ISSN 0956-8549-605

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

By

Anisha Ghosh
Oliver Linton

DISCUSSION PAPER NO 605

DISCUSSION PAPER SERIES

November 2007

Anisha Ghosh is a PhD student in the Department of Economics at London School of Economics and Political Science. She completed her BSc in Economics from Presidency College, Calcutta in 2003 and then the MRes in Economics at London School of Economics and Political Science in 2005. Oliver Linton is Professor of Econometrics in the Department of Economics, London School of Economics and Political Science. He is Co-editor of Econometric Theory and Associate Editor of: Econometrica, The Journal of Econometrics and of The Journal of the American Statistical Association.

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

Anisha Ghosh[†] Oliver Linton[‡]

this version: November 23, 2007

Abstract

This paper proposes an approach to estimating the relation between risk (conditional variance) and expected returns in the aggregate stock market that allows us to escape some of the limitations of existing empirical analyses. First, we focus on a nonparametric volatility measure that is void of any specific functional form assumptions about the stochastic process generating returns. Second, we offer a solution to the error-in-variables problem that arises because of the use of a proxy for the volatility in estimating the risk-return relation. Third, our estimation strategy involves the Generalized Method of Moments approach that overcomes the endogeneity problem in a least squares regression of an estimate of the conditional mean on the corresponding estimate of the conditional variance, that arises because both the above quantities are endogenously determined within a general equilibrium asset pricing model. Finally, we use our approach to assess the plausibility of the prominent Long Run Risks asset pricing models studied in the literature based on the restrictions that they imply on the time series properties of expected returns and conditional variances of market aggregates.

Keywords: Risk, Return, Nonparametric Volatility, Measurement Error, Generalized Method of Moments, Long Run Risks

JEL classification: C14, G12

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

[†]Department of 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

1 Introduction

The relation between the expected return and the risk of stock market investment 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.

Finance theory generally predicts a positive relationship between the risk premium on the market portfolio and the variance of its return. Prominent examples include the Intertemporal Capital Asset Pricing Model (ICAPM) of Merton (1973), and the more recent Long Run Risks (LLR) model of Bansal and Yaron (2004). However, some researchers have shown that the intertemporal mean-variance relation need not be positive theoretically, (e.g., Abel (1988), Backus and Gregory (1993), and Whitelaw (2000)).

Unfortunately, the empirical evidence on the risk-return relation is mixed and inconclusive. Bollerslev et al. (1988), Harvey (1989), Campbell and Hentschel (1992), Ghysels et al. (2005), Pastor, Sinha and Swaminathan (2006) find a positive risk-return relation, while Campbell (1987), Breen et al. (1989), Nelson (1991), Pagan and Hong (1991), Glosten et al. (1993), Whitelaw (1994), and Lettau and Ludvigson (2003) find a negative relation. French et al. (1987) find a negative relation between returns and the unpredictable component of volatility, a result they interpret as indirect evidence that ex ante volatility is positively related to ex ante excess returns but they do not find evidence of a direct connection between these variables. Scruggs (1998) and Guo and Whitelaw (2006) document a positive tradeoff 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 testing 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 relatively high degree of structure about which there is little direct empirical evidence.

Other studies have typically measured the conditional expectations underlying the conditional mean and conditional volatility 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 conditional mean and conditional volatility will be misspecified and possibly highly misleading[1].

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

In addition to the above critique, these studies typically estimate the risk-return trade-off using a least squares regression of the estimate of the conditional mean on the estimate of the conditional variance. However, 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 leading to invalid inference.

Finally, most of the literature ignores the error-in-variables problem that arises as a result of using an estimate of the conditional variance in estimating the risk-return relation; hence, inference can be misleading.

In this paper, we propose an approach to estimating the risk-return trade-off in the stock market that allows us to escape some of the limitations of existing empirical analyses. First, we focus on a nonparametric volatility measure, *realized volatility*, that is void of any specific functional form assumptions about the stochastic process generating returns and is easily computed from high-frequency intra-period returns (see Anderson, Bollerslev, Diebold, and Labys (2001), Barndorff-Nielsen and Shephard (2002, 2004)). The volatility estimates so constructed are model free and the theory of quadratic variation suggests that, under suitable conditions, realized volatility is an approximately unbiased and highly efficient estimator of return volatility.

Second, we offer a solution to the error-in-variables problem. Barndorff-Nielsen and Shephard (2002, 2004) derive the asymptotic distribution of realized volatility as an estimator of the underlying integrated volatility. We use their result and the standard setting of continuous time arbitrage-free theory based on a frictionless market to correct for the generated regressor.

Third, we estimate the risk-return trade-off parameters using the Generalized Method of Moments (GMM) approach of Hansen (1982) in the presence of measurement error in the volatility proxy. 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.

Here, we use (N) daily returns on the CRSP value-weighted index to obtain monthly and quarterly estimates of realized volatility that we use as a proxy for the conditional variance at the corresponding horizon. We then estimate the parameters of the risk-return trade-off using the GMM approach with T (monthly and quarterly, respectively) observations on the mean and variance. The measurement error issue is particularly important in this setting as the regressor is quite poorly estimated because of a small sample. We then derive the limiting distribution of the estimated coefficients. We discuss conditions under which the estimators are $\sqrt{T}$ -consistent and have an asymptotic normal distribution. We find that if $N^x/T \to \infty$, where $x > 1.5$, the estimates have the standard distribution as when there is no measurement-error problem and standard inference can be applied. Alternatively, one should bias correct, and we find that under the weaker condition that $N^x/T \to \infty$, where $x > 3$, the bias-corrected estimator has the standard limiting distribution. This improvement is particularly relevant in the case we examine where N is quite modest.

We find a statistically insignificant relation between the mean and the variance at both the monthly and quarterly frequencies. This finding is robust to the choice of instruments and across subsamples. These results motivate the hypothesis that 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 the nature of the time-variation in the relation, we split the sample into two subsamples based on the realized volatility estimates. In other words, the first subsample consists of two-third of the observations with low volatility and the second includes the remaining one-third with high volatility. Our results suggest a significantly positive relation during low volatility periods while the relation appears flat during periods of high volatility. This finding is robust across choice of samples.

Finally, we apply our methodology to assess the empirical plausibility of the LLR model of Bansal and Yaron (2004) and its extension considered by Bansal, Gallant, and Tauchen (2007). These models have a rich set of pricing implications and show promise in explaining a host of asset pricing puzzles[2]. We show that these LLR models imply a time-invariant, strictly positive, linear relation between the conditional expected excess return of stock market investment and its conditional variance, a feature that is not supported by the data.

The remainder of the paper is organized as follows. The theoretical underpinnings of realized volatility as an estimator of the conditional variance are discussed in Section 2. Section 3 describes the estimation procedure and the asymptotic distribution of the estimated model parameters is derived in Section 4. In Section 5, we describe the data and present the empirical results on the risk-return trade-off. Section 6 explores the possible time-variation in the relation. Section 7 outlines the relation between the conditional expected excess return and the conditional variance of the stock market that is implied by the LLR models and how our methodology can be applied to gauge the plausibility of these models. Section 8 concludes. The Appendix contains the proofs of our main results.

2 Nonparametric Volatility Estimator

Under the standard 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, as detailed by Back

[2]See also, Alvarez and Jerman (2005), Bansal, Dittmar and Lundblad (2005), Bansal, Kiku and Yaron (2007), Bekaert, Engstrom and Xing (2005), Hansen, Heaton, and Li (2005), Hansen and Scheinkman (2007), Kiku (2006), Lettau and Ludvigson (2005), and Malloy, Moskowitz and Vissing-Jorgensen (2004).

(1991). If, in addition, it is assumed that the sample paths are continuous, we have the Martingale Representation Theorem (e.g., Protter (1992), Karatzas and Shreve (1991)).

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 $0 \leq t \leq T$, a.s.(P),*

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

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

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

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

The integral representation (1) is equivalent to the standard stochastic differential equation specification for the logarithmic price process,

$$dp(t) = \mu(t)dt + \sigma(t)dW(t). \tag{3}$$

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

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 $m \to \infty$,*

$$\tau_{m,0} \to 0; \qquad \sup_{j\geq 1} \tau_{m,j} \to T; \qquad \sup_{j\geq 0} (\tau_{m,j+1}-\tau_{m,j}) \to 0. \tag{4}$$

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

$$\lim_{m\to\infty}\{\sum_{j\geq 1}(p(t\wedge\tau_{m,j}) - p(t\wedge\tau_{m,j-1}))^2\} \to [r,r]_t, \tag{5}$$

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. Under the maintained assumption of continuous sample path, the notional volatility equals the so-called integrated volatility. In other words, we have

Definition 1 *Notional Volatility*

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

$$\upsilon^2(t,h) \equiv [r,r]_t - [r,r]_{t-h} = [M,M]_t - [M,M]_{t-h} = \int_0^h \sigma^2(t-h+s)ds. \quad (6)$$

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

$$E[\upsilon^2(t,h)|I_{t-h}] = E[M^2(t,h)|I_{t-h}] = E[M^2(t)|I_{t-h}] - M^2(t-h). \quad (7)$$

Now, in the above setting, the conditional volatility, or expected volatility, over $[t-h,t]$, is defined by

$$\begin{aligned}
\operatorname{var}\left(r(t,h)|I_{t-h}\right) &\equiv E\left[\left\{r(t,h) - E\left(r(t,h)|I_{t-h}\right)\right\}^2 |I_{t-h}\right] \\
&= E\left[\left\{r(t,h) - E\left(\mu(t,h)|I_{t-h}\right)\right\}^2 |I_{t-h}\right] \\
&= E\left[\left\{\mu(t,h) - E\left(\mu(t,h)|I_{t-h}\right) + M(t,h)\right\}^2 |I_{t-h}\right] \\
&= E[M^2(t,h)|I_{t-h}] + E\left[\left\{\mu(t,h) - E\left(\mu(t,h)|I_{t-h}\right)\right\}^2 |I_{t-h}\right] \\
&\quad +2E\left[\left\{\mu(t,h) - E\left(\mu(t,h)|I_{t-h}\right)\right\} M(t,h)|I_{t-h}\right] \\
&= O_p(h) + O_p(h^2) + O_p(h^{3/2}), \qquad (8)
\end{aligned}$$

where the second equality follows as M is a local martingale, and the third equality follows from (1).

From equations (7) and (8), we have

$$\operatorname{var}\left(r(t,h)|I_{t-h}\right) \approx E[\upsilon^2(t,h)|I_{t-h}]. \quad (9)$$

In other words, expected volatility is well approximated by expected notional volatility. The above approximation is exact if the mean process, $\mu(t) \equiv 0$, of 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 5 to justify this approximation for the horizons, h, and sampling frequencies considered in this paper.

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

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

$$\widehat{\upsilon}^2(t,h;n) \equiv \sum_{i=1}^{n} r(t-h+(i/n)h, h/n)^2. \quad (10)$$

The realized volatility 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 $1/n$), so that it provides a volatility measure calibrated to the h-period measurement interval.

The theory of quadratic variation implies the following result (see, e.g., Andersen, Bollerslev, Diebold, and Labys (2000b, 2001a,b), Barndorff-Nielsen and Shephard (2001a, 2002a,b,c)).

Proposition 3 *The Realized Volatility provides a consistent nonparametric measure of the Notional Volatility,*

$$p\lim_{n\to\infty} \widehat{v}(t,h;n) = v^2(t,h), \quad 0 < h \leq t \leq T, \tag{11}$$

where the convergence is uniform in probability.

The following result was developed in a series of papers by Jacod (1994), Jacod and Protter (1998), Barndorff-Nielsen and Shephard (2002) and Barndorff-Nielsen and Shephard (2004):

Proposition 4 *Suppose that* $p \epsilon BSM$ *is one–dimensional and that (* $for\ all\ t < \infty$ *)* $\int_0^t \mu_s ds < \infty$, *then as* $n \to \infty$

$$n^{1/2}(\widehat{v}^2(t,h;n) - \int_0^h \sigma^2(t-h+s)ds) \to \sqrt{2}\int_0^h \sigma^4(t-h+s)dB(s), \tag{12}$$

where B *is a Brownian motion and the convergence is in law stable as a process.*

The above theorem implies that

$$n^{1/2}(\widehat{v}^2(t+h,h;n) - \int_0^h \sigma^2_{t-h+s}ds \Longrightarrow MN(0,2\int_0^h \sigma^4_{t+s}ds), \tag{13}$$

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 $\int_0^h \sigma^4_{t+s}ds$ can be consistently estimated using $(1/3)RQ_{t+h}$ where

$$RQ_{t+h} = n\sum_{i=1}^{n} r(t+(i/n)h, h/n)^4. \tag{14}$$

In particular then

$$\frac{n^{1/2}(\widehat{v}^2(t+h,h;n) - \int_0^h \sigma^2_{t+s}ds}{\sqrt{\frac{2}{3}RQ_{t+h}}} \Longrightarrow N(0,1). \tag{15}$$

This is a nonparametric result as it does not require us to specify the form of the drift or diffusion terms.

3 Model and Estimator

We focus on a linear relation between the expected returns and the conditional variance of the aggregate stock market as is implied by several prominent asset pricing models,

$$E_{t-1}\left(r_{m,t}-r_{f,t}\right)=\alpha+\beta \mathrm{var}_{t-1}\left(r_{m,t}\right), \tag{16}$$

where $r_{m,t}$ and $r_{f,t}$ are the continuously compounded returns on the stock market and the riskfree rate respectively over $[t-1,t]$.

Equations (6) and (9) in Section 2 imply that the above relation implies the following conditional moment restriction,

$$E_{t-1}\left(r_{m,t}-r_{f,t}-\alpha+\beta\upsilon_t\right)=0, \tag{17}$$

where $\upsilon_t=\int_0^1\sigma^2(t-1+s)ds$.

The above is an infeasible moment restriction as the integrated volatility, $\int_0^h\sigma^2(t-h+s)ds$, is not observable. To obtain a proxy for it, we adopt the following framework.

Let $\{r_{t_j}\}_{j=1}^{N_t}$ be intra-period (daily) continuously compounded returns on the market portfolio for each period $t=1,\ldots,T$. Suppose that

$$r_{t_j}=N_t^{-1}\mu_{t_j}+N_t^{-1/2}\sigma_{t_j}\eta_{t_j}, \tag{18}$$

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 $\left\{\mu_{t_j},\sigma_{t_j}\right\}$ 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 leverage effects. In particular η_{t_j} can affect $\sigma_{s_{j+k}}$ for $s\geqslant t$ and $k\geqslant 1$. $\sigma_{t_j}^2$ is the integral of the volatility function over a small interval, (see, e.g., Gonçalves and Meddahi (2005)). Define

$$\upsilon_t=p\lim_{N_t\to\infty}\frac{1}{N_t}\sum_{j=1}^{N_t}\sigma_{t_j}^2. \tag{19}$$

This framework is consistent with r_{t_j} being observations from a diffusion, equation (3), where W is a standard Brownian motion, and υ_t being the quadratic variation of the diffusion $\upsilon_t=\int_{t-1}^t\sigma^2(s)ds$. Let

$$\widehat{\upsilon}_t=\sum_{j=1}^{N_t}r_{t_j}^2. \tag{20}$$

Given the volatility estimator, we define the feasible moment condition

$$E_{t-1}\left(r_{m,t}-r_{f,t}-\alpha+\beta\widehat{\upsilon}_t\right)=0 \tag{21}$$

Finally, with a set of chosen conditioning variables (that could include, for instance, lagged volatilities), $\widehat{z}_{t-1}$, we have

$$E\left[G\left(r_{m,t}-r_{f,t},\widehat{v}_t,\widehat{z}_{t-1};\theta\right)\right]=0, \tag{22}$$

where $G\left(r_{m,t}-r_{f,t},\widehat{v}_t,z_{t-1};\theta\right)=\left(r_{m,t}-r_{f,t}-\alpha+\beta\widehat{v}_t\right)\otimes z_{t-1}$, and $\theta=(\alpha,\beta)'$.

Given the above set of moment restrictions, the parameters may be estimated using the GMM approach of Hansen (1982). Specifically, we define the estimator $\widehat{\theta}\in\Theta_0\subseteq\mathbb{R}^p$ as any minimizer of

$$\left\|\widehat{G}_T(\theta)\right\|_W, \quad \widehat{G}_T(\theta)=\frac{1}{T}\sum_{t=1}^{T}G\left(r_{m,t}-r_{f,t},\widehat{v}_t,\widehat{z}_{t-1};\theta\right),$$

where W is a symmetric positive definite weighting matrix and $||A||_W=(\mathrm{tr}(A^\top WA))^{1/2}$. Define also $G_T(\theta)=T^{-1}\sum_{t=1}^{T}G\left(r_{m,t}-r_{f,t},v_t,\widehat{z}_{t-1};\theta\right)$ and the infeasible GMM estimator $\widetilde{\theta}$ that minimizes $||G_T(\theta)||_W$.

4 Asymptotic Properties

The asymptotic framework has $T\to\infty$ and $N_t\to\infty$ for each t. Under certain mild assumptions (see Appendix A for details), we have,

Lemma. *Under conditions 1-4 (stated in Appendix A),*

$$T^{\alpha}\left[\max_{1\leqslant t\leqslant T}(\widehat{v}_t-v_t)\right]=o_p(1). \tag{23}$$

Note that the above result implies that $\widehat{v}_t\xrightarrow{p}v_t$ uniformly in t. Using the above result, we give the asymptotic properties of $\widehat{\theta}$ and propose a modification that has better properties.

Let $\overline{G}(\theta)=E[G\left(r_{m,t}-r_{f,t},v_t,z_{t-1};\theta\right)]$ and define

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

Then under suitable conditions the infeasible GMM estimator $\widetilde{\theta}$ satisfies

$$\sqrt{T}(\widetilde{\theta}-\theta_0)=-(\Gamma^\top W\Gamma)^{-1}\Gamma^\top W\sqrt{T}G_T(\theta_0)+o_p(1)\Longrightarrow N(0,\Sigma), \tag{24}$$

where $\Sigma=(\Gamma^\top W\Gamma)^{-1}\Gamma^\top W\Omega W\Gamma(\Gamma^\top W\Gamma)^{-1}$, (see, e.g., Pakes and Pollard (1989)). This theory does 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, in most applications G will be smooth and we shall assume this in the sequel. It is natural to suppose that the process $\{r_{m,t} - r_{f,t}, v_t\}$ is stationary and weakly dependent, e.g., strong mixing, which would support the central limit theorem in (24). We need further conditions for the precise results we obtain; these are stated in Appendix B.

Define the bias function

$$\begin{aligned} b_T(\theta_0) &= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}E\left[G_{v_tv_t}\left(r_{m,t}-r_{f,t}, v_t, z_{t-1}; \theta_0\right) IQ^t\right] \\ &\quad + \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}E\left[G_{v_{t-1}v_{t-1}}\left(r_{m,t}-r_{f,t}, v_t, z_{t-1}; \theta_0\right) IQ^{t-1}\right] \\ &\quad + \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}E\left[G_{v_{t-2}v_{t-2}}\left(r_{m,t}-r_{f,t}, v_t, z_{t-1}; \theta_0\right) IQ^{t-2}\right], \end{aligned}$$

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

$$IQ^t = p\lim_{N_t\to\infty}\frac{1}{N_t}\sum_{j=1}^{N_t}\sigma_{t_j}^4. \tag{25}$$

We have the following result established in the appendix.

THEOREM. *Under the regularity conditions given in the appendix, we have*

$$\widehat{\theta}-\theta_0 = -(\Gamma^\top W\Gamma)^{-1}\Gamma^\top W G_T(\theta_0) - (\Gamma^\top W\Gamma)^{-1}\Gamma^\top W b_T(\theta_0) + o_p(T^{-1/2}). \tag{26}$$

Furthermore, when $b_T(\theta_0) = o(T^{-1/2})$,

$$\sqrt{T}(\widehat{\theta}-\theta_0) \Longrightarrow N(0,\Sigma). \tag{27}$$

When (27) holds, standard inference can be applied. Specifically, when $G(r_{m,t} - r_{f,t}, v_t, z_{t-1}, \theta)$ is a martingale difference sequence, we take

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

Then $\widehat{\Sigma} = (\widehat{\Gamma}^\top W\widehat{\Gamma})^{-1}\widehat{\Gamma}^\top W\widehat{\Omega}W\widehat{\Gamma}(\widehat{\Gamma}^\top W\widehat{\Gamma})^{-1}$ is a consistent estimator of Σ. The condition $b_T(\theta_0) = o(T^{-1/2})$ requires that $N^{1.5}/T \to \infty$. When that 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 provided only the weaker condition that a little more than $N^3/T \to \infty$ holds. 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 (27). We can either correct the estimator or the moment condition. Define the estimated bias function

$$\begin{aligned} \widehat{b}_T(\theta) \quad = \quad & \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}G_{v_t v_t}(r_{m,t}-r_{f,t},\widehat{v}_t,\widehat{z}_{t-1},\theta_0)\widehat{IQ}^{t} \\ & +\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}G_{v_{t-1} v_{t-1}}(r_{m,t}-r_{f,t},\widehat{v}_t,\widehat{z}_{t-1},\theta_0)\widehat{IQ}^{t-1} \\ & +\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}G_{v_{t-2} v_{t-2}}(r_{m,t}-r_{f,t},\widehat{v}_t,\widehat{z}_{t-1},\theta_0)\widehat{IQ}^{t-2}, \end{aligned} \tag{28}$$

where $\widehat{IQ}^{t} = \frac{N_t}{3}\sum_{j=1}^{N_t} r_{t_j}^4$ 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 under some conditions, $\sqrt{T}(\widehat{\theta}^{bc} - \theta_0)$ has the limiting distribution in (27).[3]

5 Data Description and Empirical Results

We focus on the risk-return relation at the monthly and quarterly frequencies. The empirical analysis is based on data from the Centre for Research in Security Prices (CRSP) daily returns data file. Our market proxy is the CRSP value-weighted index (all stocks). The proxy for the riskfree rate is the one-month Treasury Bill rate (from Ibbotson Associates). The sample extends from January 1928 - December 2005. 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 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.

[3] Alternatively, define the bias corrected moment function

$$\widehat{G}_T^{bc}(\theta) = \widehat{G}_T(\theta) + \widehat{b}_T(\theta)$$

and the bias corrected estimator $\widehat{\theta}^{bc} = \arg\min_\theta ||\widehat{G}_T^{bc}(\theta)||_W$.

To set the stage, Table 1 reports summary statistics for the excess returns and the corresponding realized volatilities 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. This table reports variances instead of the more customary standard deviations because the risk-return trade-off postulates a relation between returns and their variances, and not standard deviations. Returns are slightly negatively skewed and leptokurtic. The first order autocorrelation 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. 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 autoregressive 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 (0.656 vs. 0.226). As expected, realized variance is highly skewed and leptokurtic. Most of these features of the data have been previously documented in the literature. Also, most of these characteristics of returns and the realized variances persist at the quarterly horizon.

Next, we turn to our main empirical results. The analysis in Section 4 showed that the estimation of the risk-return trade-off parameters can be posed as a GMM estimation problem, with the following moment specification,

$$E\left[G\left(r_{m,t} - r_{f,t}, \widehat{v}_t, \widehat{z}_{t-1}; \theta\right)\right] = 0, \tag{29}$$

where $G\left(r_{m,t} - r_{f,t}, \widehat{v}_t, \widehat{z}_{t-1}; \theta\right) = \left(r_{m,t} - r_{f,t} - \alpha + \beta \widehat{v}_t\right) \otimes \widehat{z}_{t-1}$, $\theta = (\alpha, \beta)'$, 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 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 equation (28)). 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 quarterly data, the estimated coefficients are mostly positive but not statistically significant. Table 3 confirms the findings in Table 2.

The rationale for using lagged integrated volatility as an instrument in Tables 2 and 3 is that it is a highly persistent process (the first order autocorrelation coefficient of the realized volatility process is 0.563 and 0.0.554 respectively, in monthly and quarterly data for the full sample). Hence, the lagged variance is useful in predicting the contemporaneous variance which enters the moment specification. This makes it a good choice of instrument improving the efficiency of the estimation exercise.

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. In Tables 4, 5, and 6, we report estimation results for these choice of instruments respectively. The Tables reveal a statistically insignificant relation, over the full sample as well as the subsamples, that is robust to the choice of instruments.

6 Time-Variation in the Risk-Return Tradeoff

A closely related literature on return predictability 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), Paye and Timmermann (2005)). This renders the forecasting relationship unstable if such shifts are not taken into account. In particular, Lettau and Van Nieuwerburgh (2006) 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 possibility, we split the sample into two subsamples based on the higher one-third quartile of the realized volatility estimates. In other words, the first subsample consists of two-third of the observations with low volatility and the second includes the remaining one-third with high volatility. Table 7 reports the estimation results for this choice of subsamples with the specification of the moment restrictions as in Table 2. This table reveals that the risk-return tradeoff is significantly positive at the monthly and quarterly horizons for the low volatility subsample. The estimated coefficients are 25.97 and 26.36 at the monthly and quarterly horizons respectively and are significant at conventional levels. However, the estimates for the high volatility period are, although positive, substantially lower than those for the low volatility period, and insignificantly different from zero for the above horizons.

For additional robustness, we repeat the above exercise excluding the time-period 1928-1954. This time-period includes periods of great economic uncertainty like the Great Depression, the World Wars I andII, and hence could potentially bias the results.

The results are presented in Table 8. The results are largely similar to those obtained in Table 7. The risk-return relation is significantly positive during the low volatility regime but appears quite unstable during high volatility periods.

7 Implications for the Long Run Risks model

Theoretical asset pricing models are aimed at explaining key stylized facts of stock market data including the 6% equity premium, the low risk-free rate, market volatility of 19% per annum, fluctuating and highly persistent conditional variance of the market return, and predictive power of price-dividend ratios for long-horizon equity returns. Bansal and Yaron (2004) introduce a "long-run risks" (LLR) state variable that simultaneously drives aggregate consumption growth and aggregate dividend growth[4]. In conjunction with Kreps and Porteus (1978) preferences, this LLR model has a rich set of pricing implications and shows promise in explaining the host of asset pricing puzzles mentioned above as well as the cross-section of expected returns of various classes of financial assets.

In particular, Bansal and Yaron (2004) model consumption, Δc_{t+1}, and dividend, Δd_{t+1}, growth rates as containing (1) a small persistent expected growth rate component, x_t, and (2) fluctuating volatility, σ_t, that captures time-varying economic uncertainty,

$$\begin{aligned}
\Delta c_{t+1} &= \mu_c + x_t + \sigma_t \eta_{t+1} \\
\Delta d_{t+1} &= \mu_d + \phi x_t + \varphi_d \sigma_t u_{j,t+1} \\
x_{t+1} &= \rho x_t + \varphi_e \sigma_t e_{t+1} \\
\sigma^2_{t+1} &= \sigma^2 + \upsilon(\sigma^2_t - \sigma^2) + \sigma_w w_{t+1} \\
& \quad \eta_{t+1}, u_{j,t+1}, e_{t+1}, w_{t+1} \sim i.i.d.N(0,1)
\end{aligned} \tag{30}$$

with the shocks η_{t+1}, $u_{j,t+1}$, e_{t+1}, w_{t+1} being mutually independent.

Using the log-linearization of returns as in Campbell and Shiller (1988), and conjecturing that the log price-dividend ratios of the unobservable consumption claim, z_t, and the observable aggregate dividend claim (the market portfolio), $z_{m,t}$, are linear in the state variables, x_t and σ^2_t,

$$\begin{aligned}
z_t &= A_0 + A_1 x_t + A_2 \sigma^2_t \\
z_{m,t} &= A_{0,m} + A_{1,m} x_t + A_{2,m} \sigma^2_t,
\end{aligned}$$

[4] See also, Alvarez and Jerman (2005), Bansal, Dittmar and Lundblad (2005), Bansal, Gallant and Tauchen (2007), Bansal, Kiku and Yaron (2007), Bekaert, Engstrom and Xing (2005), Hansen, Heaton, and Li (2005), Hansen and Scheinkman (2007), Kiku (2006), Lettau and Ludvigson (2005), and Malloy, Moskowitz and Vissing-Jorgensen (2004).

(see Bansal and Yaron (2004) for expressions for A_0, A_1, A_2, $A_{0,m}$, $A_{1,m}$, $A_{2,m}$), they derive the following expression for the equity risk premium,

$$E_t\left[r_{m,t+1}-r_{f,t}\right]=(1-\theta)\kappa_1 A_1\kappa_{1,m}A_{1,m}\varphi_e^2\sigma_t^2+(1-\theta)\kappa_1 A_2\kappa_{1,m}A_{2,m}\sigma_w^2-0.5\mathrm{var}_t\left(r_{m,t+1}\right). \tag{31}$$

Thus, the model predicts a linear relation between the ex ante expected excess returns on the market portfolio and the conditional variance of consumption growth, σ_t^2. The model also implies a linear relation between the conditional variance of consumption growth and the conditional variance of returns on the market,

$$\mathrm{var}_t\left(r_{m,t+1}\right)=\left(\kappa_{1,m}^2A_{1,m}^2\varphi_e^2+\varphi_d^2\right)\sigma_t^2+\kappa_{1,m}^2A_{2,m}^2\sigma_w^2. \tag{32}$$

Substituting the expression for σ_t^2, we have:

$$E_t\left[r_{m,t+1}-r_{f,t}\right]=\alpha+\beta\mathrm{var}_t\left(r_{m,t+1}\right) \tag{33}$$

$$\begin{aligned}
\alpha &= \frac{(\theta-1)\kappa_1 A_1\kappa_{1,m}^3A_{1,m}A_{2,m}^2\varphi_e^2\sigma_w^2}{\kappa_{1,m}^2A_{1,m}^2\varphi_e^2+\varphi_d^2}-(\theta-1)\kappa_1 A_2\kappa_{1,m}A_{2,m}\sigma_w^2\\
\beta &= \frac{-(\theta-1)\kappa_1 A_1\kappa_{1,m}A_{1,m}\varphi_e^2}{\kappa_{1,m}^2A_{1,m}^2\varphi_e^2+\varphi_d^2}-\frac{1}{2},
\end{aligned}$$

where $\theta\equiv 1-\gamma/(1-\frac{1}{\psi})$. Thus, the model predicts a time-invariant, linear relation between the conditional mean of the market portfolio and its conditional variance, a feature that is violated in the data (see Sections 5 and 6).

Bansal, Gallant, and Tauchen (2007) consider an extension of the LLR model that imposes a cointegrating restriction between the logarithms of aggregate consumption and dividend levels,

$$d_t-c_t=\mu_{dc}+s_t, \tag{34}$$

where s_t is an $I(0)$ process,

$$s_{t+1}=\lambda_{sx}x_t+\rho_s s_t+\psi_s\sigma_t z_{s,+1.}$$

With the same specification of the dynamics of the consumption growth process, the LLR variable, and the stochastic volatility as in (30), and noting that (34) implies that $\Delta d_{t+1}=\Delta c_{t+1}+\Delta s_{t+1}$, the model may be solved using similar techniques as in Bansal and Yaron (2004) to yield a time-invariant, linear relation between the ex ante expected excess return of the stock market and its conditional variance.

Hence, in order to explain the empirical finding that the risk-return relation exhibits significant time-variation, the specification of the model must be quite different from those emphasized in the existing LLR literature.

8 Conclusion

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, we focus on a nonparametric volatility measure, namely realized volatility, that is void of any specific functional form assumptions about the stochastic process governing returns. Second, we offer a solution to the error-in-variables problem that arises because of the use of a volatility proxy in the risk-return relation. Third, we estimate the risk-return trade-off parameters using the Generalized Method of Moments (GMM) approach. 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.

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 and choice of instruments. However, when the sample is split into subsamples based on the estimate of the variance, we find a positive and statistically significant relation during low volatility periods while the relation appears flat during the high volatility regime. These results are suggestive of significant time-variation in the risk-return relation.

These empirical findings are especially interesting because they run counter to the notion of a time-invariant, linear relation between volatility and expected returns at the market level that is implied by models such as the Long Run Risks Model of Bansal and Yaron (2004) and its extension by Bansal, Gallant, and Tauchen (2007) that imposes a cointegrating relationship between the logarithms of aggregate consumption and dividend levels. These models have a rich set of pricing implications and show promise in explaining a host of asset pricing puzzles as well as the cross-section of expected returns of various classes of financial assets. In order to explain the empirical finding that the risk-return relation exhibits significant time-variation, the specification of the models must be quite different from those emphasized in the existing LLR literature.

References

[1] Anderson, T.G., Bollerslev, T., Diebold, F.X., and Labys, P., (2003), "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579-625.

[2] Ang, A., and Bekaert, G., (2001), "Stock Return Predictability: Is it There?," *Manuscript*, Columbia University.

[3] Back, K., (1991), "Asset Pricing for General Processes," *Journal of Mathematical Economics*, 20, 371-395.

[4] Bansal, R., Dittmar, R.F., and Lundblad, C., (2005), "Consumption, Dividends, and the Cross-Section of Equity Returns," *Journal of Finance* 60, 1639-1672.

[5] Bansal, R., Gallant, R., and Tauchen, G., (2005), "Rational Pessimism, Rational Exhuberance, and Markets for Macro Risk," Working Paper, Duke University.

[6] Bansal, R., Kiku, D., and Yaron, A., (2006), "Risks For the Long Run: Estimation and Inference," Working Paper.

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

[8] Bekaert, G., Engstrom, E., and Xing, Y., (2005), "Risk, Uncertainty, and Asset Prices," Working Paper.

[9] Barndorff-Nielsen, O.E., and Shephard, N., (2002), "Econometric Analysis of Realized Volatility and its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society, Series B* 64, 253-280.

[10] Barndorff-Nielsen, O.E., and Shephard, N., (2004), "Econometric Analysis of Realized Covariation: High Frequency Covariance, Regression, and Correlation in Financial Economics," *Econometrica*, 72, 885-925.

[11] Bollerslev, T., Chou, R.Y., and Kroner, K.F., (1992), "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics*, 52, 5-59.

[12] Bollerslev, T., Engle, R.F., and Nelson, D.B., (1994), "ARCH Models," in: Engle, R.F., and McFadden, D. (Eds.), Handbook of Econometrics, vol. 4. Elsevier Science B.V., Amsterdam.

[13] Campbell, J.Y., (1987), "Stock Returns and the Term Structure," *Journal of Financial Economics*, 18 (2), 373-399.

[14] Campbell,J., and Cochrane, J., (1999), "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107, 205-251.

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

[16] Campbell, J.Y., and Shiller, R.J., (1988a), "Stock Prices, Earnings, and Expected Dividends," *Journal of Finance*, 43 (3), 661-676.

[17] Campbell, J.Y., and Shiller, R.J., (1988b), "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1, 195-227.

[18] Engle, R.F., (2000), "The Econometrics of Ultra-High Frequency Data," *Econometrica* 68, 1-22.

[19] Epstein, L.R., and Zin, S.E., (1989), "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica* 57, 937-969.

[20] Fama, E.F., and French, K.R., (1988), "Dividend Yields and Expected Stock Returns," *J. Financial Econ.*, 22, 3-27.

[21] Fama, E.F., and French, K.R., (1989), "Business Conditions and Expected Returns on Stocks and Bonds," *J. Financial Econ.*, 25, 23-39.

[22] Fama, E.F., and French, K.R., (1993), "Common Risk Factors in the Returns on Stocks and Bonds." *J. Financial Econ.* 33, 3-56.

[23] Ferson, W., and Foerster, S.R., (1994), "Finite Sample Properties of the Generalized Method of Moments in Tests of Conditional Asset Pricing Models," *Journal of Financial Economics*, 36, 29-55.

[24] French, K.R., Schwert, W., and Stambaugh, R.F., (1987), "*Expected Stock Returns and Volatility*," *Journal of Financial Economics*, 19 (1), 3-29.

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

[26] Glosten, L.R., Jagannathan,R., and Runkle, D.E., (1993), "On The Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48 (5), 1779-1801.

[27] Gonçalves, S., and N. Meddahi (2005). "Bootstrapping Realized Volatility," Manuscript, Universitė de Montreal.

[28] Goyal, A., and Welch, I., (2003), "Predicting the Equity Premium with Dividend Ratios," *Management Science*, 49(5), 639-654.

[29] Hansen, L.P., (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

[30] Hansen, L.P., Heaton, J., and Li, N., (2005), "Consumption Strikes Back," Working Paper, University of Chicago.

[31] Harvey, C.R. (2001), "The Specification of Conditional Expectations," *Journal of Empirical Finance*, 8 (5), 573-638.

[32] 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.

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

[34] Karatzas, I., and Shreve, S.E., (1991), "*Brownian Motion and Stochastic Calculas* (2 ed.)," Volume 113 of *Graduate Texts in Mathematics.* Berlin: Springer-Verlag.

[35] Keim, D.B., and Stambaugh, R.F., (1986), "Predicting Stock Returns in the Stock and the Bond Market," *J. Financial Econ.*, 17, 357-390.

[36] Kiku, D., (2006), "Long Run Risks and the Value Premium Puzzle," Working Paper, Duke University.

[37] Lamont, O., (1998), "Earnings and Expected Returns," *Journal of Finance*, 53, 1563-1587.

[38] Lettau, M., and Ludvigson, S., (2001), "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance*, 56, 815-849.

[39] Lettau, M., and Ludvigson, S.C., (2002), "Measuring and Modeling Variation in the Risk-Return Trade-Off," In: Ait-Sahalia, Y., and Hansen, L.P. (Eds.), *Handbook of Financial Econometrics.* North-Holland, Holland.

[40] Lettau, M., and Ludvigson, S., (2005), "Euler Equation Errors," Working Paper, New York University.

[41] Malloy, C., Moskowitz, T., and Vissing-Jorgensen, A., (2004), "Long-Run Stockholder Consumption Risk and Asset Returns," Working Paper, University of Chicago.

[42] Menzly, L., Santos, T., and Veronesi, P., (2004), "Understanding Predictability," *Journal of Political Economy*, 112(1), 1-47.

[43] Nelson, D.B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59 (2), 347-370.

[44] Parker, J.A., and Julliard, C., (2005), "Consumption Risk and the Cross Section of Expected Returns," *Journal of Political Economy*, 113(1), 185-222.

[45] Pakes, A., and D. Pollard (1989), "Simulation and the asymptotics of optimization estimators," Econometrica 57, 1027-1057.

[46] Protter, P., (2004), "Stochastic Integration and Differential Equations," New York: Springer-Verlag.

[47] Santos, T., and Veronesi, P., (2006), "Labor Income and Predictable Stock Returns," *Review of Financial Studies*, 19, 1-44.

[48] Whitelaw, R.F., (1994), "Time Variations and Covariations in the Expectation and Volatility of Stock Market Returns," *Journal of Finance*, 49, 515-541.

A Appendix

For the proof of Lemma 1 we make the following assumptions:

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

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

2. $\left[\frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t_j}^2 - p \lim_{N_t \to \infty} \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma_{t_j}^2 \right] = O_p(\frac{1}{N_t})$

3. Let $N = T^{-1} \sum_{t=1}^{T} N_t$, $\overline{N} = \max_{1 \le t \le T} N_t$ and $\underline{N} = \min_{1 \le t \le T} N_t$. $0 < \inf_T(\underline{N}/\overline{N}) < \sup_T(\overline{N}/\underline{N}) < \infty$. $N = T^{\gamma}$ for some $\gamma > 0$

4. $\left\{ \sigma_{t_j}^2 (\eta_{t_j}^2 - 1) \right\}_{j=1,t=1}^{N_t,T}$ is a strictly stationary stochastic process with finite kth moment, $k > 3$, and exponentially decaying $\alpha - mixing$ coefficient, $\alpha(k) = \exp\{-ck\}$.

A.1 Proof of Lemma 1

We have

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

Hence,

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

Consider the first term

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

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

provided $\gamma > \epsilon$, by assumption 1.

Hence,

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

Thus, the second term is of smaller order than the first and

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

Now, by the Bonferroni inequality

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

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:

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

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

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

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

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

A.2 Proof of Theorem 1

Asset pricing models imply the moment restriction

$$E\left[G(X_t, \sigma_t^2, \sigma_{t-1}^2, \sigma_{t-2}^2, \theta_0)\right] = 0,$$

where X_t are observed variables and θ is a p-vector of parameters, where $p \le q$, with true value θ_0.

The infeasible GMM estimator minimizes

$$\widetilde{\theta} = \arg\min_{\theta\in\Theta\subseteq R^p} G_T(\theta)' W_T G_T(\theta),$$

where $\{W_T\}_{T=1}^{\infty}$ is a sequence of positive definite weighting matrices.

The feasible GMM estimator minimizes

$$\widehat{\theta} = \arg\min_{\theta\in\Theta\subseteq R^p} \widehat{G}_T(\theta)' W_T \widehat{G}_T(\theta),$$

where

$$\widehat{G}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} G(X_t, \widehat{\sigma}_t^2, \widehat{\sigma}_{t-1}^2, \widehat{\sigma}_{t-2}^2, \theta_0).$$

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

Assumptions:

1. Define Estimator. For some set $\Theta \subseteq \mathbb{R}^p$,

$$\|G_T(\theta_T)\|_W = \inf_{\theta\in\Theta} \|G_T(\theta)\|_W + o_p(1),$$

2. Identification. $\|\overline{G}(\theta_0)\| = 0$, and for all $\delta > 0$ there exists $\epsilon > 0$ such that

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

3. ULLN.

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

4. The first four partial derivatives of G with respect to θ_j, $j = 1, \ldots, p$ and σ_t^2 exist and satisfy dominance conditions, namely for all vectors ν (pertaining to (σ_t^2, θ)) with $|\nu| \leq 4$, and for some sequence $\delta_T \to 0$,

$$\sup_{|x|,|x'|,|x''|\leq\delta_T} \sup_{\theta\in\Theta} \left\| D^\nu G(X_t, \sigma_t^2 + x, \sigma_{t-1}^2 + x', \sigma_{t-2}^2 + x'', \theta)\right\| \leq U_t,$$

where $EU_t < \infty$.

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

Assumptions

1. $\|G_T(\theta_T)\|_W = \inf_\theta \|G_T(\theta)\|_W + o_p(1/\sqrt{T})$;
2. The matrix
$$\Gamma(\theta) = \frac{\partial}{\partial\theta}\overline{G}(\theta)$$
is continuous in θ and is of full (column) rank at $\theta = \theta_0$.
3. For all sequences of positive numbers δ_T such that $\delta_T \to 0$,
$$\sup_{\|\theta-\theta_0\|\leq\delta_T} \|G_T(\theta) - \overline{G}(\theta)\|_W = O_p(1/\sqrt{T});$$
$$\sup_{\|\theta-\theta_0\|\leq\delta_T} \|\sqrt{T}[G_T(\theta) - \overline{G}(\theta)] - \sqrt{T}[G_T(\theta_0) - \overline{G}(\theta_0)]\|_W = o_p(1);$$
4. $\sqrt{T}G_T(\theta_0) \Longrightarrow N(0, \Omega)$
5. θ_0 is in the interior of Θ.

6. the daily processes $\{\sigma_{t_j}\}_{j=1,t=1}^{N_t,T}$ and $\{\eta_{t_j}\}_{j=1,t=1}^{N_t,T}$ are independent of the monthly/quarterly processes X_t and σ_t^2.

Under these 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

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

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,

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

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,

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

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

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

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,

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

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 is also $o_p(T^{-1/2})$ under the same conditions as the fourth term.

Hence,

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

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

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

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 $\frac{N^x}{T} \to \infty$ where $x > 1.5$. 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,

$$\sqrt{T}\left(\widehat{\theta}-\theta_0\right)\xrightarrow{d}N(0,\Sigma),\qquad \text{where } \Sigma=(\Gamma'W\Gamma)^{-1}\Gamma'W\Omega W\Gamma(\Gamma'W\Gamma)^{-1}.$$

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 (27). Define the estimated bias function

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

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

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

■

		mean	variance	skewness	kurtosis	AR(1)	AR(1-12)	T
	28:01-05:12	0.005	0.003	-0.478	9.819	0.102	0.209	936
$(r_{mt} - r_{ft})^{mon}$	28:01-66:12	0.006	0.004	-0.370	9.915	0.123	0.294	468
	67:01-05:12	0.004	0.002	-0.758	5.758	0.062	0.062	468
	28:01-05:12	0.002	2.2×10^{-5}	6.327	58.60	0.563	4.441	936
$(\widehat{v}_t)^{mon}$	28:01-66:12	0.003	3.3×10^{-5}	4.413	29.15	0.656	5.248	468
	67:01-05:12	0.002	1.0×10^{-5}	13.08	229.5	0.226	1.239	468
	28:01-05:04	0.014	0.012	0.175	9.865	-0.051	-0.141	312
$(r_{mt} - r_{ft})^{quar}$	28:01-66:04	0.017	0.016	0.456	9.950	-0.086	-0.095	156
	67:01-05:04	0.011	0.008	-0.762	4.155	0.013	-0.286	156
	28:01-05:04	0.007	0.0001	4.268	25.24	0.554	4.307	312
$(\widehat{v}_t)^{quar}$	28:01-66:04	0.009	0.0002	3.285	15.55	0.601	4.755	156
	67:01-05:04	0.005	4.6×10^{-5}	6.139	54.84	0.263	1.042	156

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. Estimates are reported for the monthly, quarterly, 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. Our market proxy is the CRSP value-weighted index (all stocks). The proxy for the riskfree rate is the one-month Treasury Bill rate (from Ibbotson Associates). The table shows the mean, variance, skewness, kurtosis, first-order serial correlation, 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.

		α	β
	1928:01-2005:12	0.006	-0.493
		(1.896)	(-0.340)
monthly	1928:01-1966:12	0.007	-0.288
		(1.648)	(-0.186)
	1967:01-2005:12	0.009	-3.071
		(1.195)	(-0.711)
	1928:1-2005:4	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	-0.021	5.904
		(-1.130)	(1.810)

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

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

		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)

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 and the J-stat for overidentifying restrictions.

		α	β
	1928:01-2005:12	-0.014	7.991
		(-0.698)	(0.919)
monthly	1928:01-1966:12	-0.006	3.795
		(-0.325)	(0.654)
	1967:01-2005:12	-0.009	7.066
		(-0.210)	(0.297)
	1928:1-2005:4	-0.145	22.24
		(-0.404)	(0.439)
quarterly	1928:1-1966:4	-0.035	5.747
		(-0.258)	(0.385)
	1967:1-2005:4	0.034	-4.520
		(1.472)	(-1.077)

Table 4. This table shows the GMM estimates for the model $E[G(.)] = 0$ where

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

		α	β
	1928:01-2005:12	0.003	0.719
		(0.798)	(0.384)
monthly	1928:01-1966:12	0.005	0.161
		(1.221)	(0.085)
	1967:01-2005:12	-0.075	44.74
		(-1.079)	(1.130)
	1928:1-2005:4	0.008	0.884
		(0.586)	(0.402)
quarterly	1928:1-1966:4	0.013	0.443
		(0.866)	(0.198)
	1967:1-2005:4	-0.331	64.23
		(-0.468)	(0.483)

Table 5. This table shows the GMM estimates for the model $E[G(.)] = 0$ where

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

		α	β
	1928:01-2005:12	-0.006	4.389
		(-0.711)	(1.272)
monthly	1928:01-1966:12	-0.011	5.731
		(-0.581)	(0.888)
	1967:01-2005:12	-0.100	58.47
		(-0.415)	(0.424)
	1928:1-2005:4	-0.020	4.764
		(-0.695)	(1.157)
quarterly	1928:1-1966:4	-0.056	8.125
		(-0.564)	(0.771)
	1967:1-2005:4	-3.982	751.0
		(-0.024)	(0.024)

Table 6. This table shows the GMM estimates for the model $E[G(.)] = 0$ where

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

		α	β
	low vol	-0.008	25.97
		(-1.298)	(3.485)
monthly	high vol	-0.016	1.019
		(-1.449)	(0.476)
	low vol	-0.035	26.36
		(-1.398)	(2.842)
quarterly	high vol	-0.032	0.491
		(-0.758)	(0.173)

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

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

The table reports the coefficient estimates along with the associated t-stats in parentheses. The statistics are shown for two subsamples that are chosen to correspond to low and high volatility periods (the sample is split based on the higher one-third quantile of the volatility estimate).

		α	β
	low vol	-0.022	46.48
		(-2.287)	(3.409)
monthly	high vol	0.0004	-2.067
		(0.0136)	(-0.222)
	low vol	-0.030	24.03
		(-1.247)	(2.402)
quarterly	high vol	-0.538	55.51
		(-0.752)	(0.783)

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

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

The table reports the coefficient estimates along with the associated t-stats in parentheses. The statistics are shown for two subsamples that are chosen to correspond to low and high volatility periods (the sample is split based on the higher one-third quantile of the volatility estimate). The sample extends from 1955:01-2005:12.