\begin{aligned} b_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}E\left[G_{v_tv_t}\left(r_{m,t}-r_{f,t}, v_t, z_{t-1}; \theta\right) IQ^t\right] \\ &\quad +\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}E\left[G_{v_{t-1}v_{t-1}}\left(r_{m,t}-r_{f,t}, v_t, z_{t-1}; \theta\right) IQ^{t-1}\right], \end{aligned}$$

where $G_{v_tv_t}$ denotes the second partial derivative of G with respect to v_t, and IQ^t is 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,*

$$\widehat{\theta}-\theta_0 = -(\Gamma^\top W\Gamma)^{-1}\Gamma^\top W G_T(\theta_0) - (\Gamma^\top W\Gamma)^{-1}\Gamma^\top W b_T(\theta_0) + o_p(T^{-1/2}), \tag{24}$$

Corollary 1. *When $b_T(\theta_0) = o(T^{-1/2})$, which requires that $N^{1.5}/T \to \infty$,*

$$\sqrt{T}(\widehat{\theta}-\theta_0) \Longrightarrow N(0,\Sigma). \tag{25}$$

When (25) holds, standard inference can be applied. Specifically, since $G(r_{m,t}-r_{f,t}, v_t, z_{t-1}, \theta)$ is a martingale difference sequence, $\widehat{\Sigma} = (\widehat{\Gamma}^\top W\widehat{\Gamma})^{-1}\widehat{\Gamma}^\top W\widehat{\Omega}W\widehat{\Gamma}(\widehat{\Gamma}^\top W\widehat{\Gamma})^{-1}$ is a consistent estimator of Σ, where

$$\begin{aligned} \widehat{\Gamma} &= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\theta}G(r_{m,t}-r_{f,t}, \widehat{v}_t, z_{t-1}, \widehat{\theta}) \\ \widehat{\Omega} &= \frac{1}{T}\sum_{t=1}^{T}G(r_{m,t}-r_{f,t}, \widehat{v}_t, z_{t-1}, \widehat{\theta})G(r_{m,t}-r_{f,t}, \widehat{v}_t, z_{t-1}, \widehat{\theta})^\top. \end{aligned}$$

The condition $b_T(\theta_0) = o(T^{-1/2})$ requires that $N^{1.5}/T \to \infty$. When this condition is not satisfied, (25) does not hold. In this case, define the bias corrected estimator

$$\widehat{\theta}^{bc} = \widehat{\theta} + (\Gamma^\top W\Gamma)^{-1}\Gamma^\top W\widehat{b}_T(\widehat{\theta}),$$

where

$$\widehat{b}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}G_{v_t v_t}(r_{m,t}-r_{f,t},\widehat{v}_t,z_{t-1},\theta)\widehat{IQ}^{t} + \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}G_{v_{t-1}v_{t-1}}(r_{m,t}-r_{f,t},\widehat{v}_t,z_{t-1},\theta)\widehat{IQ}^{t-1}, \tag{26}$$

where $\widehat{IQ}^{t} = \frac{N_t}{3}\sum_{j=1}^{N_t} r_{t_j}^4$ is a consistent estimator of the integrated quarticity.

Corollary 2. *Suppose that conditions A and B are satisfied and that $N^{\zeta}/T \to \infty$, $\zeta > 3$. Then,*

$$\sqrt{T}(\widehat{\theta}^{bc} - \theta_0) \Longrightarrow N(0,\Sigma).$$

This result is the basis of the application we conduct in the next section. In particular, it provides the basis for confidence intervals and test statistics regarding θ, and provides the methodology to take into account the potential consequences of small intraperiod samples. We have defined the methodology for general moment conditions but we next apply it to the case discussed in (20).

5 Empirical Results

5.1 Data

In our empirical analysis, we focus on the risk-return relation at the monthly, quarterly, semiannual, and annual 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 proxy for the risk free rate is the one-month Treasury Bill rate (from Ibbotson Associates). The sample extends from January 1928 - December 2005. 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, and the quarterly, semi-annual and annual returns and realized market variances are computed analogously. The monthly excess market return is the difference between the monthly market return and the monthly risk free rate, and so on.

To set the stage, Table 1 reports summary statistics for the excess returns and the corresponding realized variances for the different horizons. The table reports results for the full sample and for two subsamples of equal length. The monthly excess market return has a mean of 0.5% and a variance of 0.3% in the full sample. Returns are slightly negatively skewed and leptokurtic with the coefficients of skewness and kurtosis being -0.477 and 9.819, respectively, over the full sample.

The first order autocorrelation coefficient of monthly returns is 0.102. The average market return during $1928:01-1966:12$ is higher than that observed during $1967:01-2005:12$ (0.6% vs. 0.4%). The variance of monthly returns is also higher in the first subsample (0.4% vs. 0.2%). Both subsamples exhibit negative skewness and high kurtosis. The realized variance has a mean of 0.2% in the overall sample, which closely matches the variance of monthly returns over the same period. The mean of the variance in the first subsample is higher than in the second (0.3% vs. 0.2%), mostly because of the period of the Great Depression. The realized variance process displays considerable persistence, with an autocorrelation coefficient of 0.563 in the entire sample and has a much smaller variance compared to monthly excess returns (2.2×10^{-5} vs. 0.003). The first subsample shows more persistence in the variance process (0.657 vs. 0.226). As expected, realized variance is highly skewed and leptokurtic. Most of these characteristics of returns and realized variance persist at all the other horizons considered. The coefficient of kurtosis in excess returns and realized variances declines with the horizon (9.819 in monthly data vs. 5.618 in semi-annual data vs. 3.597 in annual data for returns and 58.47 vs. 14.36 vs. 13.78 for realized variances). The degree of skewness in the realized variance also declines with the horizon (6.317 in monthly data vs. 3.281 in semi-annual data vs. 3.102 in annual data) whereas no such trend is noticed in the coefficient of skewness 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 $2-4$. Over the sample period, the daily excess market returns are highly leptokurtic with the degree of excess kurtosis being 22.88. The evidence of very fat tails in the distribution of market returns highlights the importance of Assumption (a) in Lemma 1, which allows the volatility process to grow over time rather than restricting it to be uniformly bounded over all t and j.

Also, note that the conditional moment restriction in (16) is obtained by arguing that the integrated variance is approximately unbiased for the conditional variance. As pointed out in (6), the approximations are 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, equation (6) implies that the conditional variance of the market portfolio 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 0.0017 (0.0112) in our sample period. The second term is the conditional variance of the mean process. The squared mean of market excess returns is 1.4×10^{-5} (0.0001) 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 monthly excess market returns and the realized market variance at the monthly (quarterly) frequency is -4.5×10^{-5} (0.0071). Thus, the latter two terms in equation (6) are of smaller order than the first term lending support to the approximate unbiasedness of the integrated variance for the conditional variance in equation (16).

Next, we turn to our main empirical results. The analysis in Section 3 shows that the estimation of the risk-return trade-off parameters can be posed as a GMM estimation problem, with the following moment specification,

$$E\left[G\left(r_{m,t} - r_{f,t}, \widehat{v}_t, z_{t-1}; \theta\right)\right] = 0, \tag{27}$$

where $G\left(r_{m,t} - r_{f,t}, \widehat{v}_t, z_{t-1}; \theta\right) = \left(r_{m,t} - r_{f,t} - \alpha + \beta \widehat{v}_t\right) \otimes z_{t-1}$, $\theta = (\alpha, \beta)^\top$, and z_{t-1} is a vector of instruments. Table 2 reports results for the exactly identified case using the lagged notional (or integrated) volatility as an instrument and Table 3 reports results for an overidentified case where the first three lags of the notional volatility are used as instruments. Note that for these specifications of the moment restrictions and choice of instruments, the bias-correction is identically zero (see Theorem 6). Once again, results are reported for the full sample and two subsamples of equal length. Table 2 reveals a weak and statistically insignificant relation between the risk and the return. For monthly data, the slope coefficient is negative in the full sample as well as the subsamples but not statistically significant. This is consistent with the findings of French, Schwert, and Stambaugh (1987) and Whitelaw (1994). For lower frequency horizons, the estimated coefficients are mostly positive but not statistically significant. Table 3 confirms the findings in Table 2.

The rationale for using lagged integrated variance as an instrument in Tables 2 and 3 is that it is a highly persistent process. The first order autocorrelation coefficient of the realized variance process is 0.563, 0.554, 0.675, and 0.675, respectively, in monthly, quarterly, semiannual, and annual data for the full sample. Hence, the lagged variance is useful in predicting the contemporaneous variance which enters the moment specification (16). This makes it a good choice of instrument improving the efficiency of the estimation procedure.[5]

For additional robustness, we repeat the estimation using lagged (instead of contemporaneous) integrated variance in the moment restriction (16). This is justified under a martingale assumption on the conditional variance process and has been frequently employed in the literature. Table 4

[5] For robustness, we repeated the estimation for choice of instruments other than the lagged variance. In particular, we consider financial variables that are known to predict the mean returns. Examples include the dividend yield, the default spread and the interest rate. The results reveal a statistically insignificant relation, over the full sample as well as the subsamples, that is robust to the choice of instruments.

reports results for the exactly identified case using the lagged integrated variance as an instrument and Table 5 reports results for an overidentified case where the first three lags of the integrated variance are used as instruments. These tables confirm the findings in Tables 2 and 3. The estimated slope coefficient is negative at the monthly frequency and mostly positive in the lower frequency horizons and never statistically significant.

5.3 Time-Variation in the Risk-Return Tradeoff

To interpret the results, we turn to a closely related literature on return predictability that has reported evidence in favour of structural breaks in the OLS coefficient in the forecasting regression of returns on the lagged price-dividend ratio (e.g., Viceira (1996) and Paye and Timmermann (2006)). This renders the forecasting relationship unstable if such shifts are not taken into account. In particular, Lettau and Nieuwerburgh (2008) find evidence for two breaks in the mean of the log dividend-price ratio around 1954 and 1994. They demonstrate that if these breaks are ignored, the estimated OLS coefficient appears statistically insignificant over the full sample. However, when the sample is split into subsamples corresponding to the break dates, significant coefficient estimates are obtained in each subsample. These results suggest that if the relationship between expected returns and the conditional variance exhibits significant time variation, this could potentially render the estimated coefficient statistically insignificant when estimated over the entire sample.

Motivated by the above consideration, we split the sample into three subsamples based on the break dates in the mean of the log dividend-price ratio identified in Lettau and Nieuwerburgh (2008). Table 6 reports summary statistics for the excess returns and the corresponding realized variances for the different horizons for the three subsamples. Excess returns and realized variances have a lower mean and substantially lower realized variances in the second subsample covering January 1955 to December 1994 compared to the other subsamples (mean return of 0.4% in the second subsample vs. 0.6% in the first and last subsamples in monthly data, and mean realized variance of 0.1% vs. 0.4% and 0.2%). The mean and the variance are highest in the first subsample that mainly reflects the Great Depression. The realized variance is considerably less persistent over the second subsample (0.154 vs. 0.639 and 0.589, in monthly data).

Tables 7 and 8 report estimation results for the same specification of the moment conditions as Tables 2 and 3 but with different choice of subsamples. As in Tables 2 and 3, the estimated slope coefficient is mostly negative and statistically insignificantly different from zero in monthly data. However, the results change dramatically for the longer horizons. A close inspection of the tables reveals substantial time-variation in the risk-return tradeoff. In particular, the relationship appears quite unstable over the first subsample. The estimate changes sign from negative to positive and then becomes negative again as we move from quarterly to annual horizon in Table 7. However, the estimated coefficient is significantly positive in the second subsample covering $1954 - 1994$ in quarterly and semi-annual data. It is also positive at the annual horizon but falls short of being significantly so. This is primarily a reflection of the small number of observations, (namely, 40), used in the estimation. Finally, the estimates for the third subsample are always negative in quarterly, semi-annual, and annual data. They are statistically significant at the semi-annual horizon in Table 8 and at the annual horizon in Tables 7 and 8.

Finally, Tables 9 and 10 report estimation results for the same specification of the moment conditions as Tables 4 and 5, i.e. using lagged (instead of contemporaneous) integrated variance, but with the intra-break subperiods. The results are largely similar to those obtained in Tables 7 and 8.

6 Simulation Results

In this section, we perform Monte Carlo simulations in order to examine the finite-sample performance of the estimators. We assume that the continuously compounded returns on the market are generated by a diffusion process of the form (2)

$$dp(t) = \mu(t)dt + \sigma(t)dW_1(t). \tag{28}$$

Note that our nonparametric estimation approach, in Sections 3 and 4, does not require us to specify the form of the drift, $\mu(t)$, or diffusion terms, $\sigma(t)$, in (28), 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, Andersen, Benzoni, and Lund (2002))

$$d\log\sigma^2(t) = -0.0136\left(a_2 + \log\sigma^2(t)\right)dt + 0.1148dW_2(t). \tag{29}$$

The second model is the $GARCH(1,1)$ diffusion (see, Andersen and Bollerslev (1998))

$$d\sigma^2(t) = 0.035\left(a_3 - \sigma^2(t)\right)dt + 0.144\sigma^2(t)dW_3(t). \tag{30}$$

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 model for $\mu(t)$ is motivated by the time-invariant, linear relation between the conditional expected excess return of the market portfolio and its conditional variance considered in this paper.

Hence, we consider the following model for $\mu(t)$

$$\mu(t) = \alpha + \beta\sigma^2(t). \tag{31}$$

Note that time-aggregating (31) over the interval $[t-1, t]$ and taking conditional expectations of both sides with respect to time $t-1$ information set delivers the conditional moment restriction (16). In the simulations, we set $\alpha = 0$ and try a few different values for β, namely, $\beta = 2$, 5, -2, and -5, in order to examine the size and power of the estimation approach.

Finally, as in the empirical application in Section 5, we assume that the above specifications of the diffusion process generates high frequency (daily) returns on the market. The parameters a_2 and a_3 in (29) and (30) 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. 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, and the quarterly, semi-annual and annual returns and realized market variances are calculated analogously. These are then used in the GMM estimation problem (27), with the lagged realized variance being used as an instrument, to estimate the parameter vector, $\theta = (\alpha, \beta)^\top$. This procedure is repeated 500 times.

Table 11 reports the simulation results for the lognormal model for the diffusion term, $\sigma(t)$. Panel A of Table 11 reports results for $\alpha = 0$ and $\beta = 5$ while Panel B does the same for $\alpha = 0$ and $\beta = 2$. The first row of Panel A corresponds to $N = 22$ high frequency data within each of $T = 936$ time periods. This choice corresponds to the historically available daily data within monthly time periods that we use in our estimation in Section 5. The second and third columns of the table report the mean of the estimates of the intercept, α, and slope coefficient, β, respectively, the standard errors of the estimates in parentheses, and the 90% confidence intervals in square brackets, across the 500 simulations. The first row of Panel A reveals that the mean α across simulations is 0.000 with a standard error of 0.004. The mean β across simulations is 4.957 with a standard error of 1.576. The 90% confidence interval for β lies entirely in the positive axis. Note that the point estimate of β obtained in the historical sample at the monthly horizon is -0.493 in Table 1.2. The last column of Table 11 reports the probability of observing an estimate of β at least as small as the one obtained in the historical sample if the true data generating process is described by equations (28), (29), and (31). In other words, this gives the probability of observing a negative relation between the equity premium and the conditional variance of the market return of the magnitude found in the data if the market returns were generated by a diffusion process with a positive relation ($\beta = 5$) between its conditional mean and variance. The first row suggests that this probability is miniscule at 0.0%.

The values for N in the second, third, and fourth rows are chosen to correspond to available daily data within quarterly, semi-annual, and annual time periods, respectively, and T to correspond to available quarterly, semi-annual, and annual time periods, respectively. These rows reveal that the mean of α is -0.001, 0.001, and 0.002, and that of β is 5.11, 4.92, and 4.95, respectively, for these choices of N and T. The standard errors increase with decrease in T. The standard error of α increases from 0.004 for $T = 936$ to 0.014 for $T = 312$ to 0.034 for $T = 156$ to 0.103 for $T = 78$, while that for β increases from 1.576 for $T = 936$ to 1.873 for $T = 312$ to 2.286 for $T = 156$ to 3.541 for $T = 78$. However, the mean estimate of β is statistically significantly positive in all four rows on Panel A. Moreover, even for $T = 312$ and 156, the 90% confidence intervals for β lies entirely in the positive axis. These results suggest that if there were indeed a positive relation between the conditional mean and variance, our estimation results would capture it.

Table 11, Panel B reports the same results as Panel A but for $\alpha = 0$ and $\beta = 2$. The first row of this panel reveals that the mean β for $N = 22$ and $T = 936$ is 2.032 with standard error of 1.631. While the 90% confidence interval for β includes points in the negative axis, the probability of observing a point estimate at least as small as the value observed in the data is only 6.2%. The simulation results in this Panel get weaker with decrease in T. Rows 2, 3, and 4 in this Panel show that lowering T from 312 to 156 to 78 raises the standard errors of the α and β estimates and the probability of observing an estimate of β as small as or smaller than the value obtained in the historical sample rises to 16.8%, 20.4%, and 18.2%, respectively.

Table 12 reports the simulation results for the $GARCH(1,1)$ model for the diffusion term, $\sigma(t)$. Panel A reports results for $\alpha = 0$ and $\beta = 5$ while Panel B does the same for $\alpha = 0$ and $\beta = 2$. The results are largely similar to those in Table 11. Panel A shows that lowering T from 936 to 312 to 156 to 78 raises the standard errors of the α and β estimators. However, the 90% confidence intervals reveal that the lower limits of these intervals are bigger than the point estimates of β obtained in the historical sample for all four horizons considered (see Table 2). Moreover, the last column suggests that the probability of observing a point estimate of β at least as small as the value observed in the historical sample at the corresponding horizons are very small at 0.4%, 2.4%, 2.2%, and 4.2%, respectively. The simulation results are weaker in Panel B that corresponds to $\alpha = 0$ and $\beta = 2$. The lower limits of the 90% confidence intervals are smaller than the point estimates of β obtained in the historical sample for all four horizons considered (see Table 2). The last column suggests that the probability of observing a point estimate of β at least as small as the value observed in the historical sample at the corresponding horizons are bigger at 13.6%, 22.4%, 25.8%, and 22.0%, respectively.

7 Conclusion and Extensions

This paper proposes an approach to estimating the risk-return tradeoff in the stock market that allows us to escape some of the limitations of existing empirical analyses. First, it does not require any specific functional form assumptions, either about the conditional mean or the conditional variance. We focus on a nonparametric measure of expost return variability, namely integrated variance, that is approximately unbiased for the conditional variance. This latent variance measure may be consistently and accurately estimated using the so-called realized variance, which is easily computed from high frequency intra-period returns. Second, we estimate the risk-return trade-off parameters using the Generalized Method of Moments (GMM) approach. The unbiasedness property of the integrated variance provides a moment restriction under the null hypothesis of a linear relation between the conditional expected excess returns of the stock market and its conditional variance. This approach overcomes the endogeneity problem inherent in a least squares regression of an estimate of the conditional mean on the estimate of the conditional variance as both these quantities are simultaneously determined. Third, we offer a solution to the measurement error problem that arises because of the use of a proxy for the variance (realized variance in place of integrated variance), an issue that has thus far been ignored in the literature.

The results indicate a weak, statistically insignificant relation between the conditional mean and the conditional variance of the stock market return. This finding is robust across different return horizons (monthly, quarterly, semiannual, and annual) and choice of instruments. However, when the sample is split into three subsamples based on the break dates in the mean of the log price-dividend ratio of the market, as identified in Lettau and Nieuwerburgh (2008), significant coefficient estimates, albeit opposite in sign, are obtained in the latter two subsamples. The relation appears quite flat and unstable in the first subsample that covers the period of the Great Depression and World War II. These results are suggestive of significant time-variation in the risk-return relation.

Finally, the econometric framework developed in this paper is quite general with several other potential applications in asset pricing. 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)). 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

Abel, A. B. (1988): "Stock Prices Under Time-Varying Dividend Risk: An Exact Solution in an Infinite Horizon General Equilibrium Model," *Journal of Monetary Economics*, 22, 375–393.

Andersen, T., L. Benzoni, and J. Lund (2002): "An Empirical Investigation of Continuous Time Equity Return Models," *Journal of Finance*, 57, 1239–1284.

Andersen, T., and T. Bollerslev (1998): "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, 885–905.

Andersen, T., T. Bollerslev, and F. Diebold (2002): "Parametric and Nonparametric Volatility Measurement," in *Handbook of Financial Econometrics (forthcoming)*, ed. by Y. Ait-Sahalia, and L. Hansen. Holland.

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

Back, K. (1991): "Asset Pricing for General Processes," *Journal of Mathematical Economics*, 20, 371–395.

Backus, K., and A. Gregory (1993): "Theoretical Relations Between Risk Premiums and Conditional Variances," *Journal of Business and Economic Statistics*, 11, 177–185.

Bansal, R., and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, pp. 1481–1509.

Barndorff-Nielsen, O., and N. Shephard (2002): "Econometric Analysis of Realized Volatility and its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society*.

——— (2004): "Econometrics of testing for jumps in financial economics using bipower variation," *Journal of Financial Econometrics*, 2, 1–37.

Bosq, D. (1998): *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, Lecture Notes in Statistics. Springer-Verlag, New York.

Campbell, J., and L. Hentschel (1992): "No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns," *Journal of Financial Economics*, 31(3), 281–318.

Campbell, J. Y. (1987): "Stock Returns and the Term Structure," *Journal of Financial Economics*, 18(2), 373–99.

Campbell, J. Y., and J. H. Cochrane (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107(2), 205–51.

Constantinides, G., and A. Ghosh (2008): "Asset Pricing Tests with Long Run Risks in Consumption Growth," NBER Working Paper.

Corradi, V., and W. Distaso (2006): "Semiparametric Comparison of Stochastic Volatility Models via Realized Measure," *Review of Economic Studies*, 73, 635–677.

Corradi, V., W. Distaso, and N. Swanson (forthcoming): "Predictive Density Estimators for Daily Volatility Based on the Use of Realized Measures," *Journal of Econometrics*.

French, K., W. Schwert, and R. Stambaugh (1987): "Expected Stock Returns and Volatility,," *Journal of Financial Economics*, 19(1), 3–29.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2005): "There is a Risk-Return Trade-Off After All," *Journal of Financial Economics*, 76, 509–548.

Glosten, L., R. Jagannathan, and D. Runkle (1993): "On The Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–1801.

Gonçalves, S., and N. Meddahi (2009): "Bootstrapping Realized Volatility," *Econometrica*, 77, 283–306.

Guo, H., and R. Whitelaw (2006): "Uncovering the Risk-Return Relation in the Stock Market," *Journal of Finance*, 61, 1433–1463.

Hansen, L., and S. Richard (1987): "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," *Econometrica*, 55, 587–614.

Hansen, L. P., and R. J. Hodrick (1980): "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy*, 88, 829–853.

Harvey, C. (2001): "The Specification of Conditional Expectations," *Journal of Empirical Finance*, 8, 573–638.

Jacod, J. (1994): "Limit of Random Measures Associated with the Increments of a Brownian Semimartingale," Preprint Number 120, Laboratoire de Probabilities, Universite Pierre et Marie Curie, Paris.

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

Judd, K. L. (1998): *Numerical Methods in Economics*. Cambridge, MA, M.I.T. Press.

Lettau, M., and S. Ludvigson (2003): "Measuring and Modeling Variation in the Risk-Return Trade-Off," in *Handbook of Financial Econometrics*, ed. by Y. Ait-Sahalia, and L. Hansen. North-Holland, Holland.

Lettau, M., and S. V. Nieuwerburgh (2008): "Reconciling the Return Predictability Evidence," *Review of Financial Studies*, 21.

Linton, O., and B. Perron (2003): "The shape of the risk premium: Evidence from a semiparametric-mean GARCH model," *Journal of Business and Economic Statistics*, pp. 354–367.

Ludvigson, S., and S. Ng (2007): "The Empirical Risk-Return Relation: A Factor Analysis Approach," *Journal of Financial Economics*, 83, 171–222.

Lundblad, C. (2005): "The Risk-Return Tradeoff in the Long Run: 1836-2003," *Journal of Financial Economics*, 85, 123–150.

Merton, R. C. (1973): "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867–86.

Nelson, D. (1991): "Conditional Heteroskedasticity in Asset Returns: A New Approach," 59(2), 347–370.

Newey, W. K., and D. McFadden (1994): "Large sample estimation and hypothesis testing," in *Handbook of Econometrics*, ed. by R. F.Engle, and D. McFadden, vol. 4. Elsevier Press.

Pakes, A., and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.

Pastor, L., M. Sinha, and B. Swaminathan (2008): "Estimating the Intertemporal Risk-Return Tradeoff Using the Implied Cost of Capital," *Journal of Finance*, 63, 2859–2897.

Paye, B., and A. Timmermann (2006): "Instability of Return Prediction Models," *Journal of Empirical Finance*, 13(3).

Protter, P. (2004): *Stochastic Integration and Differential Equations*. Springer-Verlag, New York.

Scruggs, J. (1998): "Resolving the Puzzling Intertemporal Relation Between the Market Risk Premium and the Conditional Market Variance: A Two-Factor Approach," *Journal of Finance*, 53, 575–603.

Scruggs, J., and P. Glabadanidis (2003): "Risk Premia and the Dynamic Covariance Between Stock and Bond Returns," *Journal of Financial and Quantitative Analysis*, 37, 27–62.

Viceira, L. (1996): "Testing for Structural Change in the Predictability of Asset Returns," Unpublished Manuscript, Harvard University.

Whitelaw, R. (1994): "Time Variations and Covariations in the Expectation and Volatility of Stock Market Returns," *Journal of Finance*, 49, 515–541.

——— (2000): "Stock Market Risk and Return: An Equilibrium Approach," *Review of Financial Studies*, 13, 521–547.

A Appendix

In what follows, we define $\sigma_t^2 \equiv \upsilon_t$.

A.1 Proof of Lemma 1

We have

$$\widehat{\sigma}_t^2 = \sum_{j=1}^{N_t} r_{t_j}^2 = \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t_j}^2 (\eta_{t_j}^2 - 1) + \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t_j}^2 .$$

Hence,

$$\widehat{\sigma}_t^2 - \sigma_t^2 = \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t_j}^2 (\eta_{t_j}^2 - 1) + \left[\frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t_j}^2 - p \lim_{N_t \to \infty} \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t_j}^2 \right] . \tag{32}$$

Consider the first term in equation (32):

$$E\left[\frac{1}{N_t}\sum_{j=1}^{N_t}\sigma_{t_j}^2(\eta_{t_j}^2-1)\right]=\frac{1}{N_t}\sum_{j=1}^{N_t}E\left[E\left\{\sigma_{t_j}^2(\eta_{t_j}^2-1)\right\}|\mathcal{F}_{t_{j-1}}\right]=0$$

$$\begin{aligned}\operatorname{var}\left[\frac{1}{N_t}\sum_{j=1}^{N_t}\sigma_{t_j}^2(\eta_{t_j}^2-1)\right] &= \frac{1}{N_t^2}\sum_{j=1}^{N_t}E\left[\sigma_{t_j}^4(\eta_{t_j}^2-1)^2\right] \\ &\quad +\frac{1}{N_t^2}\sum_{j=1}^{N_t}\sum_{\substack{k=1\\ j\neq k}}^{N_t}E\left[\sigma_{t_j}^2\sigma_{t_k}^2(\eta_{t_j}^2-1)(\eta_{t_k}^2-1)\right] \\ &= \frac{2}{N_t^2}\sum_{j=1}^{N_t}E\left[\sigma_{t_j}^4\right]\leqslant\frac{2}{N_t}MT^{\epsilon}=O\left(\frac{T^{\epsilon}}{N_t}\right)=o(1),\end{aligned}$$

provided $\gamma>\epsilon$, by Assumption A1. Therefore,

$$\frac{1}{N_t}\sum_{j=1}^{N_t}\sigma_{t_j}^2(\eta_{t_j}^2-1)=O_p\left(\sqrt{\frac{T^{\epsilon}}{N_t}}\right).$$

Thus, the second term in equation (32) is of smaller order than the first provided $\delta>\frac{1}{2}(1-\frac{\epsilon}{\gamma})$, and we have,

$$\widehat{\sigma}_t^2-\sigma_t^2\approx\frac{1}{N_t}\sum_{j=1}^{N_t}\sigma_{t_j}^2(\eta_{t_j}^2-1).$$

Now, by the Bonferroni inequality

$$\begin{aligned}\Pr\left[\max_{1\leq t\leq T}|\widehat{\sigma}_t^2-\sigma_t^2|>\delta\right] &\leq \sum_{t=1}^{T}\Pr\left[\left|\frac{1}{N_t}\sum_{j=1}^{N_t}\sigma_{t_j}^2(\eta_{t_j}^2-1)\right|>\delta\right]. \\ &= \sum_{t=1}^{T}\Pr\left[\left|\frac{1}{N_t}\sum_{j=1}^{N_t}\sigma_{t_j}^2(\eta_{t_j}^2-1)\right|>\delta,\ \max_{\substack{1\leqslant t\leqslant T\\ 1\leqslant j\leqslant N_t}}\sigma_{t_j}^2<N_t\right]+ \\ &\quad \sum_{t=1}^{T}\Pr\left[\left|\frac{1}{N_t}\sum_{j=1}^{N_t}\sigma_{t_j}^2(\eta_{t_j}^2-1)\right|>\delta,\ \max_{\substack{1\leqslant t\leqslant T\\ 1\leqslant j\leqslant N_t}}\sigma_{t_j}^2\geqslant N_t\right].\end{aligned}$$

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

$$
\begin{aligned}
\Pr\left[\left|\frac{1}{N_t}\sum_{j=1}^{N_t}\sigma_{t_j}^2(\eta_{t_j}^2-1)\right|>\delta\right] &= \Pr\left[\left|\sum_{j=1}^{N_t}\sigma_{t_j}^2(\eta_{t_j}^2-1)\right|>N_t\delta\right] \\
&\leqslant a_1\exp\left(-\frac{q\delta^2}{25m_2^2+5c\delta}\right)+a_2(k)\alpha\left(\left[\frac{N_t}{q+1}\right]\right)^{\frac{2k}{2k+1}},
\end{aligned}
$$

where:

$$
a_1 = 2\frac{N_t}{q}+2\left(1+\frac{\delta^2}{25m_2^2+5c\delta}\right), \quad \textit{with } m_2^2=\max_{1\leqslant j\leqslant N_t} E\left[\sigma_{t_j}^2(\eta_{t_j}^2-1)\right]
$$

$$
a_2(k) = 11N_t\left(1+\frac{5m_k^{\frac{k}{2k+1}}}{\delta}\right), \quad \textit{with } m_k=\max_{1\leqslant j\leqslant N_t}\left\|\sigma_{t_j}^2(\eta_{t_j}^2-1)\right\|_k
$$

for each $N_t \geqslant 2$, each integer $q \in \left[1, \frac{N_t}{2}\right]$, each $\delta > 0$, and each $k \geqslant 3$. $c > 0$ depends on the distribution of the time series.

Assuming $N_t = N$ for all t, and given assumption 4, we have

$$
\begin{aligned}
&\sum_{t=1}^{T}\Pr\left[\left|\frac{1}{N}\sum_{j=1}^{N_t}\sigma_{t_j}^2(\eta_{t_j}^2-1)\right|>\delta\right] \\
\leqslant\ & Ta_1\exp\left(-\frac{q\delta^2}{25m_2^2+5c\delta}\right)+Ta_2(k)\alpha\left(\left[\frac{N}{q+1}\right]\right)^{\frac{2k}{2k+1}} \\
\to\ & 2T\left(\frac{N}{q}+1\right)\exp\left(-\frac{q\delta^2}{25m_2^2}\right)+11NT\left(1+\frac{5m_k^{\frac{k}{2k+1}}}{\delta}\right)\exp\left(-\frac{c'N}{q+1}\frac{2k}{2k+1}\right)
\end{aligned}
$$

as $\delta \to 0$. Putting $\delta_T = \frac{T^\epsilon}{N^{1/2}} \to 0$ provided $\gamma > 2\epsilon$, and $q = N^\theta$, $\theta < 1$, we have

$$
\sum_{t=1}^{T}\Pr\left[\left|\frac{1}{N}\sum_{j=1}^{N_t}\sigma_{t_j}^2(\eta_{t_j}^2-1)\right|>\delta_T\right]\to 0,
$$

provided $\theta > 1 - \frac{2\epsilon}{\gamma}$.

Consider now the second term,

$$
\begin{aligned}
\sum_{t=1}^{T} \Pr\left[\left|\frac{1}{N}\sum_{j=1}^{N_t}\sigma_{t_j}^2(\eta_{t_j}^2-1)\right| > \delta,\ \max_{\substack{1\leqslant t\leqslant T\\ 1\leqslant j\leqslant N_t}} \sigma_{t_j}^2 \geqslant N\right] &\leqslant \sum_{t=1}^{T} \Pr\left[\max_{\substack{1\leqslant t\leqslant T\\ 1\leqslant j\leqslant N_t}} \sigma_{t_j}^2 \geqslant N\right] \\
&= T\Pr\left[\max_{\substack{1\leqslant t\leqslant T\\ 1\leqslant j\leqslant N_t}} \sigma_{t_j}^2 \geqslant N\right] \\
&\leqslant T\sum_{t=1}^{T}\sum_{j=1}^{N_t} \Pr\left[\sigma_{t_j}^2 \geqslant N\right] \\
&\leqslant T^2 N\frac{E\left(\sigma_{t_j}^{2k}\right)}{N^k} \\
&\to 0,
\end{aligned}
$$

provided $2+\gamma-\gamma k<0$, i.e., $\gamma > \frac{2}{k-1}$. Thus, in order to have $\gamma<1$, we require $k>3$.

Thus, it follows that

$$
\max_{1\leq t\leq T} |\widehat{\sigma}_t^2 - \sigma_t^2| = O_p\left(\frac{T^\epsilon}{N^{1/2}}\right).
$$

So, provided $\alpha+\epsilon < \frac{\gamma}{2}$, the result follows. ■

A.2 Proof of Theorem 1

A.2.1 Consistency of $\widehat{\theta}$

We just verify the ULLN condition. By the triangle inequality

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

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

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

By the Mean Value Theorem,

$$\begin{aligned}
\widehat{G}_T(\theta) - G_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T} G_{\sigma_t}(X_t, \overline{\sigma}_t^2, \overline{\sigma}_{t-1}^2, \overline{\sigma}_{t-2}^2, \theta)(\widehat{\sigma}_t^2 - \sigma_t^2) + \\
&\quad \frac{1}{T}\sum_{t=1}^{T} G_{\sigma_{t-1}}(X_t, \overline{\sigma}_t^2, \overline{\sigma}_{t-1}^2, \overline{\sigma}_{t-2}^2, \theta)(\widehat{\sigma}_{t-1}^2 - \sigma_{t-1}^2) + \\
&\quad \frac{1}{T}\sum_{t=1}^{T} G_{\sigma_{t-2}}(X_t, \overline{\sigma}_t^2, \overline{\sigma}_{t-1}^2, \overline{\sigma}_{t-2}^2, \theta)(\widehat{\sigma}_{t-2}^2 - \sigma_{t-2}^2),
\end{aligned}$$

where $\overline{\sigma}_t$ is intermediate between $\widehat{\sigma}_t^2$ and σ_t^2, and so on. Furthermore, on the set A_T,

$$\begin{aligned}
\sup_{\theta\in\Theta} \|\widehat{G}_T(\theta) - G_T(\theta)\|_W &\leqslant 3\sup_{\theta\in\Theta}\left\|\frac{1}{T}\sum_{t=1}^{T} G_{\sigma}(X_t, \overline{\sigma}_t^2, \overline{\sigma}_{t-1}^2, \overline{\sigma}_{t-2}^2, \theta)(\widehat{\sigma}_t^2 - \sigma_t^2)\right\|_W \\
&\leq 3\epsilon_T \frac{1}{T}\sum_{t=1}^{T} U_t = o_p(1).
\end{aligned}$$

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

A.2.2 Asymptotic Normality

Under our conditions

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

For the asymptotic expansion our proof parallels the work of Pakes and Pollard (1989). We expand the estimated moment condition out to third order

$$\begin{aligned}
\widehat{G}_T(\theta_0) - G_T(\theta_0) &= \frac{1}{T}\sum_{t=1}^{T} G_{\sigma_t}(X_t, \sigma_t^2, \sigma_{t-1}^2, \sigma_{t-2}^2, \theta_0)(\widehat{\sigma}_t^2 - \sigma_t^2) \quad [3] \qquad (33)\\
&+\frac{1}{2T}\sum_{t=1}^{T} G_{\sigma_t\sigma_t}(X_t, \sigma_t^2, \sigma_{t-1}^2, \sigma_{t-2}^2, \theta_0)(\widehat{\sigma}_t^2 - \sigma_t^2)^2 \quad [3] \\
&+\frac{1}{2T}\sum_{t=1}^{T} G_{\sigma_t\sigma_{t-1}}(X_t, \sigma_t^2, \sigma_{t-1}^2, \sigma_{t-2}^2, \theta_0)(\widehat{\sigma}_t^2 - \sigma_t^2)(\widehat{\sigma}_{t-1}^2 - \sigma_{t-1}^2) \quad [6] \\
&+\frac{1}{6T}\sum_{t=1}^{T} G_{\sigma_t\sigma_t\sigma_t}(X_t, \overline{\sigma}_t^2, \overline{\sigma}_{t-1}^2, \overline{\sigma}_{t-2}^2, \theta_0)(\widehat{\sigma}_t^2 - \sigma_t^2)^3 \quad [3] \\
&+\frac{1}{6T}\sum_{t=1}^{T} G_{\sigma_t\sigma_t\sigma_{t-1}}(X_t, \overline{\sigma}_t^2, \overline{\sigma}_{t-1}^2, \overline{\sigma}_{t-2}^2, \theta_0)(\widehat{\sigma}_t^2 - \sigma_t^2)^2\ (\widehat{\sigma}_{t-1}^2 - \sigma_{t-1}^2), \quad [24]
\end{aligned}$$

where $\overline{\sigma}_t^2$ is intermediate between $\widehat{\sigma}_t^2$ and σ_t^2 and so on. The symbol [3] indicates the sum of the term given plus 3 similar terms obtained via partial differentiation with respect to the other arguments.

Consider the first term in (33),

$$\frac{1}{T}\sum_{t=1}^{T} G_{\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)(\widehat{\sigma}_t^2-\sigma_t^2) = \frac{1}{T}\sum_{t=1}^{T} G_{\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\frac{1}{N}\sum_{j=1}^{N}\sigma_{t_j}^2(\eta_{t_j}^2-1)$$

We have

$$\begin{aligned} & E\left[\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{j=1}^{N} G_{\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\sigma_{t_j}^2(\eta_{t_j}^2-1)\right] \\ = \; & \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{j=1}^{N} E\left[G_{\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\right] E\left[\sigma_{t_j}^2(\eta_{t_j}^2-1)\right] \end{aligned}$$

by assumption 6.

Also,

$$\begin{aligned} & \operatorname{var}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{j=1}^{N} G_{\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\sigma_{t_j}^2(\eta_{t_j}^2-1)\right] \\ = \; & \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{j=1}^{N} E\left[G_{\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)^2\right] E\left[\sigma_{t_j}^4(\eta_{t_j}^2-1)^2\right] + \\ & \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{\substack{s=1\\ t\neq s}}^{T}\sum_{j=1}^{N}\sum_{\substack{k=1\\ j\neq k}}^{N} E\left[G_{\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)G_{\sigma_s}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\right] \\ & \times E\left[\sigma_{t_j}^2\sigma_{t_k}^2(\eta_{t_j}^2-1)(\eta_{t_k}^2-1)\right] \\ = \; & \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{j=1}^{N} E\left[G_{\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)^2\right] E\left[\sigma_{t_j}^4(\eta_{t_j}^2-1)^2\right] \\ = \; & 2E\left[G_{\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)^2\right]\frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{j=1}^{N} E\left[\sigma_{t_j}^4\right] \\ \leqslant \; & \frac{1}{NT}MT^{\epsilon} = O\left(\frac{T^{\epsilon}}{NT}\right) = o\left(\frac{1}{T}\right), \end{aligned}$$

provided $\gamma > \epsilon$.

Next, consider the second term in (33),

$$
\begin{aligned}
&\frac{1}{2T}\sum_{t=1}^{T} G_{\sigma_t\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)(\widehat{\sigma}_t^2-\sigma_t^2)^2 \\
=\ &\frac{1}{2T}\sum_{t=1}^{T} G_{\sigma_t\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\left[\frac{1}{N}\sum_{j=1}^{N}\sigma_{t_j}^2(\eta_{t_j}^2-1)\right]^2 \\
=\ &\frac{1}{2T}\sum_{t=1}^{T}\frac{1}{N^2}\sum_{j=1}^{N} G_{\sigma_t\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\sigma_{t_j}^4(\eta_{t_j}^2-1)^2 \\
&+\frac{1}{2T}\sum_{t=1}^{T}\frac{1}{N^2}\sum_{j=1}^{N}\sum_{\substack{k=1\\ j\neq k}}^{N} G_{\sigma_t\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\sigma_{t_j}^2\sigma_{t_k}^2(\eta_{t_j}^2-1)(\eta_{t_k}^2-1).
\end{aligned}
$$

Hence

$$
\begin{aligned}
&E\left[\frac{1}{2T}\sum_{t=1}^{T} G_{\sigma_t\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)(\widehat{\sigma}_t^2-\sigma_t^2)^2\right] \\
=\ &\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N^2}\sum_{j=1}^{N} E\left[G_{\sigma_t\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\right]E\left[\sigma_{t_j}^4\right] \\
\simeq\ &\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}E\left[G_{\sigma_t\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\right]E\left[IQ^t\right]
\end{aligned}
$$

and

$$\begin{aligned}
& \operatorname{var}\left[\frac{1}{2T}\sum_{t=1}^{T} G_{\sigma_t\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)(\widehat{\sigma}_t^2-\sigma_t^2)^2\right] \\
= & \frac{1}{4T^2N^4}\sum_{t=1}^{T}\sum_{j=1}^{N} E\left[G_{\sigma_t\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\right]^2 E\left[\sigma_{t_j}^8(\eta_{t_j}^2-1)^4\right] \\
& +\frac{1}{4T^2N^4}\sum_{t=1}^{T}\sum_{\substack{s=1\\ t\neq s}}^{T}\sum_{j=1}^{N}\sum_{\substack{k=1\\ j\neq k}}^{N} E\left[G_{\sigma_t\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)G_{\sigma_s\sigma_s}(X_t,\sigma_s^2,\sigma_{s-1}^2,\sigma_{s-2}^2,\theta_0)\right] \\
& \times E\left[\sigma_{t_j}^4\sigma_{s_k}^4(\eta_{t_j}^2-1)^2(\eta_{s_k}^2-1)^2\right] \\
& -\left[\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}E\left[G_{\sigma_t\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\right]E\left[IQ^t\right]\right]^2 \\
& +\frac{1}{4T^2N^4}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{\substack{k=1\\ j\neq k}}^{N} E\left[G_{\sigma_t\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\right]^2 E\left[\sigma_{t_j}^4\sigma_{t_k}^4(\eta_{t_j}^2-1)^2(\eta_{t_k}^2-1)^2\right] \\
\to & \; 0,
\end{aligned}$$

provided $\gamma > \epsilon$.

Next, consider the third term in (33),

$$\begin{aligned}
& \frac{1}{2T}\sum_{t=1}^{T} G_{\sigma_t\sigma_{t-1}}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)(\widehat{\sigma}_t^2-\sigma_t^2)(\widehat{\sigma}_{t-1}^2-\sigma_{t-1}^2) \\
= & \frac{1}{2T}\sum_{t=1}^{T} G_{\sigma_t\sigma_{t-1}}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\frac{1}{N}\sum_{j=1}^{N}\sigma_{t_j}^2(\eta_{t_j}^2-1)\frac{1}{N}\sum_{j=1}^{N}\sigma_{t-1_j}^2(\eta_{t-1_j}^2-1) \\
= & \frac{1}{2TN^2}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{i=1}^{N} G_{\sigma_t\sigma_{t-1}}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\sigma_{t_j}^2(\eta_{t_j}^2-1)\sigma_{t-1_i}^2(\eta_{t-1_i}^2-1).
\end{aligned}$$

Hence,

$$E\left[\frac{1}{2T}\sum_{t=1}^{T} G_{\sigma_t\sigma_{t-1}}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)(\widehat{\sigma}_t^2-\sigma_t^2)(\widehat{\sigma}_{t-1}^2-\sigma_{t-1}^2)\right]=0$$

and

$$
\begin{aligned}
& \operatorname{var}\left[\frac{1}{2T}\sum_{t=1}^{T} G_{\sigma_t\sigma_{t-1}}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)(\widehat{\sigma}_t^2-\sigma_t^2)(\widehat{\sigma}_{t-1}^2-\sigma_{t-1}^2)\right] \\
= & \frac{1}{4T^2N^4}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{i=1}^{N} E\left[G_{\sigma_t\sigma_{t-1}}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\right]^2 E\left[\sigma_{t_j}^4\sigma_{t-1_i}^4(\eta_{t_i}^2-1)^2(\eta_{t-1_j}^2-1)^2\right] \\
= & E\left[G_{\sigma_t\sigma_{t-1}}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\right]^2 \frac{1}{T^2N^2}\sum_{t=1}^{T} E\left[\frac{1}{N}\sum_{j=1}^{N}\sigma_{t_j}^4\frac{1}{N}\sum_{i=1}^{N}\sigma_{t-1_i}^4\right] \\
\leqslant & E\left[G_{\sigma_t\sigma_{t-1}}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\right]^2 \frac{1}{T^2N^2}\sum_{t=1}^{T} M^2T^{2\epsilon} = O\left(\frac{T^{2\epsilon}}{TN^2}\right) = o\left(\frac{1}{T}\right),
\end{aligned}
$$

provided $\gamma > \epsilon$.

Consider next the fourth term in (33),

$$
\begin{aligned}
& \frac{1}{6T}\sum_{t=1}^{T} G_{\sigma_t\sigma_t\sigma_t}(X_t,\overline{\sigma}_t^2,\overline{\sigma}_{t-1}^2,\overline{\sigma}_{t-2}^2,\theta_0)(\widehat{\sigma}_t^2-\sigma_t^2)^3 \\
\leq & \left(\max_{1\leq t\leq T}\left|\widehat{\sigma}_t^2-\sigma_t^2\right|\right)^3 \frac{1}{6T}\sum_{t=1}^{T}\sup_{|x|,|x'|,|x''|\leq\delta_T}\left|G_{\sigma\sigma\sigma}(X_t,\sigma_t^2+x,\sigma_{t-1}^2+x,\sigma_{t-2}^2+x,\theta_0)\right| \\
= & O_p(T^{-3\alpha}).
\end{aligned}
$$

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

Finally, the final term in (33) is also $o_p(T^{-1/2})$ under the same conditions as the fourth term. Hence,

$$
\begin{aligned}
\widehat{G}_T(\theta_0) \simeq\; & G_T(\theta_0) + \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}E\left[G_{\sigma_t\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\right]E\left[IQ^t\right] \\
& + \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}E\left[G_{\sigma_{t-1}\sigma_{t-1}}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\right]E\left[IQ^{t-1}\right] \\
& + \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}E\left[G_{\sigma_{t-2}\sigma_{t-2}}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\right]E\left[IQ^{t-2}\right] \\
=\; & G_T(\theta_0) + b_T(\theta_0).
\end{aligned}
$$

Therefore, we have

$$
\widehat{\theta}-\theta_0 = -(\Gamma^\top W\Gamma)^{-1}\Gamma^\top W G_T(\theta_0) - (\Gamma^\top W\Gamma)^{-1}\Gamma^\top W b_T(\theta_0) + o_p(T^{-1/2}).
$$

Case 1: $\sqrt{T} b_T(\theta_0) = o_p(1)$

$$\begin{aligned} b_T(\theta_0) &= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}E\left[G_{\sigma_t\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\right]E\left[IQ^t\right] \\ &+\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}E\left[G_{\sigma_{t-1}\sigma_{t-1}}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\right]E\left[IQ^{t-1}\right] \\ &+\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}E\left[G_{\sigma_{t-2}\sigma_{t-2}}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\right]E\left[IQ^{t-2}\right]. \end{aligned}$$

Now,

$$\begin{aligned} &\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}E\left[G_{\sigma_t\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\right]E\left[IQ^t\right] \\ \leqslant\ & E\left[G_{\sigma_t\sigma_t}(X_t,\sigma_t^2,\sigma_{t-1}^2,\sigma_{t-2}^2,\theta_0)\right]\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}MT^{\epsilon} \\ =\ & O\left(\frac{T^{\epsilon}}{N}\right) = o\left(T^{-1/2}\right), \end{aligned}$$

provided $\gamma > \epsilon + 1/2$. This requires $\frac{N^\zeta}{T} \to \infty$ where $\zeta > 1.5$. In this case,

$$\sqrt{T}\left(\widehat{\theta}-\theta_0\right) = -(\Gamma^\top W\Gamma)^{-1}\Gamma^\top W\sqrt{T}G_T(\theta_0) + o_p(1).$$

Hence,

$$\sqrt{T}\left(\widehat{\theta}-\theta_0\right) \xrightarrow{d} N(0,\Sigma), \qquad \text{where } \Sigma = (\Gamma^\top W\Gamma)^{-1}\Gamma^\top W\Omega W\Gamma(\Gamma^\top W\Gamma)^{-1}.$$

Case 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 $\widehat{\theta} + (\Gamma^\top W\Gamma)^{-1}\Gamma^\top W b_T(\theta_0)$ 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 (46).

Define the estimated bias function

$$\begin{aligned}
\widehat{b}_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}G_{\sigma_t\sigma_t}(X_t,\widehat{\sigma}_t^2,\widehat{\sigma}_{t-1}^2,\widehat{\sigma}_{t-2}^2,\theta_0)\widehat{IQ}^{t} \\
&\quad +\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}G_{\sigma_{t-1}\sigma_{t-1}}(X_t,\widehat{\sigma}_t^2,\widehat{\sigma}_{t-1}^2,\widehat{\sigma}_{t-2}^2,\theta_0)\widehat{IQ}^{t-1} \\
&\quad +\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}G_{\sigma_{t-2}\sigma_{t-2}}(X_t,\widehat{\sigma}_t^2,\widehat{\sigma}_{t-1}^2,\widehat{\sigma}_{t-2}^2,\theta_0)\widehat{IQ}^{t-2}, \\
\text{where } \widehat{IQ}^{t} &= \frac{N_t}{3}\sum_{j=1}^{N_t} r_{t_j}^4
\end{aligned}$$

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

$$\widehat{\theta}^{bc} = \widehat{\theta} + (\widehat{\Gamma}^\top W\widehat{\Gamma})^{-1}\widehat{\Gamma}^\top W\widehat{b}_T(\widehat{\theta}).$$

Then,

$$\sqrt{T}(\widehat{\theta}^{bc} - \theta_0) \Longrightarrow N(0,\Sigma),$$

provided that

$$\sqrt{T}\widehat{b}_T(\widehat{\theta}) - \sqrt{T}b_T(\theta_0) = o_p(1).$$

■

B Tables

Table 1: Descriptive Statistics

		mean	variance	skewness	kurtosis	AR(1)	AR(1-12)	T
$(r_{mt}-r_{ft})^{mon}$	1928:01-2005:12	0.005	0.003	-0.477	9.819	0.102	0.209	936
	1928:01-1966:12	0.006	0.004	-0.369	9.873	0.123	0.296	468
	1967:01-2005:12	0.004	0.002	-0.755	5.734	0.062	0.061	468
$(\widehat{v}_t)^{mon}$	1928:01-2005:12	0.002	2.2×10^{-5}	6.317	58.47	0.563	4.449	936
	1928:01-1966:12	0.003	3.3×10^{-5}	4.399	29.02	0.657	5.261	468
	1967:01-2005:12	0.002	1.0×10^{-5}	13.04	218.5	0.226	1.263	468
$(r_{mt}-r_{ft})^{quar}$	1928:1-2005:4	0.014	0.012	0.175	9.802	-0.051	-0.145	312
	1928:1-1966:4	0.017	0.016	0.451	9.823	-0.086	-0.082	156
	1967:1-2005:4	0.011	0.008	-0.755	4.102	0.013	-0.296	156
$(\widehat{v}_t)^{quar}$	1928:1-2005:4	0.007	0.0001	4.248	25.08	0.554	4.342	312
	1928:1-1966:4	0.009	0.0002	3.254	15.35	0.601	4.849	156
	1967:1-2005:4	0.005	4.6×10^{-5}	6.080	54.14	0.264	1.162	156
$(r_{mt}-r_{ft})^{sm}$	1928:1-2005:2	0.028	0.022	-0.671	5.618	0.114	-0.381	156
	1928:1-1966:2	0.035	0.030	-0.811	5.420	0.175	-0.402	78
	1967:1-2005:2	0.022	0.014	-0.256	3.005	-0.014	-0.404	78
$(\widehat{v}_t)^{sm}$	1928:1-2005:2	0.014	0.0004	3.281	14.36	0.675	4.104	156
	1928:1-1966:2	0.018	0.0007	2.475	8.439	0.724	5.085	78
	1967:1-2005:2	0.011	0.0001	3.281	16.95	0.307	0.030	78
$(r_{mt}-r_{ft})^{year}$	1928-2005	0.057	0.040	-0.742	3.597	0.078	0.095	78
	1928-1966	0.069	0.052	-0.763	3.462	0.124	0.184	39
	1967-2005	0.045	0.028	-0.754	2.706	0.022	-0.256	39

Table 1. Summary statistics of logarithmic excess returns and realized variance. The table reports summary statistics of the continuously compounded excess returns on the stock market and the associated realized variance. Our market proxy is the CRSP value-weighted index (all stocks on the NYSE, AMEX, and NASDAQ). The proxy for the risk free rate is the one-month Treasury Bill rate (from Ibbotson Associates). Estimates are reported for the monthly, quarterly, semiannual, and annual frequencies. Monthly returns are calculated by compounding daily returns within calendar months. Monthly realized volatilities are constructed by cumulating squares of daily returns within each month, and so on. The table shows the mean, variance, skewness, kurtosis, first-order autocorrelation, and the sum of the first 12 autocorrelations,

AC(1-12), for each of the variables. The statistics are shown for the full sample and for two subsamples of equal length.

Table 2: Results Using Contemporaneous Variance

		α	β
	1928:01-2005:12	0.006	-0.493
		(1.896)	(-0.340)
monthly	1928:01-1966:12	0.007	-0.305
		(1.702)	(-0.197)
	1967:01-2005:12	0.009	-3.071
		(1.195)	(-0.711)
	1928:01-2005:12	0.011	0.472
		(0.926)	(0.248)
quarterly	1928:1-1966:4	0.018	-0.073
		(1.245)	(-0.036)
	1967:1-2005:4	-0.019	5.785
		(-1.060)	(1.786)
	1928:1-2005:4	0.021	0.473
		(0.984)	(0.276)
$semi-annually$	1928:1-1966:2	0.032	0.135
		(1.107)	(0.072)
	1967:1-2005:2	-0.025	4.447
		(-0.564)	(1.041)
	1928:1-2005:2	0.054	0.005
		(1.470)	(0.004)
annually	1928-1966	0.068	-0.129
		(1.326)	(-0.094)
	1967-2005	0.029	0.755
		(0.312)	(0.170)

Table 2. This table shows the GMM estimates for the model
$E[G(.)] = 0$ where

$$G = \begin{pmatrix} r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_{t+1} \\ \left(r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_{t+1}\right) v_t \end{pmatrix}$$

Table 3: Results For Over-identified System

		α	β
	1928:01-2005:12	0.005	-0.135
		(1.780)	(-0.099)
monthly	1928:01-1966:12	0.007	-0.420
		(2.011)	(-0.288)
	1967:01-2005:12	0.0001	1.920
		(0.026)	(0.598)
	1928:1-2005:4	0.013	0.102
		(1.054)	(0.049)
quarterly	1928:1-1966:4	0.018	-0.110
		(1.151)	(-0.049)
	1967:1-2005:4	-0.004	2.806
		(-0.214)	(0.839)
	1928:1-2005:4	0.024	0.206
		(1.156)	(0.130)
$semi-annually$	1928:1-1966:2	0.030	0.085
		(1.104)	(0.050)
	1967:1-2005:2	0.005	1.611
		(0.130)	(0.425)
	1928:1-2005:2	0.049	0.509
		(1.469)	(0.422)
annually	1928-1966	0.070	0.369
		(1.593)	(0.305)
	1967-2005	0.028	0.804
		(0.255)	(0.155)

Table 3. This table shows the estimates for the model $E[G(.)] = 0$ where

$$G = \begin{pmatrix} r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_t \\ (r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_t)\, v_{t-1} \\ (r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_t)\, v_{t-2} \\ (r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_t)\, v_{t-3} \end{pmatrix}$$

The table reports the coefficient estimates along with the associated t-stats in parentheses.

Table 4: Results Using Lagged Variance

		est.		Bias Corrected est.	
		α	β	α	β
	1928:01-2005:12	0.005	-0.277	0.005	-0.324
		(2.715)	(-0.344)	(2.767)	(-0.402)
monthly	1928:01-1966:12	0.006	-0.189	0.006	-0.217
		(2.138)	(-0.186)	(2.166)	(-0.213)
	1967:01-2005:12	0.005	-0.712	0.005	-0.889
		(2.233)	(-1.068)	(2.373)	(-1.333)
	1928:1-2005:4	0.012	0.262	0.012	0.275
		(1.851)	(0.250)	(1.836)	(0.263)
quarterly	1928:1-1966:4	0.017	-0.044	0.017	-0.047
		(1.945)	(-0.036)	(1.948)	(-0.038)
	1967:1-2005:4	0.003	1.526	0.003	1.542
		(0.406)	(1.881)	(0.395)	(1.900)
	1928:1-2005:2	0.024	0.319	0.024	0.330
		(1.541)	(0.275)	(1.531)	(0.284)
semiannually	1928:1-1966:2	0.032	0.098	0.032	0.101
		(1.487)	(0.072)	(1.484)	(0.075)
	1967:1-2005:2	0.008	1.350	0.007	1.400
		(0.524)	(1.273)	(0.489)	(1.320)
	1928-2005	0.054	0.003	0.054	0.003
		(1.924)	(0.004)	(1.924)	(0.004)
annually	1928-1966	0.067	-0.090	0.067	-0.091
		(1.598)	(-0.094)	(1.598)	(-0.095)
	1967-2005	0.039	0.274	0.038	0.304
		(1.039)	(0.178)	(1.022)	(0.197)

Table 7. This table shows the estimates for the model $E[G(r_{m,t+1} - r_{f,t}, v_t, v_{t-1}, v_{t-2}, \theta_0)] = 0$ where

$$G = \begin{pmatrix} r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_{t-1} \\ (r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_{t-1})\, v_{t-1} \end{pmatrix}$$

The table reports the coefficient estimates along with the associated t-stats in parentheses.

Table 5: Results For Over-Identified System

		est.		Bias Corrected est.	
		α	β	α	β
	1928:01-2005:12	0.005	-0.117	0.005	-0.130
		(2.529)	(-0.135)	(2.546)	(-0.151)
monthly	1928:01-1966:12	0.007	-0.271	0.007	-0.295
		(2.423)	(-0.265)	(2.449)	(-0.289)
	1967:01-2005:12	0.004	-0.246	0.004	-0.303
		(1.630)	(-0.261)	(1.671)	(-0.321)
	1928:1-2005:4	0.013	0.134	0.013	0.140
		(1.548)	(0.095)	(1.542)	(0.099)
quarterly	1928:1-1966:4	0.017	-0.054	0.017	-0.057
		(1.569)	(-0.033)	(1.571)	(-0.035)
	1967:1-2005:4	0.004	1.343	0.003	1.545
		(0.543)	(1.523)	(0.398)	(1.752)
	1928:1-2005:2	0.024	0.172	0.024	0.177
		(1.553)	(0.148)	(1.548)	(0.153)
semiannually	1928:1-1966:2	0.030	0.072	0.030	0.074
		(1.381)	(0.057)	(1.380)	(0.058)
	1967:1-2005:2	0.012	0.967	0.011	1.081
		(0.741)	(0.831)	(0.667)	(0.930)
	1928-2005	0.054	0.331	0.054	0.337
		(2.015)	(0.397)	(2.010)	(0.403)
annually	1928-1966	0.075	0.230	0.075	0.233
		(1.975)	(0.263)	(1.972)	(0.266)
	1967-2005	0.041	0.167	0.041	0.182
		(0.955)	(0.092)	(0.947)	(0.100)

Table 8. This table shows the estimates for the model $E[G(r_{m,t+1} - r_{f,t}, v_t, v_{t-1}, v_{t-2}, \theta_0)] = 0$ where

$$G = \begin{pmatrix} r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_{t-1} \\ (r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_{t-1})\, v_{t-1} \\ (r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_{t-1})\, v_{t-2} \\ (r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_{t-1})\, v_{t-3} \end{pmatrix}$$

The table reports the coefficient estimates along with the associated t-stats in parentheses and the J-stat for overidentifying restrictions.

Table 6: Statistics for Intra-Break Periods

		mean	variance	skewness	kurtosis	AR(1)	AR(1-12)	T
	1928:01-1954:12	0.006	0.005	-0.322	8.136	0.122	0.303	324
$(r_{mt} - r_{ft})^{mon}$	1955:01-1994:12	0.004	0.002	-0.708	6.372	0.082	-0.058	480
	1995:01-2005:12	0.006	0.002	-0.945	4.244	0.043	0.353	132
	1928:01-1954:12	0.004	4.5×10^{-5}	3.687	21.12	0.639	4.927	324
$(\widehat{v}_t)^{mon}$	1955:01-1994:12	0.001	8.6×10^{-6}	16.45	322.0	0.154	0.609	480
	1995:01-200512	0.002	6.1×10^{-6}	1.991	6.893	0.589	3.596	132
	1928:1-1954:4	0.017	0.021	0.500	8.043	-0.107	-0.106	108
$(r_{mt} - r_{ft})^{quar}$	1955:1-1994:4	0.011	0.007	-0.970	4.978	0.070	-0.365	160
	1995:1-2005:4	0.018	0.008	-0.561	2.871	-0.080	0.125	44
	1928:1-1954:4	0.012	0.0003	2.645	10.87	0.573	4.331	108
$(\widehat{v}_t)^{quar}$	1955:1-1994:4	0.004	3.6×10^{-5}	8.643	93.58	0.151	0.266	160
	1995:1-2005:4	0.007	3.9×10^{-5}	1.365	4.264	0.524	3.189	44
	1928:1-1954:2	0.034	0.038	-0.714	4.550	0.199	-0.610	54
$(r_{mt} - r_{ft})^{sm}$	1955:1-1994:2	0.023	0.015	-0.371	3.202	-0.139	-0.368	80
	1995:1-2005:2	0.036	0.009	-0.636	1.979	0.599	-0.379	22
	1928:1-1954:2	0.023	0.0009	1.912	5.619	0.700	4.558	54
$(\widehat{v}_t)^{sm}$	1955:1-1994:2	0.008	7.9×10^{-5}	5.288	37.91	0.207	0.392	80
	1995:1-2005:2	0.015	0.0001	1.021	3.060	0.409	0.053	22
	1928-1954	0.068	0.066	-0.768	3.016	0.164	-0.153	27
$(r_{mt} - r_{ft})^{year}$	1955-1994	0.045	0.025	-0.634	3.045	-0.184	-0.350	40
	1995-2005	0.072	0.034	-0.601	1.647	0.276	-	11
	1928-1954	0.047	0.003	1.712	5.088	0.666	4.983	27
$(\widehat{v}_t)^{year}$	1955-1994	0.016	0.0002	3.186	15.62	0.157	-0.189	40
	1995-2005	0.030	0.0004	0.373	1.530	0.597	-	11

Table 9. Summary statistics of excess returns and realized variance. The table reports summary statistics of excess returns on the stock market and the associated realized variance. The estimates are obtained in four ways, using monthly, quarterly, semi-annual, and yearly returns. Monthly returns are calculated by compounding daily returns within calendar months. Monthly realized volatilities are constructed by cumulating squares of daily returns within each month, and so on. Our market proxy is the CRSP value-

weighted index. The proxy for the risk free rate is the one-month Treasury Bill rate. The table shows the mean, variance, skewness, kurtosis, first-order serial correlation, and the sum of the first 12 autocorrelations, for each of the variables. The statistics are shown for three subsamples that are chosen to correspond to the structural break dates in the mean of the log dividend-price ratio as identified in Lettau and Van Nieuwerburgh (2007).

Table 7: Results Using Contemporaneous Variance

		α	β
	1928:01-1954:12	0.007	-0.252
		(1.167)	(-0.153)
monthly	1955:01-1994:12	0.013	-6.852
		(1.739)	(-1.207)
	1995:01-2005:12	0.006	-0.092
		(0.868)	(-0.030)
	1928:1-1954:4	0.018	-0.147
		(0.914)	(-0.066)
quarterly	1955:1-1994:4	-0.043	13.79
		(-2.067)	(3.340)
	1995:1-2005:4	0.024	-0.851
		(0.881)	(-0.208)
	1928:1-1954:2	0.032	0.070
		(0.761)	(0.034)
$semi-annually$	1955:1-1994:2	-0.100	15.62
		(-1.881)	(2.252)
	1995:1-2005:2	0.125	-5.969
		(2.267)	(-1.480)
	1928-1954	0.067	-0.163
		(0.849)	(-0.103)
annually	1955-1994	-0.178	14.20
		(-0.978)	(1.245)
	1995-2005	0.244	-5.799
		(2.567)	(-1.663)

Table 10. This table shows the estimates for the model $E[G(r_{m,t+1} - r_{f,t}, v_t, v_{t-1}, v_{t-2}, \theta_0)] = 0$ where

$$G = \begin{pmatrix} r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_t \\ (r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_t)\, v_{t-1} \end{pmatrix}$$

The table reports the coefficient estimates along with the associated t-stats in parentheses.

Table 8: Results For Over-Identified System

		α	β
	1928:01-1954:12	0.007	-0.448
		(1.456)	(-0.290)
monthly	1955:01-1994:12	0.0004	2.547
		(0.065)	(0.484)
	1995:01-2005:12	0.007	-0.310
		(0.877)	(-0.094)
	1928:1-1954:4	0.018	-0.206
		(0.843)	(-0.085)
quarterly	1955:1-1994:4	-0.042	13.53
		(-1.977)	(2.750)
	1995:1-2005:4	0.045	-3.875
		(1.931)	(-1.015)
	1928:1-1954:2	0.026	0.099
		(0.653)	(0.053)
$semi-annually$	1955:1-1994:2	-0.089	14.24
		(-1.664)	(2.017)
	1995:1-2005:2	0.150	-7.654
		(4.057)	(-2.793)
	1928-1954	0.077	0.239
		(1.151)	(0.182)
annually	1955-1994	0.012	2.117
		(0.095)	(0.252)
	1995-2005	0.292	-7.398
		(3.115)	(-2.362)

Table 11. This table shows the estimates for the model $E[G(r_{m,t+1} - r_{f,t}, v_t, v_{t-1}, v_{t-2}, \theta_0)] = 0$ where

$$G = \begin{pmatrix} r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_t \\ (r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_t)\, v_{t-1} \\ (r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_t)\, v_{t-2} \\ (r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_t)\, v_{t-3} \end{pmatrix}$$

The table reports the coefficient estimates along with the associated t-stats in parentheses.

Table 9: Results Using Lagged Variance

		est.		Bias Corrected est.	
		α	β	α	β
	1928:01-1954:12	0.006	-0.161	0.006	-0.186
		(1.520)	(-0.153)	(1.543)	(-0.177)
monthly	1955:01-1994:12	0.005	-1.502	0.006	-1.327
		(2.686)	(-2.111)	(2.873)	(-2.661)
	1995:01-2005:12	0.006	-0.054	0.006	-0.065
		(1.283)	(-0.030)	(1.289)	(-0.036)
	1928:1-1954:4	0.018	-0.084	0.018	-0.090
		(1.441)	(-0.066)	(1.447)	(-0.071)
quarterly	1955:1-1994:4	0.003	2.072	0.002	2.466
		(0.435)	(1.907)	(0.225)	(2.270)
	1995:1-2005:4	0.021	-0.446	0.022	-0.490
		(1.311)	(-0.205)	(1.331)	(-0.225)
	1928:1-1954:2	0.032	0.049	0.032	0.051
		(1.050)	(0.034)	(1.048)	(0.036)
semiannually	1955:1-1994:2	-0.002	3.174	-0.007	3.723
		(-0.138)	(1.886)	(-0.400)	(2.212)
	1995:1-2005:2	0.072	-2.428	0.074	-2.571
		(2.708)	(-1.128)	(2.788)	(-1.195)
	1928-1954	0.064	-0.108	0.065	-0.111
		(1.049)	(-0.104)	(1.052)	(-0.107)
annually	1955-1994	0.014	2.027	0.009	2.325
		(0.439)	(1.473)	(0.289)	(1.689)
	1995-2005	0.173	-3.429	0.177	-3.576
		(2.608)	(-1.146)	(2.673)	(-1.196)

Table 12. This table shows the estimates for the model
$E[G(r_{m,t+1} - r_{f,t}, v_t, v_{t-1}, v_{t-2}, \theta_0)] = 0$ where

$$G = \begin{pmatrix} r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_{t-1} \\ (r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_{t-1})\, v_{t-1} \end{pmatrix}$$

The table reports the coefficient estimates along with the associated t-stats in parentheses.

Table 10: Results For Over-Identified System

		est.		Bias Corrected est.	
		α	β	α	β
	1928:01-1954:12	0.005	-0.117	0.005	-0.130
		(2.529)	(-0.135)	(2.546)	(-0.151)
monthly	1955:01-1994:12	0.007	-0.271	0.007	-0.295
		(2.423)	(-0.265)	(2.449)	(-0.289)
	1995:01-2005:12	0.004	-0.246	0.004	-0.303
		(1.630)	(-0.261)	(1.671)	(-0.321)
	1928:1-1954:4	0.017	-0.108	0.017	-0.113
		(1.163)	(-0.064)	(1.167)	(-0.067)
quarterly	1955:1-1994:4	0.003	2.220	0.001	2.629
		(0.352)	(1.861)	(0.136)	(2.204)
	1995:1-2005:4	0.034	-2.130	0.035	-2.278
		(2.176)	(-0.888)	(2.247)	(-0.949)
	1928:1-1954:2	0.027	0.084	0.027	0.086
		(0.851)	(0.062)	(0.849)	(0.063)
semiannually	1955:1-1994:2	-0.001	3.061	-0.005	3.568
		(-0.085)	(1.871)	(-0.330)	(2.181)
	1995:1-2005:2	0.092	-3.795	0.095	-3.967
		(3.837)	(-1.711)	(3.943)	(-1.788)
	1928-1954	0.082	0.125	0.082	0.126
		(1.475)	(0.135)	(1.473)	(0.137)
annually	1955-1994	0.005	2.567	-0.0005	2.931
		(0.154)	(1.621)	(-0.010)	(1.851)
	1995-2005	0.202	-4.405	0.205	-4.519
		(2.595)	(-1.239)	(2.638)	(-1.271)

Table 13. This table shows the estimates for the model $E[G(r_{m,t+1} - r_{f,t}, v_t, v_{t-1}, v_{t-2}, \theta_0)] = 0$ where

$$G = \begin{pmatrix} r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_{t-1} \\ (r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_{t-1})\, v_{t-1} \\ (r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_{t-1})\, v_{t-2} \\ (r_{m,t+1} - r_{f,t+1} - \alpha - \beta v_{t-1})\, v_{t-3} \end{pmatrix}$$

The table reports the coefficient estimates along with the associated t-stats in parentheses and the J-stat for overidentifying restrictions.

Table 11: lognormal diffusion

	Panel A: $\alpha = 0, \beta = 5$		
	α	β	$Prob(\beta \leqslant \widehat{\beta})$
$N = 22, T = 936$	0.000	4.957	0.000
	(0.004)	(1.576)	
	$[-0.007, 0.007]$	$[2.427, 7.586]$	
$N = 66, T = 312$	-0.001	5.112	0.008
	(0.014)	(1.873)	
	$[-0.024, 0.021]$	$[2.093, 8.425]$	
$N = 132, T = 156$	0.001	4.916	0.030
	(0.034)	(2.286)	
	$[-0.051, 0.056]$	$[1.241, 8.733]$	
$N = 264, T = 78$	0.002	4.945	0.048
	(0.103)	(3.541)	
	$[-0.157, 0.164]$	$[-0.478, 10.02]$	
	Panel B: $\alpha = 0, \beta = 2$		
	α	β	$Prob(\beta \leqslant \widehat{\beta})$
$N = 22, T = 936$	-0.000	2.032	0.062
	(0.004)	(1.631)	
	$[-0.007, 0.008]$	$[-0.641, 4.770]$	
$N = 66, T = 312$	-0.001	2.048	0.168
	(0.014)	(1.891)	
	$[-0.028, 0.021]$	$[-1.110, 5.132]$	
$N = 132, T = 156$	0.001	1.934	0.204
	(0.035)	(2.387)	
	$[-0.059, 0.057]$	$[-1.721, 5.977]$	
$N = 264, T = 78$	0.001	2.023	0.182
	(0.096)	(3.238)	
	$[-0.148, 0.174]$	$[-3.012, 7.166]$	

The table reports simulation results for the lognormal diffusion. Panel A reports results for $\alpha = 0, \beta = 5$, while Panel B reports the same for $\alpha = 0, \beta = 2$. The first column reports the choice of the number of high and low-frequency observations. The choice corresponds to the corresponding numbers in the historical sample. The second and third columns report the mean across 500 simulations along with the standard deviation across the simulations in parentheses and the simulated 90% confidence interval in square brackets of the α and β estimates, respectively. The fourth column reports the probability of obtaining an estimate of β smaller than or equal to the value obtained in the historical sample.

Table 12: GARCH(1,1) diffusion

	$Panel\ A: \alpha = 0, \beta = 5$		
	α	β	$Prob(\beta \leqslant \widehat{\beta})$
$N = 22, T = 936$	0.000	4.839	0.004
	(0.005)	(2.064)	
	$[-0.007, 0.0078]$	$[1.436, 8.203]$	
$N = 66, T = 312$	-0.000	5.013	0.024
	(0.014)	(2.182)	
	$[-0.022, 0.023]$	$[1.493, 8.333]$	
$N = 132, T = 156$	-0.002	5.168	0.022
	(0.032)	(2.576)	
	$[-0.054, 0.046]$	$[1.402, 9.447]$	
$N = 264, T = 78$	-0.000	5.006	0.042
	(0.074)	(2.973)	
	$[-0.120, 0.112]$	$[0.362, 9.610]$	
	$Panel\ B: \alpha = 0, \beta = 2$		
	α	β	$Prob(\beta \leqslant \widehat{\beta})$
$N = 22, T = 936$	0.000	1.853	0.136
	(0.005)	(2.214)	
	$[-0.008, 0.008]$	$[-1.828, 5.822]$	
$N = 66, T = 312$	-0.000	2.095	0.224
	(0.014)	(2.229)	
	$[-0.023, 0.024]$	$[-1.428, 5.419]$	
$N = 132, T = 156$	0.000	1.965	0.258
	(0.032)	(2.554)	
	$[-0.048, 0.050]$	$[-2.092, 6.209]$	
$N = 264, T = 78$	-0.003	2.134	0.220
	(0.075)	(2.930)	
	$[-0.126, 0.105]$	$[-2.284, 7.090]$	

The table reports simulation results for the GARCH(1,1) diffusion. Panel A reports results for $\alpha = 0, \beta = 5$, while Panel B reports the same for $\alpha = 0, \beta = 2$. The first column reports the choice of the number of high and low-frequency observations. The choice corresponds to the corresponding numbers in the historical sample. The second and third columns report the mean across 500 simulations along with the standard deviation across the simulations in parentheses and the simulated 90% confidence interval in square brackets of the α and β estimates, respectively. The fourth column reports the probability of obtaining an estimate of β smaller than or equal to the value obtained in the historical sample.