Ghosh, Anisha; Linton, Oliver Bruce

Working Paper

Estimation with mixed data frequencies: A bias-correction approach

cemmap working paper, No. CWP65/19

Provided in Cooperation with:
Institute for Fiscal Studies (IFS), London

Suggested Citation: Ghosh, Anisha; Linton, Oliver Bruce (2019) : Estimation with mixed data frequencies: A bias-correction approach, cemmap working paper, No. CWP65/19, Centre for Microdata Methods and Practice (cemmap), London, https://doi.org/10.1920/wp.cem.2019.6519

cemmap

centre for microdata methods and practice

Estimation with Mixed Data Frequencies: A Bias-Correction Approach

Anisha Ghosh
Oliver Linton

The Institute for Fiscal Studies
Department of Economics,
UCL

cemmap working paper CWP65/19

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Estimation with Mixed Data Frequencies: A Bias-Correction Approach *

Anisha Ghosh† Oliver Linton‡

November 17, 2019

Abstract

We propose a solution to the measurement error problem that plagues the estimation of the relation between the expected return of the stock market and its conditional variance due to the latency of these conditional moments. We use intra-period returns to construct a nonparametric proxy for the latent conditional variance in the first step which is subsequently used as an input in the second step to estimate the parameters characterizing the risk-return tradeoff via a GMM approach. We propose a bias-correction to the standard GMM estimator derived under a double asymptotic framework, wherein the number of intra-period returns, N, as well as the number of low frequency time periods, T, simultaneously go to infinity. Simulation exercises show that the bias-correction is particularly relevant for small values of N which is the case in empirically realistic scenarios. The methodology lends itself to additional applications, such as the empirical evaluation of factor models, wherein the factor betas may be estimated using intra-period returns and the unexplained returns or alphas subsequently recovered at lower frequencies.

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

JEL Classification Codes: C14, G12

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

†Desautels Faculty of Management, McGill University; anisha.ghosh@mcgill.ca

‡University of Cambridge; obl20@cam.ac.uk.

I Introduction

For a number of central questions in financial economics, the use of mixed data frequencies enables more effective utilization of all available data, thereby offering the promise of sharper inference. A prominent example is the estimation of the relation between the expected excess return on the stock market and its conditional variance (hereafter referred to as the risk-return tradeoff). This risk-return relation is typically estimated at monthly, quarterly, annual, or even lower frequencies. However, a critical input to this estimation problem, namely the (latent) conditional variance, can be estimated more efficiently using higher frequency intra-period (e.g., daily) data. Therefore, using mixed data frequencies, wherein the market variance at lower frequencies is estimated using higher frequency returns and the risk-return relation at lower frequencies is subsequently recovered using the estimated market variance, seems attractive. A second example involves the empirical evaluation of factor models for the cross section of asset returns. Just as with the risk-return tradeoff, the performance evaluation of the underlying risk factors in factor models are often conducted at monthly or lower frequencies. Therefore, the factor betas for a chosen set of test assets can be first computed using intra-period returns, and the alphas generated by the factor model can then be obtained using these estimated betas as inputs.

The goal of this paper is to develop an asymptotic theory for such inference methods involving mixed data frequencies. A large literature in financial econometrics has studied the limiting behavior of variance (and covariance) estimators using high frequency data. These studies rely on so-called *infill* asymptotics, whereby the number of intra-period observations within a finite time horizon is assumed to grow. Our approach, on the other hand, relies on a double asymptotic framework, where both the number of intra period observations (n) within a lower frequency time period as well as the number of lower frequency periods (T) are simultaneously assumed to be large. Moreover, in many empirically realistic scenarios, N may be small relative to T. Examples include recovering the risk-return relation at the monthly frequency over the entire post war sample 1947–2018: in this case $N = 22$ (corresponding to daily data within a month) and $T = 864$ months; or the evaluation of factor models over long sample periods. Higher frequency than daily data are not available for such long sample periods, yet assessing the performance of models over long time periods is often considered crucial to help shed light on their relative strengths and weaknesses. In such scenarios, the estimation errors in the estimated market variance or the factor betas may be non-trivial. Our methodology proposes a bias-correction to improve the performance of the estimators in such cases.

We illustrate our methodology in the context of recovering the risk-return tradeoff on the aggregate equity market portfolio. 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. Despite its central importance to both theory and practice, the empirical evidence on the risk-return relation is mixed and inconclusive. Bollerslev, Engle, and Wooldridge (1988), Harvey (1989), Harrison and Zhang (1999), Ghysels, Santa-Clara, and Valkanov (2005), Lundblad (2005), Ludvigson and Ng (2007), and Pastor, Sinha, and Swaminathan (2008) find a positive risk-return relation, while Campbell (1987), Glosten, Jagannathan, and Runkle (1993), Whitelaw (1994), Harvey (2001), and Brandt and Kang (2004) find a negative relation. Still others find mixed and inconclusive evidence like French, Schwert, and Stambaugh (1987), Nelson (1991), and Campbell and Hentschel (1992).[1]

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. The results have been found to be very sensitive to the particular model specification. Other studies have typically measured the conditional expectations underlying the conditional mean and conditional variance as projections onto predetermined conditioning variables. Practical constraints, such as choosing among a few conditioning variables, introduce an element of arbitrariness into the econometric modeling of expectations and can lead to omitted information estimation bias.

More recently, researchers have proposed a nonparametric proxy for the latent expost variance, namely the integrated variance, that is void of any specific functional form assumptions about the stochastic process generating returns.[2] The integrated variance is unbiased for the latent conditional variance and, although latent, it may be consistently estimated using the *realized variance* that is computed as the sum of squares of high-frequency intraperiod returns. Proxying the latent integrated variance with the realized variance has been used in French, Schwert, and Stambaugh (1987) and Bandi and Perron (2008) in estimating the risk-return relation. But these studies have ignored the measurement error that arises because of the use of realized variance as a proxy for the integrated variance. The goal of the current paper is to offer an explicit solution to the measurement error problem, that is likely

[1] Bandi and Perron (2008) find that the relation is difficult to detect at short horizons, but becomes strong at longer horizons of six to ten years.

[2] For an excellent survey of this extensive literature, see Andersen, Bollerslev, and Diebold (2002). See also Barndorff-Nielsen and Shepherd (2002), Andersen, Bollerslev, Diebold, and Labys (2003), and Ait-Sahalia and Jacod (2014).

to be particularly relevant when the number of intra period observations used to compute the realized variance is modest.

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 to estimate the risk-return parameters in a general GMM setting. We derive the limiting distribution of the estimated parameters under this double asymptotic framework. We find that under fairly strong conditions on n and T, the estimates are $\sqrt{T}$ -consistent and have the standard distribution as when there is no measurement-error. However, if the above condition is not satisfied, there is an asymptotic bias that would invalidate this approximation. In that case, we find that under weaker conditions on n and T, a bias-corrected estimator has the standard limiting distribution. This improvement is particularly relevant in the empirical case we examine where n is quite modest (e.g., daily returns within a month or quarter).

The above is an important methodological contribution to the extant literature on high-frequency volatility estimation. Most work has currently been about just estimating that quantity itself and using it to compare discrete time models in settings where the noise is small. Our approach is concerned with small sample issues when using estimated realized volatility as regressors in the estimation of parameters associated with the unobserved quadratic variation. This involves a useful extension of the existing asymptotic results for realized volatility[3] 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.

Our paper is related to Bollerslev and Zhou (2002) who propose estimating the parameters of stochastic volatility models by matching the sample moments of realized volatility to the corresponding population moments of integrated volatility using a GMM approach. Corradi and Distaso (2006) extend Bollerslev and Zhou (2002) by providing sufficient conditions under which the measurement error in the realized variance can be ignored asymptotically, in a double asymptotic framework, that simultaneously allows n and T to approach infinity. Todorov (2009) further extends the framework to allow for infinite activity price jumps that can exhibit arbitrary time-variation. The analyses in these papers differs from ours in two notable respects. First, they consider precise functional-form specifications for the stochastic volatility process that depend on a vector of unknown parameters to be estimated, while our approach does not take a stance on the specification of the volatility process. Second, their GMM estimation strategy is formally justified under the assumption that the number of high frequency intra period observations, n, used to estimate the latent integrated variance

[3] See Barndorff-Nielsen and Shepherd (2002).

converges to infinity faster than the number of lower frequency observations, T, used to calculate the sample moments of the variance. Such an assumption is unrealistic in our setting, where the number of (daily) returns used to compute the (e.g., monthly, quarterly) realized variance is, in fact, substantially smaller than the number of data points used to compute the sample moments underlying the risk-return relation.

Bollerslev, Patton, and Quaedvlieg (2016) also recognize the role of measurement error in the realized variance and point out that is biases the estimated autoregressive coefficient in forecasting regressions for the variance. They recommend allowing for a time-varying autoregressive coefficient that takes into account the temporal variation in the measurement error based on the asymptotic distribution theory for realized variance. Our analysis differs from this study in that it focuses on the risk-return relation and provides an explicit expression for the bias in the estimated coefficient of the risk-return relation that arises because of the time-varying measurement error in the realized variance under fairly general assumptions, along with a feasible estimator for the bias.

The uniformity of the estimation error in the realized variance necessary in our double asymptotic framework is also invoked in Kanay and Kristensen (2016) who focus on a two-step approach to estimating stochastic volatility models, whereby a spot volatility estimated in step one is plugged into a given existing estimation method for a fully observed diffusion model in step two. The uniformity of the approximation error is also invoked in Li and Patton (2015) to evaluate the forecast performance for a board set of latent risk measures, such as volatility, beta, correlation, or jump variation. Our paper goes further to show how this result can be used to obtain a bias-corrected estimator for the risk-return relation.

In the empirical analysis, we focus on the risk-return relation at the monthly frequency. We use n daily returns of the CRSP value-weighted stock market index to obtain monthly estimates of the realized variance. We then estimate the parameters of the risk-return relation using the GMM approach with T monthly observations on the realized excess market returns and realized variance. We find a negative relation between the mean and the variance that is statistically significant over the entire post war sample $1947-2018$. Moreover, we find that the bias-correction that we propose is instrumental in delivering the strongly statistically significant results. This finding is robust to the choice of instruments. Upon inclusion of the lagged variance and the lagged market return as additional regressors, we obtain a positive relation between the conditional mean and variance of the market return. This is consistent with the findings in Lettau and Ludvigson (2010) who highlight the difference between the unconditional correlation between the expected market return and its variance and the conditional (on the lagged return and variance) correlation.

The remainder of the paper is organized as follows. The econometric framework and estimation methodology are described in Section II. Section III derives the asymptotic properties of the GMM estimator in the presence of measurement error. In Section IV, we perform Monte-Carlo simulations to examine the finite-sample performance of the estimator. Section V presents the empirical results and Section VI concludes with possible directions for future research. The Appendix contains the proofs of our main results.

II Estimation Methodology

II.1 Model and Hypothesis

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

$$E\left(r_{m,t} - r_{f,t} | \mathcal{F}_{t-1}\right) = b_0 + b_1\left(x_{t-1}\right) \operatorname{var}\left(r_{m,t} | \mathcal{F}_{t-1}\right) + b_2^{\intercal} Z_{t-1}, \tag{1}$$

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

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

$$E\left[r_{m,t} - r_{f,t} - b_0 - b_1\left(x_{t-1}\right) \operatorname{var}\left(r_{m,t} | \mathcal{F}_{t-1}\right) - b_2^{\intercal} Z_{t-1} | \mathcal{F}_{t-1}\right] = 0. \tag{2}$$

The parameters in the above moment restriction can be estimated using a standard GMM approach with a long time series of observations on $r_{m,t}$, $r_{f,t}$, Z_{t-1}, x_{t-1}, and $\operatorname{var}\left(r_{m,t} | \mathcal{F}_{t-1}\right)$. Note that Z_{t-1} and x_{t-1} may include lagged values of the conditional variance, and the function $b_1(x_{t-1})$ may be potentially nonlinear. The main difficulty in the estimation process arises because of the unobservability of the conditional variance, $\text{var}(r_{m,t}|\mathcal{F}_{t-1})$, and its lags. Our strategy is to replace this quantity by a feasible approximately unbiased nonparametric estimator computed from higher frequency data. We next describe the framework in which this procedure makes sense.

We describe here our general approach to estimating volatility from higher frequency data and how it fits in with the low frequency testing strategy. We suppose that we observe high frequency returns r_{t_j}, $j = 1, \ldots, n_t$ for each t (where $t = 1, \ldots, T$). We suppose that they are generated by the following (sequence of) discrete-time model(s)

$$r_{t_j} = n_t^{-1}\mu_{t_j} + n_t^{-1/2}\sigma_{t_j}\eta_{t_j}, \tag{3}$$

where η_{t_j} is stationary and ergodic and, furthermore, η_{t_j} and $\eta_{t_j}^2 - 1$ are martingale difference sequences with respect to $\mathcal{F}_{t_{j-1}}$, where $\mathcal{F}_{t_{j-1}}$ contains all information upto time t_{j-1}, including μ_{t_j}, σ_{t_j}. 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 and volatility feedback effects in intra period returns. In particular, η_{t_j} can affect $\sigma_{s_{j+k}}$ for $s = t$, $k \geqslant 1$ and $s > t$, $k \geqslant 0$. We do not assume Gaussianity for the innovation process, so that the conditional distribution of returns can be heavy tailed. This framework is broadly consistent with observed returns being the discretized approximation to the continuously compounded returns $r_{t_j}^* = p_{t_j}^* - p_{t_{j-1}}^*$, where the true underlying efficient log-price p^* follows the continuous time diffusion

$$dp_t^* = \mu(p_t^*)dt + \sigma(p_t^*)dW_t, \tag{4}$$

for functions $\mu(.)$, $\sigma(.)$, and Brownian motion W. Clearly, if $\mu(.) \equiv 0$ and $\sigma(.) = \sigma$ (a constant), we have $p_t^* = \sigma W_t$, so that $r_{t_j}^*$ are independent and normally distributed and $r_{t_j} = r_{t_j}^*$ so that $\eta_{t_j} \sim N(0,1)$ and are i.i.d. More generally, one can show (under some conditions) that, with probability one, $r_{t_j} = r_{t_j}^* + o(n_t^{-\rho})$ for some $\rho > 1$ (Euler, Milstein approximations; see, e.g., Gonçalves and Meddahi (2009), Mykland and Zhang (2009)).[4] The process (3) is consistent with a stochastic volatility process as in Gonçalves and Meddahi (2009, section 4): in their case, high frequency returns are mutually independent but heterogeneous, conditioning on the drift and volatility functions. Our process can be also seen as an example of the discrete time approximations developed in Nelson (1990) where we replaced his generic sequence h by the specific one n_t^{-1}. For example, compare (3) with his expression 2.22 (with

[4]We will use the continuous time theory to justify some of our methodology. We recognize that the approximation we make here in principle could affect our results, but remark that the more complicated higher order approximations to the discrete time process may not necessarily work better in practice.

$c = 0)$ ${}_h r_{(k-1)h:kh} =_h \sigma_{kh} \times_h Z_{kh}$, where ${}_h Z_{kh} \sim N(0, h)$ and ${}_h\sigma^2_{(k+1)h}$ has some particular dynamic specification. Under appropriate conditions, special cases of our process can be shown to converge to a stochastic volatility process (and our estimator below would converge to the quadratic variation of that limiting diffusion process).

We do not explicitly allow for jumps in our stochastic process: we are treating large observations through the traditional discrete time lens, whereby a heavy-tailed distribution for η_{t_j} would lead frequently to large values of r_{t_j}, that captures some aspects of the continuous time notion of a jump. We also do not explicitly allow for microstructure noise in the observed prices since high frequency observations refer to daily data in our application – a frequency at which microstructure effects are arguably negligible.

We next define the ex-post measure of return variation for period t in this framework. We assume that the following probability limit exists uniformly in t:

$$\sigma_t^2 \equiv p \lim_{n_t \to \infty} \overline{\sigma}_t^2, \text{ where } \overline{\sigma}_t^2 = \frac{1}{n_t} \sum_{j=1}^{n_t} \sigma_{t_j}^2, \tag{5}$$

where σ_t^2 can be stochastic, and that the convergence occurs so fast that the error term from replacing σ_t^2 by $\overline{\sigma}_t^2$ is negligible. If the underlying model were the diffusion process (4), then $\sigma_t^2 = \int_0^1 \sigma^2(t-1+s)ds$ is the integrated variance and the approximation in Equation (5) is indeed good.

The integrated variance is approximately an (ex-ante) unbiased estimator for the conditional variance (see, e.g., Protter (2004)), so that

$$\text{var}\,(r_{m,t}|\mathcal{F}_{t-1}) \approx E[\sigma_t^2|\mathcal{F}_{t-1}],$$

with strict equality if $\mu(.) \equiv 0$ in Equation (4). The unbiasedness property of the integrated variance gives us the following *infeasible* moment restriction:

$$E\left[r_{m,t} - r_{f,t} - b_0 - b_1\left(x_{t-1}\right)\sigma_t^2 - b_2^{\intercal} Z_{t-1}|\mathcal{F}_{t_{-1}}\right] = 0. \tag{6}$$

We will use this moment condition as the basis for estimation.

II.2 Estimation Procedure

We concentrate on the following realized variance estimator computed from the high frequency intra period returns:

$$\widehat{\sigma}_t^2 = \sum_{j=1}^{n_t} r_{t_j}^2. \tag{7}$$

In the diffusion case, the theory of quadratic variation implies that the realized variance provides a consistent nonparametric measure of the integrated variance (see, e.g., Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2002)): $p\lim_{n_t\to\infty}\widehat{\sigma}_t^2 = \sigma_t^2$, where the convergence is uniform in probability (over $t = 1, \ldots, T$). Also, Jacod (1994), Jacod and Protter (1998), and Barndorff-Nielsen and Shephard (2002) develop the following asymptotic distribution theory for realized variance as an estimator of the integrated variance: $n_t^{1/2}(\widehat{\sigma}_t^2 - \sigma_t^2) \Rightarrow \sqrt{2}(\int_0^1 \sigma^2(t-1+s)dB(t-1+s))$ as $n_t \to \infty$, where B is a Brownian motion that is independent of W in Equation (4) and the convergence is in law stable as a process. This result implies that $n_t^{1/2}(\widehat{\sigma}_t^2 - \sigma_t^2) \Longrightarrow MN(0, 2\int_0^1 \sigma^4(t-1+s)ds)$, where MN denotes a mixed Gaussian distribution. Barndorff-Nielsen and Shephard (2002) showed that the above result can be used in practice as the integrated quarticity $IQ_t \equiv \int_0^1 \sigma^4(t-1+s)ds$ can be consistently estimated using $(1/3)RQ_t$, where

$$RQ_t = n_t \sum_{j=1}^{n_t} r_{t_j}^4.$$

It further follows that $(1.5RQ_t^{-1}n_t)^{1/2}(\widehat{\sigma}_t^2 - \sigma_t^2) \Longrightarrow N(0,1)$. This is a nonparametric result as it does not require the specification of the form of the drift, $\mu(.)$, or the diffusion, $\sigma(.)$, in Equation (4). The integrated quarticity plays an important role in our bias correction procedure below under our model assumptions.

Under the model (3), we have by the martingale CLT (Hall and Heyde (1980, Corollary 3.1),

$$n_t^{1/2}(\widehat{\sigma}_t^2 - \overline{\sigma}_t^2) \Longrightarrow MN(0, v_t),$$

where $v_t = p\lim_{n_t\to\infty} \frac{1}{n_t}\sum_{j=1}^{n_t} \sigma_{t_j}^4 \vartheta_{t_j}$, where $\vartheta_{t_j} = E[\eta_{t_j}^4|\mathcal{F}_{t_{j-1}}] - 1$. Under some additional conditions, $RQ_t \to \lim_{n_t\to\infty} \frac{1}{n_t}\sum_{j=1}^{n_t} \sigma_{t_j}^4 E[\eta_{t_j}^4|\mathcal{F}_{t_{j-1}}]$. Under stronger conditions (for example, suppose that $\sigma_{t_j}^2$ are deterministic or stochastic but independent of the process $\{\eta_{t_j}\}$), then, $v_t = \sigma_t^4(\kappa - 1)$, where $\kappa = E[\eta_{t_j}^4]$. If η_{t_j} were standard Gaussian, which follows from (4), then $\kappa = 3$. Furthermore, in this case we will have $((\kappa-1)RQ_t/\kappa n_t)^{-1/2}(\widehat{\sigma}_t^2 - \sigma_t^2) \Longrightarrow N(0,1)$. We do not assume Gaussianity for our innovation process, although we do make this assumption to define a simple bias correction method.

Plugging the realized variance into the infeasible moment restriction (6), we obtain the *feasible* moment restriction:

$$E\left[r_{m,t} - r_{f,t} - b_0 - b_1(x_{t-1})\widehat{\sigma}_t^2 - b_2^\intercal Z_{t-1}|\mathcal{F}_{t-1}\right] = 0. \quad (8)$$

Finally, with a set of chosen instruments, Y_{t-1} (that could include, for instance, lagged variances), we have the unconditional moment restrictions:

$$E\left[\left(r_{m,t}-r_{f,t}-b_0-b_1\left(x_{t-1}\right)\widehat{\sigma}_t^2-b_2^\intercal Z_{t-1}\right)Y_{t-1}\right]=0. \tag{9}$$

Defining $Y_t \equiv \left(y_t', \sigma_t^2, \sigma_{t-1}^2, \ldots, \sigma_{t-p}^2\right)$, $Z_t \equiv \left(z_t', \sigma_t^2, \sigma_{t-1}^2, \ldots, \sigma_{t-p}^2\right)$, where y_t' and z_t' are the observable components of Y_t and Z_t, respectively, $X_t \equiv \left(r_{m,t}, r_{f,t}, y_t', z_t', x_t\right)$, $V_t = (\sigma_t^2, \sigma_{t-1}^2, \ldots, \sigma_{t-p}^2)^\intercal$, and $\widehat{V}_t = (\widehat{\sigma}_t^2, \widehat{\sigma}_{t-1}^2, \ldots, \widehat{\sigma}_{t-p}^2)^\intercal$, we can rewrite the feasible moment restriction as:

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

where $\theta = (b_0, b_1^\intercal, b_2^\intercal)^\intercal$ with true value θ_0.[5] The above set of moment restrictions are expressed entirely in terms of observable variables and, therefore, the parameter vector θ may be estimated using the GMM approach. Specifically, we define the estimator $\widehat{\theta}_T \in \Theta$ as the minimizer of

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

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

III Asymptotic Properties

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 $t = 1, 2, \ldots, T$. Empirically, n_t is really only moderate size (e.g., daily in most empirical asset pricing applications where models need to be estimated over long time periods when higher frequency data were not available) and so the quality of the asymptotic approximation is likely to be an issue. We show how to address this issue by providing a bias correction method that improves the approximation error.

We first present a lemma that involves a useful extension of the existing asymptotic results obtained for realized volatility in Barndorff-Nielsen and Shephard (2002). This lemma is

[5] We choose to present a theory for a general GMM estimator, rather than for the precise moment conditions in Equation (9). This is because our theory is sufficiently general and amenable to other applications, in addition to the linear (in variance) moment specification considered in this paper. For instance, our approach can accommodate potential nonlinearities in the relation between the conditional means and variances of stock returns – an implication of several popular equilibrium theories. Similarly, an extension of our framework can be used to estimate conditional factor pricing models like the conditional Capital Asset Pricing Model or the conditional Fama-French three and five factor models. In the latter extension, the underlying latent variables are the conditional betas (multivariate for multifactor models) that could be replaced with realized betas to make the estimation feasible.

concerned with the uniformity of the estimation error. The existing financial econometrics literature on nonparametric volatility estimation has focused on estimating financial market volatility over a finite time horizon, typically daily or monthly. In these applications, it suffices to establish consistency of the estimator over the finite time interval. In our present application, however, the number of finite-length time periods tends to infinity, thereby requiring a stronger consistency result. In this paper, we apply the methodology to estimate the empirical risk-return relation. However, the results are considerably general and might be useful in other contexts that require financial market variance estimation over successive time periods.

Our first result establishes the consistency of $\widehat{\sigma}_t^2$ for σ_t^2, uniformly in t. To derive the result, we make the following regularity assumptions.

Assumptions A

1. *The process μ_{t_j} is uniformly bounded. 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. *For some $\gamma > \max\left\{\frac{2}{k-1}, 2\epsilon\right\}$*, where ϵ and k are as in Assumptions A1 and A4, respectively,

$$n_t = O(T^{\gamma}) \text{ for all } t$$

3. *For $\theta \geqslant 2$, there exists a probability limit σ_t^{θ} for each t such that,*

$$\max_{1 \le t \le T} \left| \frac{1}{n_t} \sum_{j=1}^{n_t} \sigma_{t_j}^{\theta} - \sigma_t^{\theta} \right| = O_p(n_t^{-\lambda})$$

for some $\frac{1}{2}\left(1 - \frac{\epsilon}{\gamma}\right) < \lambda < 1$.

4. *The process $\{\eta_{t_j}\}_{j=1,t=1}^{n_t,T}$ is stationary and ergodic and has finite k^{th} moment for some large $k > 6$.*

5. *The process $\{\eta_{t_j}\}_{j=1,t=1}^{n_t,T}$ is i.i.d with mean zero, variance one, and finite k^{th} moment for some large $k > 6$. Let $\kappa = E\left(\eta_{t_j}^4\right)$.*

Remarks. (i) Condition A1 controls the behaviour of the volatility process over long time spans. One possibility is to require that the process $\sigma_{t_j}^2$ is uniformly bounded over all t and all j and all sample paths, but this is a little strong. Instead, we shall control the rate of growth of the maximum value this process can achieve over many periods. Let $m_t = \sum_{j=1}^{n_t} \sigma_{t_j}^4 / n_t$ denote the intraperiod second moment of volatilities. Suppose, for example, that the stochastic process m_t was stationary and Gaussian, then $\max_{1\leq t\leq T} m_t$ would grow to infinity at a logarithmic rate. We shall allow instead this process to grow at an algebraic rate that is much faster than logarithmic. Over the sample period $1947-2018$, daily excess market returns are highly leptokurtic with the degree of excess kurtosis being 16.5. The evidence of very fat tails in the distribution of returns highlights the importance of this assumption.

(ii) We are not assuming Gaussianity of η_{t_j} and we are not exploiting the structure of an underlying continuous time model so we need to make strong assumptions like A3. Note that this assumption is similar to Assumption H of Goncalves and Meddahi (2009) with the added feature that we need to control this error uniformly over the low frequency time span of our data. This assumption is consistent with many sample schemes. For example, suppose that $\sigma_{t_j}^{\theta}$ were *i.i.d.*, then we can argue that $\frac{1}{\sqrt{n_t}}\left(\sum_{j=1}^{n_t}\sigma_{t_j}^{\theta}-E\sigma_t^{\theta}\right)$ is asymptotically normal for each t. In that case, we would have to control the growth rate of $\max\limits_{1\leq t\leq T}\frac{Z_t}{\sqrt{n_t}}$, which can easily be shown to be of order $(\log T)^{1/2}/\sqrt{n_t}$, and so given our assumptions on the relative magnitude of T and n_t, this term is of smaller order in probability.

(iii) Since Assumption A4 is with regard to the standardized return series η_{t_j}, it is not so strong as requiring that returns themselves have many moments.

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

Lemma 1 *Suppose that Assumptions A1-A4 hold. Then, for* $\alpha<\frac{\gamma}{2}-\epsilon$, *we have*

$$T^{\alpha}\max_{1\leqslant t\leqslant T}(\widehat{\sigma}_t^2-\sigma_t^2)=o_p(1). \tag{10}$$

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

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

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

$$\sqrt{T}(\widetilde{\theta}_T-\theta_0)=-(\Gamma^{\top}W\Gamma)^{-1}\Gamma^{\top}W\sqrt{T}G_T(\theta_0)+o_p(1)\Longrightarrow N(0,\Sigma), \tag{11}$$

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

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

Assumptions B

1. *As* $T \to \infty$,

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

2. *The matrix* $\Gamma(\theta) = \frac{\partial}{\partial \theta^\top} \overline{G}(\theta)$ *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* $\delta_T \to 0$,

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

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

4. *As* $T \to \infty$,

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

5. *The true parameter* θ_0 *is in the interior of* Θ.
6. *For some* $\omega > 0$,

$$\sup_{T \ge 1} \frac{1}{T} \sum_{t=1}^{T} E\left|G(X_t, V_t; \theta_0)\right|^{2+\omega} < \infty$$

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

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

where $EU_t < \infty$.

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

The following theorem provides an asymptotic expansion for the estimator $\widehat{\theta}_T$. Appendix A.2 provides a formal proof of this result. Let $G_{\sigma_t^2\sigma_t^2}$ denote the second partial derivative of G with respect to σ_t^2.

Theorem 1 *Suppose that conditions A1-A4 and B are satisfied. Then,*

$$\widehat{\theta}_T - \theta_0 = -(\Gamma^\top W \Gamma)^{-1}\Gamma^\top W G_T(\theta_0) - (\Gamma^\top W \Gamma)^{-1}\Gamma^\top W b_T(\theta_0) + o_p(T^{-1/2}), \tag{12}$$

where

$$b_T(\theta_0) = \sum_{k=0}^{p} \frac{1}{2T} \sum_{t=1}^{T} E\left[G_{\sigma_{t-k}^2\sigma_{t-k}^2}(X_t, V_t; \theta_0)(\widehat{\sigma}_{t-k}^2 - \sigma_{t-k}^2)^2\right].$$

The first term on the right hand side of Equation (12) is the standard one that arises in a GMM procedure in the absence of any measurement error. The second term, on the other hand, is an additional bias term that arises because of the measurement error in realized variance as an estimator of the integrated variance. A few comments are in order for this bias term. First, note that the quantity $b_T(\theta)$ is of order $\frac{T^\epsilon}{n}$ in probability (based on Assumption A1). Therefore, its relative magnitude depends on the assumption we make connecting n and T, i.e. on the relative growth rates of n and T. The direction of the bias depends on the partial derivatives of the moment conditions. Second, the bias in non zero if and only if the second derivatives of the moment restrictions with respect to the market variances (contemporaneous or lagged) are non zero. For example, the seminal Capital Asset Pricing Model (CAPM) predicts a linear relationship between the conditional expected excess return of the stock market and its conditional variance, with the coefficient equal to the coefficient of risk aversion of the average investor in the economy. In this case, if the comtemporaneous integrated variance is used as an approximately unbiased estimator of the conditional variance, the moment restriction in Equation (6) reduces to $E\left[(r_{m,t} - r_{f,t} - b_0 - b_1\sigma_t^2)\,|\mathcal{F}_{t_{-1}}\right] = 0$, where b_1 is the risk aversion coefficient, and, therefore, Equation (12) implies that the bias term is identically equal to zero. On the other hand, if lags of the conditional variance are included as additional right hand side variables in the risk return relation – a specification advocated by Whitelaw (1994), Brandt and Kang (2004), and Ludvigson and Ng (2007) – then the bias term in non zero when lagged variances are also included in the set of instruments used in the estimation. Also, the bias term in non zero when there are non linearities in the risk return relation (e.g., a quadratic relation) – a specification implied by several prominent asset pricing models.

We obtain the following result.

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

$$\sqrt{T}(\widehat{\theta}_T - \theta_0) \Longrightarrow N(0, \Sigma). \tag{13}$$

Note that this requires $n^x/T \to \infty$, where $x > \frac{1}{\gamma}$ and $\gamma > \varepsilon + \frac{1}{2}$. This condition for the asymptotically negligibility of the bias can perhaps be more conveniently expressed as $T^{\varepsilon+\frac{1}{2}}/n \to 0$. If σ_{t_j} is assumed to be uniformly bounded, i.e. $\epsilon = 0$ in Assumption A1, then this condition reduces to $\sqrt{T}/n \to 0$, and is reminiscent of a similar condition in Bai and Ng (2006) under which they show that factor estimation uncertainty does not matter for factor-augmented models. When (13) holds, standard inference can be applied. Specifically, since $G(X_t, V_t; \theta_0)$ 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

$$\begin{aligned} \widehat{\Gamma} &= \frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \theta^\top} G\left(X_t, \widehat{V}_t; \widehat{\theta}_T\right) \\ \widehat{\Omega} &= \frac{1}{T} \sum_{t=1}^{T} G\left(X_t, \widehat{V}_t; \widehat{\theta}_T\right) G\left(X_t, \widehat{V}_t; \widehat{\theta}_T\right)^\top . \end{aligned}$$

When the condition in Corollary 1 is not satisfied, i.e., $b_T(\theta_0) \neq o(T^{-1/2})$, then the bias term in Equation (12) may not vanish asymptotically and, consequently, Equation (13) does not hold. Define the bias corrected estimator, $\widehat{\theta}_T^{bc}$, as

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

where

$$\widehat{b}_T(\widehat{\theta}_T) = \sum_{k=0}^{p} \frac{1}{2T} \sum_{t=1}^{T} G_{\sigma_{t-k}^2 \sigma_{t-k}^2}(X_t, V_t; \widehat{\theta}_T) \frac{\kappa - 1}{\kappa} \frac{RQ_{t-k}}{n_{t-k}},$$

and $RQ_t = n_t \sum_{j=1}^{n_t} r_{t_j}^4$. In this case, we have the following result.

Corollary 2 *Suppose that* $\sqrt{T}\widehat{b_T}(\widehat{\theta}_T) = \sqrt{T}b_T(\theta_0) + o_p(1)$. *Then, under Assumptions A1-A5 and B, we have*

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

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

IV Simulation Results

We perform Monte Carlo simulations to examine the finite-sample performance of the estimators of θ. We assume that the continuously compounded returns on the market portfolio are generated by Equation (3) from Section II (that we restate here for convenience):

$$r_{t_j} = \frac{1}{n_t}\mu_{t_j} + \frac{1}{n_t^{1/2}}\sigma_{t_j}\eta_{t_j}^{(1)}.$$

Note that our nonparametric estimation approach, described in Sections II and III, does not require us to specify the functional forms of either the drift, μ_{t_j}, or the diffusion, σ_{t_j}, processes in the above equation. In other words, the approach remains valid for any particular functional form specifications for these stochastic processes, provided they satisfy Assumptions A and B.

Our modeling of μ_{t_j} is motivated by the empirical specification of the risk-return relation, Equation (1), considered in this paper. In particular, we assume that the instantaneous conditional mean μ_{t_j} is linear in the conditional variance $\sigma^2_{t_j}$ and the lagged variance $\sigma^2_{(t-1)_j}$:

$$\mu_{t_j} = b_0 + b_1\sigma^2_{t_j} + b_2\sigma^2_{(t-1)_j}. \tag{14}$$

We consider two different models for $\sigma^2_{t_j}$ that have been employed extensively in the literature and shown to provide a good fit to the dynamic properties of returns. The first specification is motivated by the $GARCH(1,1)$ diffusion (see, e.g., Andersen and Bollerslev (1998)):

$$\sigma^2_{t_j} - \sigma^2_{t_{j-1}} = \frac{1}{n_t}0.035\left(0.636 - \sigma^2_{t_{j-1}}\right) + \frac{1}{n_t^{1/2}}0.236\sigma^2_{t_{j-1}}\eta_{t_j}^{(2)}. \tag{15}$$

The second model for $\sigma^2_{t_j}$ is motivated by the lognormal diffusion (see, e.g., Andersen, Benzoni, and Lund (2002)):

$$\log\left(\sigma^2_{t_j}\right) - \log\left(\sigma^2_{t_{j-1}}\right) = -\frac{1}{n_t}0.0136\left(0.8382 + \log\left(\sigma^2_{t_{j-1}}\right)\right) + \frac{1}{n_t^{1/2}}0.1148\eta^{(3)}_{t_j}. \tag{16}$$

In Equations (15) and (16), the innovations to the variance processes, $\eta^{(2)}_{t_j}$ and $\eta^{(3)}_{t_j}$, are assumed to be $i.i.d$ $N(0,1)$. We present simulation results when $\eta^{(2)}_{t_j}$ and $\eta^{(3)}_{t_j}$ are assumed to be independent of the return innovation $\eta^{(1)}_{t_j}$, i.e., there are no leverage and volatility feedback effects, as well as when these effects are present.

The lower frequency period-t market return, $r_{m,t}$, is computed as the sum of higher frequency intra period returns, r_{t_j}, i.e. $r_{m,t} = \frac{1}{n_t}\sum_{j=1}^{n_t} r_{t_j}$. Note that the conditional expectation of $r_{m,t}$, with respect to the information set available at time $t-1$, delivers the moment restrictions in Equation (1), with $b_1(x_{t-1}) = b_1$ and $Z_{t-1} = \sigma^2_{t-1}$. Similarly, the lower frequency realized market variance is computed as the sum of squares of the higher frequency intra period returns. These are then used in the GMM estimation problem (9), to estimate the parameter vector, θ. This procedure is repeated across 1000 simulated samples for different combinations of n_t and T.

To illustrate the finite sample performance of our proposed bias correction, we first focus on the simplest specification of the moment restrictions that require a non-zero bias correction, namely $b_0 = 0$, $b_1 = 0$, and $b_2 = 2$ in Equation (14). This implies a risk-return relation given by $E_{t-1}\left[r_{m,t} - b_0 - b_2\sigma^2_{t-1}\right] = 0$. We use the lagged variance σ^2_{t-1} as an instrument. Table I reports the simulation results for the GARCH(1,1) model for the variance, where $\eta^{(2)}_{t_j}$ is assumed to be independent of $\eta^{(1)}_{t_j}$. Consider first Panel A. Each row of Panel A reports results for $n_t \equiv n = 22$ and a different value of T. In the context of our empirical application, for example, this corresponds to estimating the risk-return relation at the monthly frequency, using daily returns to estimate the monthly variance. $T = 864$ corresponds to the length of the historical time series (72 years). We report results for smaller and larger values of T to show the effect of increasing the number of lower frequency time periods on the performance of the estimators.

Panel A, Row 1 corresponds to $n = 22$ high frequency data points within each of $T = 500$ time periods. The second and third columns report the mean and 95% confidence interval (in square brackets below) of the estimators of b_0 and b_2, respectively, across the 1000 simulations. The fourth and fifth columns report the same statistics as columns two and three, respectively, but for the bias-corrected estimators of these parameters. Panel A, Row 1 reveals that the bias correction proposed in Section III substantially reduces the bias in estimating the risk-return tradeoff coefficient, b_2. The mean of the standard GMM estimator $\widehat{b}_2$ across the simulations is 1.13 while the mean of the bias-corrected estimator, $\widehat{b}_2^{bc}$, is 1.33 - much closer to the population value of 2. Alternatively stated, the bias of the standard estimator is 30% higher at 0.87 compared to the bias of 0.67 for the bias-corrected estimator. Rows 2, 3, and 4 of Panel A show that increasing T to 1000, 5000, and $10,000$, respectively, monotonically increases the bias of both the standard estimator as well as that of the bias-corrected one. Note that this is not surprising because the bias is $O\left(\frac{T^{\epsilon}}{n}\right)$ and, therefore, is expected to increase as T is increased for a given n. However, note that, for each value of T, the bias-corrected estimator has a smaller bias than the standard GMM estimator.

Consider next Panel B. Each row of Panel B reports results for $T = 1000$ and a different value of n. This allows us to study the effect of increasing the number of high frequency data points within a lower frequency time period on the performance of the estimators. The results in Panel B show that, similar to Panel A, the bias-corrected estimator of b_2 has a smaller bias for all the values of n compared to the standard estimator and that, not surprisingly, the difference between the two estimators diminish with increasing n for a given T.

Note that Table I presents results ruling out leverage and volatility feedback effects. We repeated our simulations allowing for these effects, in particular, by setting the correlation between $\eta_{t_j}^{(2)}$ and $\eta_{t_j}^{(1)}$ to be -0.5. The results, presented in Table II, remain almost identical to those obtained in Table 1.

Table III reports simulation results for the lognormal model for the variance, when leverage and volatility feedback effects are ruled out. The results are largely similar to those obtained in Tables 1-2 for the GARCH(1,1) model. Panels A and B shows that the bias of the bias-corrected estimator is smaller than that of the standard GMM estimator for all combinations of n and T, particularly when n is small relative to T. Once again, the results remain virtually unchanged in the presence of leverage and volatility feedback effects. These are omitted for brevity.

Overall, the simulation results point toward the superior performance of the bias-corrected estimator relative to the standard GMM estimator. And, importantly, the improvement is particularly pronounced in the empirically relevant scenario, where the size of the intra-period sample is small compared to the number of low frequency time periods.

V Empirical Results

In our empirical analysis, we focus on the risk-return relation at the monthly frequency. 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 of all stocks on the NYSE, AMEX, and NASDAQ. The sample extends over the entire post war period, 1947:01 - 2018:12. The monthly market return is obtained as the sum of daily continuously compounded market returns within the month and the realized monthly market variance as the sum of squares of the daily continuously compounded market returns. Figure 1 displays the time series of the monthly returns and realized volatility.

The analysis in Section II shows that the estimation of the risk-return trade-off parameters can be posed as a GMM estimation problem, with the moment specification in Equation (9), that we restate here for convenience:

$$E\left[\left(r_{m,t} - r_{f,t} - b_0 - b_1\left(x_{t-1}\right)\widehat{\sigma}_t^2 - b_2^{\intercal} Z_{t-1}\right) Y_{t-1}\right] = 0,$$

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

Our first specification is obtained by setting $b_1\left(x_{t-1}\right) = b_1$ (a scalar constant), $b_2^{\intercal} = 0^{\intercal}$, and using the lagged variance, $\widehat{\sigma}_{t-1}^2$, as an instrument. The rationale for using the lagged variance as an instrument is that the variance process is highly persistent. The first order autocorrelation coefficient of the realized variance process is 0.52 in monthly data for the full sample. Hence, the lagged variance is useful in predicting the contemporaneous variance which enters the moment specification. This makes it a good choice of instrument, improving the efficiency of the estimation procedure. This specification produces an exactly identified system of two moment restrictions in two unknown parameters, $\theta = (b_0, b_1)$, to be estimated. Note that for this specification of the moment restrictions and choice of instruments, the bias-correction is identically zero (see Theorem 1). Table IV, Panel A reports the estimation results. The first row presents results over the entire available sample period, while Rows 2 and 3 do the same for two non-overlapping subsamples of equal length. Row 1 shows that, over the full sample, the estimated coefficient of the conditional variance b_1 is negative and statistically significant at the 5% level of significance. Rows 2 and 3 show that while the coefficient is statistically indistinguishable from zero over the first subsample covering the period 1947:01–1982:12, it is statistically significantly negative in the second subperiod covering 1983:01–2018:12.

Our second specification is identical to the first, except that the lagged (instead of contemporaneous) variance is used in the moment restriction. This specification is obtained by setting $b_1\left(x_{t-1}\right) = 0$ and $Z_{t-1} = \widehat{\sigma}_{t-1}^2$. Unlike the first specification, for this specification of the moment restrictions and choice of instruments, Theorem 1 suggests a bias-correction to the standard GMM estimator to improve its performance. Table IV, Panel B reports the estimation results. Row 1 shows that, over the full sample, both the standard GMM estimator as well bias-corrected estimator of the coefficient b_2 are negative and statistically significant. The bias-correction increases the magnitude of the estimate, thereby further increasing its statistical significance relative to that of the standard GMM estimator. As in Panel A, Row 2 shows that a statistically insignificant risk-return relation is obtained in the first subsample. Row 3, however, shows that, for the second subsample, the standard GMM estimator of the risk-return relation is significantly negative. We obtain even stronger results for the bias-corrected estimator.

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

Our final specification is obtained by setting $b_1(x_{t-1}) = b_1$ (a scalar constant) and $Z_{t-1} = \left(r_{m,t-1}, \widehat{\sigma}^2_{t-1}\right)$. This specification is motivated by the findings in Whitelaw (1994), Brandt and Kang (2004), and Lettau and Ludvigson (2010) that the lagged conditional mean and conditional variance are a statistically important feature of the empirical risk-return relation and that it is important to distinguish between the unconditional correlation and the conditional correlation (conditional on the lagged mean and variance) between the first two moments of the market return to uncover the relation between them. Thus, we have four parameters to estimate. We use two lags of the variance and one lag of the market return as instruments giving an exactly identified system of four moment restrictions in four parameters. The results in Table IV, Panel C show that the estimated coefficient of the conditional variance b_1 is positive for the full sample (Row 1) as well as in each subsample (Rows 2-3). This is consistent with the findings in Lettau and Ludvigson (2010) who find a negative unconditional correlation between the expected market return and its conditional variance but a positive conditional correlation (conditional on the lagged mean and variance) between these moments. However, while the latter paper finds the positive conditional correlation to be strongly statistically significant, our estimate of the conditional correlation is not statistically significant. The bias-correction increases the magnitude of the point estimate but not sufficiently relative to its standard error to make it statistically significant.

Panels A, B, and C of Table V report results similar to those in Table IV, but for an over-identified system that includes the one-month Treasury Bill rate and the default spread as additional instruments. The results are very similar to those obtained in Table IV. Specifically, the unconditional correlation between the expected market return and its conditional variance is typically estimated to be negative (Panels A-B), while their conditional correlation to be positive (Panel C). Moreover, unlike in Table IV where the conditional correlation is found to not be statistically significant, statistical significance is obtained for the over-identified system over the full sample (Table V, Panel C, Row 1). And the bias correction aids in increasing the statistical significance of the estimator.

Finally, in Table VI, we present the estimates of the unconditional risk-return relation separately over the expansionary and recessionary phases of the business cycle. We split the sample into two subsamples. The first subsample consists of those months that are in NBER-designated expansion periods. The second subsample, on the other hand, includes those months that are in NBER-designated recessionary episodes. Panel A presents the results for the same specification as in Panels A of Tables IV-V. The risk-return tradeoff is estimated to be negative during both expansions and recessions. Similar negative estimates are obtained in Panel B, that presents the results for the same specification as in Panels B of Tables IV-V. In this case, the bias-correction increases the magnitudes of the estimated coefficients making them statistically significant at conventional levels of significance.

VI Conclusion and Extensions

In this paper, 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 in order to estimate the risk-return relation. 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 of the risk-return relation.

We derive the limiting distribution of the estimated coefficients under this double asymptotic framework. We find that under fairly strong conditions on n and T, the estimates are $\sqrt{T}$ -consistent and have the standard distribution as when there is no measurement-error. However, if the above condition is not satisfied, there is an asymptotic bias that would invalidate this approximation. In that case, we find that under weaker conditions on n and T, a bias-corrected estimator has the standard limiting distribution. This improvement is particularly relevant in the empirical case we examine where n is quite modest relative to T.

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

References

[1] Ait-Sahalia, Y., and Jacod, J., (2014), "High-Frequency Financial Econometrics," Princeton University Press.

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

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

[4] Andersen, T.G., Bollerslev, T., Diebold, F.X., (2002), "Parametric and Nonparametric Volatility Measurement," In: Ait-Sahalia, Y., and Hansen, L.P. (Eds.), *Handbook of Financial Econometrics*. North-Holland, Holland.

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

[6] Bai, J., and Ng, S., (2006), "Evaluating Latent and Observed Factors in Macroeconomics and Finance," *Journal of Econometrics*, 131, 507-537.

[7] Bandi, F. and B. Perron (2008) Long-run Risk-Return Trade-offs, Journal of Econometrics, 143, 2008, 349-374.

[8] 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.

[9] Barndorff-Nielsen, O.E., and Shephard, N., (2004), "Econometrics of testing for jumps in financial economics using bipower variation," *Journal of Financial Econometrics 2, 1-37.*

[10] Barndorff-Nielsen, O.E., and Shephard, N., (2006), "Power and Bi-power variation with Stochastic volatility and Jumps," *Journal of Financial Econometrics 4, 1-30.*

[11] Bollerslev, T., Engle, E., and Wooldridge, J., (1988), "A Capital Asset Pricing Model with Time-Varying Covariances," *Journal of Political Economy*, 96, 116-131.

[12] Bollerslev, T., Patton, A. J., and Quaedvlieg, R., (2016), "Exploiting the Errors: A Simple Approach for Improved Volatility Forecasting," *Journal of Econometrics*, 192(1), 1-18.

[13] Bollerslev, T., and Zhou, H., (2002), "Estimating Stochastic Volatility Diffusion Using Conditional Moments of Integrated Volatility," *Journal of Econometrics*, 109, 33-65.

[14] Bosq, D., (1998), "Nonparametric Statistics for Stochastic Processes: Estimation and Prediction," Lecture Notes in Statistics 110, Springer-Verlag.

[15] Brandt, M.W., and Kang, Q., (2004), "On the Relation Between the Conditional Mean and Volatility of Stock Returns: A Latent VAR Approach," *Journal of Financial Economics*, 72, 217-257.

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

[17] 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.

[18] Corradi, V., and W. Distaso (2006). Semi-Parametric Comparison of Stochastic Volatility Models using Realized Measures. *Review of Economic Studies* 73, 635–667.

[19] de la Pena, V. H., (1999), "A General Class of Exponential Inequalities for Martingales and Ratios," *The Annals of Probability*, 27, 537-564.

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

[21] 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.

[22] 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.

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

[24] Graham, J.R., and Harvey, C.R., (2008), "Expectations of Equity Risk Premia, Volatility, and Asymmetry from a Corporate Finance Perspective," *Reviewof Financial Studies*, 21(4), 1455-1508.

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

[26] Hall, P., and Heyde, C. C., (1980), "Martingale Limit Theory and its Applications," Academic Press, New York.

[27] Harrison, P., and Zhang, H.H., (1999), "An Investigation of the Risk and Return Relation at Long Horizons," *Review of Economics and Statistics*, 81(3), 399-408.

[28] Harvey, C.R. (1989), "Time-Varying Conditional Covariances in Tests of Asset Pricing Models," *Journal of Financial Economics*, 24, 289-371.

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

[30] 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.

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

[32] Kanaya, S., and Kristensen, D., (2016), "Estimation of Stochastic Volatility Models by Nonparametric Filtering," *Econometric Theory*, 32(4), 861-916.

[33] Lettau, M., and Ludvigson, S.C., (2010), "Measuring and Modeling Variation in the Risk-Return Trade-off," *Handbook of Financial Econometrics*, ed. by Yacine Ait-Sahalia and Lars P. Hansen, vol. 1, 617-690. Elsevier Science B.V., North Holland, Amsterdam.

[34] Li, J., and Patton, A. J., (2015), "Asymptotic Inference about Predictive Ability using High Frequency Data," *Working paper.*

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

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

[37] Mykland, P. A. and Zhang, L. (2009), Inference for Continuous Semimartingales Observed at High Frequency. Econometrica, 77: 1403–1445. doi: 10.3982/ECTA7417

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

[39] Pakes, A., and D. Pollard (1989), “Simulation and the asymptotics of optimization estimators,” *Econometrica*, 57, 1027-1057.

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

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

[42] Scruggs, J.T., (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.

[43] Todorov, V., (2009), "Estimation of Continuous-Time Stochastic Volatility Models with Jumps Using High-Frequency Data," *Journal of Econometrics*, 148, 131-148.

[44] 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

A.1 Proof of Lemma 1

The proof of Lemma 1 relies on the exponential inequality of de La Peña (1999, Theorem 1.2A) for martingale difference sequences, which we repeat here for convenience.

THEOREM. *Let* $(X_j, \mathcal{F}_j)$, $j = 1, 2, \ldots, n$, *be a martingale difference sequence with* $E(X_j|\mathcal{F}_{j-1}) = 0$ *and* $\sigma_j^2 = E(X_j^2|\mathcal{F}_{j-1})$ *and let* $V_n^2 = \sum_{j=1}^n \sigma_j^2$. *Furthermore, suppose that for some* c

$$\Pr\left[|X_j| \leq c|\mathcal{F}_{j-1}\right] = 1.$$

Then

$$\Pr\left[\sum_{j=1}^n X_j \geq x, V_n^2 \leq y\right] \leq \exp\left(-\frac{x^2}{2(y+cx)}\right). \tag{17}$$

Lemma 1. *Suppose that Assumptions A1-A4 hold. Then, we have*

$$\text{(a)} \max_{1\leq t\leq T}\left|(\hat{\sigma}_t^2 - \sigma_t^2) - \frac{1}{n_t}\sum_{j=1}^{n_t}\sigma_{t_j}^2\left(\eta_{t_j}^2 - 1\right)\right| = O_p(T^{-\lambda\gamma}),$$

$$\text{(b) } T^\alpha \max_{1\leq t\leq T}\left|\hat{\sigma}_t^2 - \sigma_t^2\right| = o_p(1) \text{ for } \alpha < \gamma/2 - \varepsilon.$$

Proof of Lemma 1. From (3) we obtain

$$\sum_{j=1}^{n_t} r_{t_j}^2 = \frac{1}{n_t}\sum_{j=1}^{n_t}\mu_{t_j}^2 + \frac{1}{n_t}\sum_{j=1}^{n_t}\sigma_{t_j}^2\eta_{t_j}^2 + \frac{2}{n_t^{3/2}}\sum_{j=1}^{n_t}\mu_{t_j}\sigma_{t_j}\eta_{t_j}.$$

Therefore,

$$\begin{aligned}\widehat{\sigma}_t^2 - \sigma_t^2 &= \underbrace{\left(\frac{1}{n_t}\sum_{j=1}^{n_t}\sigma_{t_j}^2 - \sigma_t^2\right)}_{O_p\left(n^{-\lambda}\right)} + \underbrace{\frac{1}{n_t}\sum_{j=1}^{n_t}\sigma_{t_j}^2(\eta_{t_j}^2 - 1)}_{O_p\left(\sqrt{\frac{T^\epsilon}{n}}\right)} + \underbrace{\frac{1}{n_t^2}\sum_{j=1}^{n_t}\mu_{t_j}^2}_{O_p(n^{-1})} + \underbrace{\frac{2}{n_t^{3/2}}\sum_{j=1}^{n_t}\mu_{t_j}\sigma_{t_j}\eta_{t_j}}_{O_p(n^{-1})} \\ &= J_{1t} + J_{2t} + J_{3t} + J_{4t}.\end{aligned} \tag{18}$$

We have $\max_{1\leq t\leq T}|J_{1t}| = O_p(n^{-\lambda})$ by Assumption A3. We have $\max_{1\leq t\leq T}|J_{3t}| = O_p(n^{-1})$ under our conditions. Furthermore, $\max_{1\leq t\leq T}|J_{4t}| = O_p(n^{-1})$ by a similar argument to the sequel under our conditions. Thus, J_{3t} and J_{4t} are smaller than J_{1t} since $\lambda < 1$. This establishes Lemma 1(a).

Consider next the term J_{2t}. This is a sum of martingale differences, which satisfies for each t

$$\frac{1}{n_t}\sum_{j=1}^{n_t} E\left[\sigma_{t_j}^2(\eta_{t_j}^2-1)|\mathcal{F}_{t_{j-1}}\right]=0 \tag{19}$$

$$\begin{aligned}\frac{1}{n_t^2}\sum_{j=1}^{n_t} E\left[\sigma_{t_j}^4(\eta_{t_j}^2-1)^2|\mathcal{F}_{t_{j-1}}\right] &= \frac{1}{n_t}\left(\frac{1}{n_t}\sum_{j=1}^{n_t}\sigma_{t_j}^4 E\left[(\eta_{t_j}^2-1)^2|\mathcal{F}_{t_{j-1}}\right]\right) \\ &\leq \frac{C}{n_t}\frac{1}{n_t}\sum_{j=1}^{n_t}\sigma_{t_j}^4 \\ &\leqslant \frac{C}{n}MT^\epsilon \\ &= O\left(\frac{T^\epsilon}{n}\right)=o(1),\ \text{ since } \gamma>\epsilon.\end{aligned} \tag{20}$$

Note that the second-to-last line derives from Assumption A1. Therefore, $J_{2t}=O_p\left(\sqrt{\frac{T^\epsilon}{n}}\right)$ uniformly over t.

Let $X_j(t)=\sigma_{t_j}^2(\eta_{t_j}^2-1)$ and write $X_j(t)=X_j^+(t)+X_j^-(t)$, where:

$$\begin{aligned} X_j^+(t) &= X_j(t)1\left(|X_j(t)|\leq\sqrt{\frac{n}{\log n}}\right)-E\left[X_j(t)1\left(|X_j(t)|\leq\sqrt{\frac{n}{\log n}}\right)|\mathcal{F}_{j-1}\right] \\ X_j^-(t) &= X_j(t)1\left(|X_j(t)|>\sqrt{\frac{n}{\log n}}\right)-E\left[X_j(t)1\left(|X_j(t)|>\sqrt{\frac{n}{\log n}}\right)|\mathcal{F}_{j-1}\right].\end{aligned}$$

Then we have

$$\max_{1\leq t\leq T}\left|\frac{1}{n_t}\sum_{j=1}^{n_t}\sigma_{t_j}^2(\eta_{t_j}^2-1)\right|\leq\max_{1\leq t\leq T}\left|\frac{1}{n_t}\sum_{j=1}^{n_t}X_j^+(t)\right|+\max_{1\leq t\leq T}\left|\frac{1}{n_t}\sum_{j=1}^{n_t}X_j^-(t)\right|=I_1+I_2.$$

For I_1 we can apply (17). Write $\sigma_j^{+2}(t)=E[X_j^{+2}(t)|\mathcal{F}_{j-1}]$ and $V_{n_t}^{2+}=\sum_{j=1}^{n_t}\sigma_j^{+2}(t)$. Then with probability one $\max_{1\leq t\leq T}V_{n_t}^{2+}\leq CnT^\epsilon$ for some constant C. By the Bonferroni inequality and (17)

$$\begin{aligned}
\Pr\left[\max_{1\le t\le T}\left|\sum_{j=1}^{n_t} X_j^{+}(t)\right| \ge cT^{\epsilon/2}\sqrt{n\log n}\right] &\le 2T\Pr\left[\left|\sum_{j=1}^{n_t} X_j^{+}(t)\right| \ge cT^{\epsilon/2}\sqrt{n\log n}, V_{n_t}^{2+}\le CnT^{\varepsilon}\right] \\
&\le 2T\exp\left(-\frac{c^2T^{\epsilon}n\log n}{2(CnT^{\varepsilon}+\sqrt{\frac{n}{\log n}}cT^{\varepsilon/2}\sqrt{n\log n})}\right) \\
&= 2T\exp\left(-\frac{c^2\log n}{2(C+cT^{-\epsilon/2})}\right) \\
&= 2Tn^{-\rho}
\end{aligned}$$

for some $\rho > 0$. Choosing ρ large enough ensures this term is small, which means that $I_1 = O_P\left(\sqrt{\frac{T^{\epsilon}\log n}{n}}\right)$. Regarding I_2, we have

$$\begin{aligned}
\Pr\left[\max_{1\le t\le T}\left|\sum_{j=1}^{n_t} X_j^{-}(t)\right| \ge cT^{\epsilon/2}\sqrt{n\log n}\right] &\le \Pr\left[\max_{1\le t\le T}\max_{1\le j\le n_t}\left|\sigma_{t_j}^2(\eta_{t_j}^2-1)\right| > c\sqrt{\frac{n}{\log n}}\right] \\
&\le Tnc'\frac{(\log n)^{k/2}E\left[(\eta_{t_j}^2-1)^k\right]}{n^{k/2}} \\
&= o(1)
\end{aligned}$$

by the Markov inequality applied to the stationary process $\eta_{t_j}^2 - 1$. The last equality holds provided k is taken large enough.

Thus, it follows that

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

So, provided $\alpha < \frac{\gamma}{2} - \epsilon$, the result of Lemma 1(b) follows.

■

A.2 Proof of Theorem 1

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

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

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

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

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

By the Mean Value Theorem, for a set of mean values $\overline{V}_t$

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

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

$$\begin{aligned}\sup_{\theta\in\Theta} \|\widehat{G}_T(\theta) - G_T(\theta)\|_W &= \sup_{\theta\in\Theta} \left\|\frac{1}{T}\sum_{t=1}^{T} G_V\left(X_t, \overline{V}_t; \theta\right)^\intercal (\widehat{V}_t - V_t)\right\|_W \\ &\le \dim\left(V_t\right) \delta_T \frac{1}{T}\sum_{t=1}^{T} U_t = o_p(1),\end{aligned}$$

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

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

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

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

For the asymptotic expansion, our proof parallels the work of Pakes and Pollard (1989). We expand the estimated moment condition out to third order. Therefore, for the i-th moment restriction, we have

$$\begin{aligned}\widehat{G}_{i,T}(\theta_0) - G_{i,T}(\theta_0) &= \frac{1}{T}\sum_{t=1}^{T} G_{i,V}\left(X_t, V_t; \theta_0\right)^{\intercal}(\widehat{V}_t - V_t) \\ &+ \frac{1}{2T}\sum_{t=1}^{T}(\widehat{V}_t - V_t)^{\intercal} G_{i,VV}\left(X_t, V_t; \theta_0\right)(\widehat{V}_t - V_t) \\ &+ \frac{1}{6T}\sum_{t=1}^{T} G_{i,VVV}\left(X_t, \overline{V}_t; \theta_0\right)(\widehat{V}_t - V_t)^{\otimes 3}\end{aligned}$$

where $\overline{V}_t$ is intermediate between $\widehat{V}_t$ and V_t.

Consider the first term

$$\frac{1}{T}\sum_{t=1}^{T} G_{i,V}\left(X_t, V_t; \theta_0\right)^{\intercal}(\widehat{V}_t - V_t) = \frac{1}{T}\sum_{t=1}^{T}\sum_{k=0}^{p} G_{i,\sigma^2_{t-k}}(X_t, V_t; \theta_0)\frac{1}{n_{t-k}}\sum_{j=1}^{n_t}\sigma^2_{t-k_j}(\eta^2_{t-k_j} - 1).$$

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

$$\begin{aligned} & E\left[\frac{1}{T}\sum_{t=1}^{T} G_{i,\sigma^2_{t-k}}(X_t, V_t; \theta_0)\frac{1}{n_{t-k}}\sum_{j=1}^{n_t}\sigma^2_{t-k_j}(\eta^2_{t-k_j} - 1)\right] \\ =\ & \frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_{t-k}}\sum_{j=1}^{n_t} E\left[G_{i,\sigma^2_{t-k}}(X_t, V_t; \theta_0) E\left(\sigma^2_{t-k_j}(\eta^2_{t-k_j} - 1) | X_t, V_t\right)\right] = 0.\end{aligned}$$

Also,

$$\begin{aligned} & \operatorname{var}\left[\frac{1}{T}\sum_{t=1}^{T} G_{i,\sigma^2_{t-k}}(X_t, V_t; \theta_0)\frac{1}{n_{t-k}}\sum_{j=1}^{n_t}\sigma^2_{t-k_j}(\eta^2_{t-k_j} - 1)\right] \\ =\ & E\left[\frac{1}{T^2}\sum_{t=1}^{T}\frac{1}{n^2_{t-k}}\sum_{j=1}^{n_t} G^2_{i,\sigma^2_{t-k}}(X_t, V_t; \theta_0)\sigma^4_{t-k_j}(\eta^2_{t-k_j} - 1)^2\right] \\ \leqslant\ & \frac{C}{n_{t-k}T} E\left[\left(\frac{1}{T}\sum_{t=1}^{T} G^2_{i,\sigma^2_{t-k}}(X_t, V_t; \theta_0)\right)\left(\frac{1}{n_t}\sum_{j=1}^{n_t}\sigma^4_{t-k_j}\right)\right] \\ \leqslant\ & \frac{C}{n_{t-k}T} E\left(U_t\right) M T^{\epsilon} = O\left(\frac{T^{\epsilon}}{n_{t-k}T}\right) = o\left(\frac{1}{T}\right), \text{ provided } \gamma > \epsilon.\end{aligned}$$

Next, consider the second term,

$$
\begin{aligned}
& \frac{1}{2T}\sum_{t=1}^{T}(\widehat{V}_t - V_t)^\intercal G_{VV}(X_t, V_t;\theta_0)(\widehat{V}_t - V_t) \\
= \;& \frac{1}{2T}\sum_{t=1}^{T}\left[\frac{1}{n_t}\sum_{j=1}^{n_t}\sigma^2_{t_j}(\eta^2_{t_j}-1),\ldots,\frac{1}{n_t}\sum_{j=1}^{n_t}\sigma^2_{t-p_j}(\eta^2_{t-p_j}-1)\right] G_{VV}(X_t, V_t;\theta_0) \\
& \left[\frac{1}{n_t}\sum_{j=1}^{n_t}\sigma^2_{t_j}(\eta^2_{t_j}-1),\ldots,\frac{1}{n_t}\sum_{j=1}^{n_t}\sigma^2_{t-p_j}(\eta^2_{t-p_j}-1)\right]^\intercal \\
= \;& \sum_{k=0}^{p}\sum_{l=0}^{p}\frac{1}{2T}\sum_{t=1}^{T}\frac{1}{n_{t-k}n_{t-l}}\sum_{j=1}^{n_t} G_{\sigma^2_{t-k}\sigma^2_{t-l}}(X_t, V_t;\theta_0)\sigma^2_{t-k_j}(\eta^2_{t-k_j}-1)\sigma^2_{t-l_j}(\eta^2_{t-l_j}-1).
\end{aligned}
$$

We have

$$
\begin{aligned}
& E\left[\frac{1}{2T}\sum_{t=1}^{T}(\widehat{V}_t - V_t)^\intercal G_{VV}(X_t, V_t;\theta_0)(\widehat{V}_t - V_t)\right] \\
= \;& \frac{1}{2T}\sum_{t=1}^{T}\operatorname{tr}\left(E\left\{G_{VV}(X_t, V_t;\theta_0)E\left[(\widehat{V}_t - V_t)(\widehat{V}_t - V_t)^\intercal | X_t, V_t\right]\right\}\right) \\
= \;& \sum_{k=0}^{p}\frac{1}{2T}\sum_{t=1}^{T}E\left[G_{\sigma^2_{t-k}\sigma^2_{t-k}}(X_t, V_t;\theta_0)(\widehat{\sigma}^2_{t-k}-\sigma^2_{t-k})^2\right].
\end{aligned}
$$

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

Finally, we consider the third order terms,

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

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

Hence,

$$\begin{aligned}\widehat{G}_T(\theta_0) &\simeq G_T(\theta_0) + \sum_{k=0}^{p} \frac{1}{2T} \sum_{t=1}^{T} E\left[G_{\sigma^2_{t-k}\sigma^2_{t-k}}(X_t, V_t; \theta_0)(\widehat{\sigma}^2_{t-k} - \sigma^2_{t-k})^2\right] \\ &= G_T(\theta_0) + b_T(\theta_0).\end{aligned}$$

Therefore, we have

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

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

$$b_T(\theta_0) = E\left[\sum_{k=0}^{p} \frac{1}{2T} \sum_{t=1}^{T} G_{\sigma^2_{t-k}\sigma^2_{t-k}}(X_t, V_t; \theta_0)(\widehat{\sigma}^2_{t-k} - \sigma^2_{t-k})^2\right]$$

Now,

$$\begin{aligned}& E\left[\frac{1}{2T} \sum_{t=1}^{T} G_{\sigma^2_{t-k}\sigma^2_{t-k}}(X_t, V_t; \theta_0)(\widehat{\sigma}^2_{t-k} - \sigma^2_{t-k})^2\right] \\ \leq\ & \frac{C}{2} E\left[\left(\frac{1}{T} \sum_{t=1}^{T} G_{\sigma^2_{t-k}\sigma^2_{t-k}}(X_t, V_t; \theta_0)\right) \frac{1}{n_t} \left(\frac{1}{n_t} \sum_{t=1}^{T} \sigma^4_{t_j}\right)\right] \\ \leq\ & \frac{C}{2} E\left(U_t\right) \frac{1}{n_t} M T^{\epsilon} \\ =\ & O\left(\frac{T^{\epsilon}}{N}\right) = o\left(T^{-1/2}\right),\end{aligned}$$

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

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

Hence,

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

Corollary 2: When the above condition is not satisfied, we may not have $T^{1/2}$ consistency because of the asymptotic bias. However, we show that a bias corrected estimator $\widehat{\theta} + (\Gamma^\top W \Gamma)^{-1}\Gamma^\top W b_T(\theta_0)$ 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 (13). Define the estimated bias function

$$\begin{aligned} \widehat{b}_T(\theta) &= \sum_{k=0}^{p} \frac{1}{2T} \sum_{t=1}^{T} G_{\sigma^2_{t-k}\sigma^2_{t-k}}(X_t, V_t; \theta_0) \frac{\kappa - 1}{\kappa} \frac{RQ_{t-k}}{n_{t-k}} \\ \text{where } RQ_t &= n_t \sum_{j=1}^{n_t} r^4_{t_j} \end{aligned}$$

and $\kappa = 3$ if η_{t_j} is normally distributed. Then, for the bias corrected estimator

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

we have that

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

provided

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

Using similar arguments as above, this requires $\frac{n^x}{T} \to \infty$ where $x > \frac{1}{\gamma}$, where $\gamma > \epsilon + \frac{1}{2} - \alpha$.
■

Table I: Simulation Results for GARCH(1,1) Diffusion

	Bias-Uncorrected Estimators		Bias-Corrected Estimators	
	b_0	b_2	b_0^{bc}	b_2^{bc}
Panel A: Effect of Changing T Keeping N Fixed				
$N = 22; T = 500$	.41	1.13	.28	1.33
	$[.18, .99]$	$[.64, 1.42]$	$[.04, .89]$	$[.69, 1.73]$
$N = 22; T = 1,000$	.44	1.08	.33	1.24
	$[.23, 1.03]$	$[.51, 1.34]$	$[.09, .97]$	$[.54, 1.60]$
$N = 22; T = 5,000$	.56	.90	.49	.99
	$[.31, 1.15]$	$[.30, 1.20]$	$[.20, 1.13]$	$[.31, 1.39]$
$N = 22; T = 10,000$	.62	.81	.56	.89
	$[.35, 1.14]$	$[.23, 1.15]$	$[.25, 1.12]$	$[.24, 1.31]$
Panel B: Effect of Changing N Keeping T Fixed				
$N = 22; T = 1,000$	.44	1.08	.33	1.24
	$[.23, 1.03]$	$[.51, 1.34]$	$[.09, .97]$	$[.54, 1.60]$
$N = 66; T = 1,000$	.22	1.55	.16	1.64
	$[.06, .70]$	$[1.05, 1.80]$	$[-.01, .65]$	$[1.09, 1.95]$
$N = 132; T = 1,000$	.12	1.75	.08	1.81
	$[.01, .38]$	$[1.43, 1.95]$	$[-.03, .36]$	$[1.45, 2.04]$
$N = 264; T = 1,000$	.06	1.87	.04	1.90
	$[-.04, .24]$	$[1.63, 2.07]$	$[-.06, .23]$	$[1.65, 2.11]$

The table reports simulation results for the system $E_{t-1}\left[r_{m,t} - b_0 - b_2\sigma_{t-1}^2\right] = 0$. High frequency returns are assumed to be generated by Equation (3), where the diffusion term $\sigma_{t_j}^2$ is assumed to follow a GARCH(1,1) process, while the drift term $\mu_{t_j} = b_0 + b_2\sigma_{(t-1)_j}^2$, with $b_0 = 0$ and b_2=2. The table reports the mean and the 95% confidence interval (in square brackets) of the standard GMM parameter estimates and the bias-corrected estimates across 2000 simulations.

Table II: Simulation Results for GARCH(1,1) Diffusion With Leverage Effects

	Bias-Uncorrected Estimators		Bias-Corrected Estimators	
	b_0	b_2	b_0^{bc}	b_2^{bc}
Panel A: Effect of Changing T Keeping N Fixed				
$N = 22; T = 500$	.39	1.14	.26	1.33
	[.17, .96]	[.64, 1.41]	[.03, .87]	[.68, 1.73]
$N = 22; T = 1,000$	.42	1.08	.31	1.24
	[.21, .90]	[.61, 1.35]	[.08, .83]	[.66, 1.62]
$N = 22; T = 5,000$	.55	.88	.48	.98
	[.30, 1.10]	[.28, 1.21]	[.19, 1.07]	[.30, 1.39]
$N = 22; T = 10,000$	.61	.81	.55	.89
	[.34, 1.12]	[.25, 1.16]	[.23, 1.11]	[.26, 1.32]
Panel B: Effect of Changing N Keeping T Fixed				
$N = 22; T = 1,000$	.42	1.08	.31	1.24
	[.21, .90]	[.61, 1.35]	[.08, .83]	[.66, 1.62]
$N = 66; T = 1,000$	.19	1.57	.13	1.66
	[.05, .61]	[1.06, 1.80]	[−.03, .55]	[1.09, 1.95]
$N = 132; T = 1,000$	.11	1.76	.07	1.81
	[−.01, .39]	[1.36, 1.97]	[−.06, .36]	[1.38, 2.05]
$N = 264; T = 1,000$	.05	1.87	.03	1.90
	[−.06, .25]	[1.59, 2.08]	[−.08, .23]	[1.61, 2.13]

The table reports simulation results for the system $E_{t-1}\left[r_{m,t} - b_0 - b_2\sigma^2_{t-1}\right] = 0$. High frequency returns are assumed to be generated by Equation (3), where the diffusion term $\sigma^2_{t_j}$ is assumed to follow a GARCH(1,1) process, while the drift term $\mu_{t_j} = b_0 + b_2\sigma^2_{(t-1)_j}$, with $b_0 = 0$ and b_2=2. The table reports the mean and the 95% confidence interval (in square brackets) of the standard GMM parameter estimates and the bias-corrected estimates across 2000 simulations.

Table III: Simulation Results for Lognormal Diffusion

	Bias-Uncorrected Estimators		Bias-Corrected Estimators	
	b_0	b_2	b_0^{bc}	b_2^{bc}
Panel A: Effect of Changing T Keeping N Fixed				
$N = 22; T = 500$	.36	1.18	.19	1.48
	$[.13, .76]$	$[.87, 1.46]$	$[.01, .51]$	$[1.12, 1.84]$
$N = 22; T = 1,000$	.34	1.21	.19	1.46
	$[.17, .61]$	$[.99, 1.40]$	$[.04, .42]$	$[1.17, 1.72]$
$N = 22; T = 5,000$	.34	1.20	.21	1.40
	$[.24, .48]$	$[1.01, 1.30]$	$[.11, .39]$	$[1.13, 1.57]$
$N = 22; T = 10,000$	.34	1.19	.22	1.39
	$[.26, .47]$	$[1.00, 1.28]$	$[.13, .39]$	$[1.12, 1.52]$
Panel B: Effect of Changing N Keeping T Fixed				
$N = 22; T = 1,000$	.34	1.21	.19	1.46
	$[.17, .61]$	$[.99, 1.40]$	$[.04, .42]$	$[1.17, 1.72]$
$N = 66; T = 1,000$	.15	1.65	.06	1.80
	$[.05, .30]$	$[1.46, 1.83]$	$[-.04, .19]$	$[1.59, 2.01]$
$N = 132; T = 1,000$	.08	1.81	.03	1.90
	$[-.01, .20]$	$[1.63, 2.00]$	$[-.06, .14]$	$[1.71, 2.11]$
$N = 264; T = 1,000$	.04	1.90	.02	1.95
	$[-.04, .14]$	$[1.73, 2.08]$	$[-.07, .11]$	$[1.77, 2.14]$

The table reports simulation results for the system $E_{t-1}\left[r_{m,t} - b_0 - b_2\sigma^2_{t-1}\right] = 0$. High frequency returns are assumed to be generated by Equation (3), where the diffusion term $\sigma^2_{t_j}$ is assumed to follow a lognormal process, while the drift term $\mu_{t_j} = b_0 + b_2\sigma^2_{(t-1)_j}$, with $b_0 = 0$ and $b_2 = 2$. The table reports the mean and the 95% confidence interval (in square brackets) of the standard GMM parameter estimates and the bias-corrected estimates across 2000 simulations.

Table IV: Estimation Results at the Monthly Frequency

Sample	Bias-Uncorrected Estimators				Bias-Corrected Estimators			
	b_0	b_1	b_2	b_3	b_0^{bc}	b_1^{bc}	b_2^{bc}	b_3^{bc}
Panel A: $E_{t-1}\left[r_{m,t} - b_0 - b_1\upsilon_t\right] = 0$								
1947:01-2018:12	.01 (6.13)	−2.41 (−2.21)						
1947:01-1982:12	.01 (1.26)	1.43 (.28)						
1983:01-2018:12	.02 (5.16)	−2.79 (−2.42)						
Panel B: $E_{t-1}\left[r_{m,t} - b_0 - b_1\upsilon_{t-1}\right] = 0$								
1947:01-2018:12	.01 (7.18)	−1.27 (−2.35)			−.01 (7.46)	−1.47 (−2.79)		
1947:01-1982:12	.01 (2.59)	.72 (.28)			.01 (2.52)	.88 (.34)		
1983:01-2018:12	.01 (5.56)	−1.42 (−2.61)			.01 (5.83)	−1.66 (−3.05)		
Panel C: $E_{t-1}\left[r_{m,t} - b_0 - b_1\upsilon_t - b_2\upsilon_{t-1} - b_3 r_{m,t-1}\right] = 0$								
1947:01-2018:12	−.004 (−.04)	10.5 (.96)	−6.19 (−.90)	.14 (1.14)	−.004 (−.30)	16.7 (.97)	−10.3 (−.93)	.16 (.87)
1947:01-1982:12	−.02 (−.79)	33.4 (1.13)	−13.6 (−1.05)	.30 (1.39)	−.03 (−.93)	52.0 (1.36)	−25.1 (−1.51)	.40 (1.42)
1983:01-2018:12	−.003 (−.17)	9.81 (.83)	−5.84 (−.84)	.15 (.75)	−.008 (−.35)	15.9 (.86)	−9.85 (−.88)	.17 (.55)

The table reports estimation results at the monthly frequency for exactly identified systems.

Table V: Estimation Results at the Monthly Frequency for Over-Identified Systems

	Bias-Uncorrected Estimators					Bias-Corrected Estimators				
	b_0	b_1	b_2	b_3	Jstat	b_0^{bc}	b_1^{bc}	b_2^{bc}	b_3^{bc}	Jstat
Panel A: $E_{t-1}\left[r_{m,t} - b_0 - b_1 \upsilon_t\right] = 0$										
1947:01-2018:12	.01 (5.14)	−1.38 (−1.17)			9.1 (.01)					
1947:01-1982:12	.01 (1.89)	−.63 (−.12)			11.4 (.003)					
1983:01-2018:12	.02 (4.66)	−2.28 (−1.96)			3.22 (.20)					
Panel B: $E_{t-1}\left[r_{m,t} - b_0 - b_1 \upsilon_{t-1}\right] = 0$										
1947:01-2018:12	.01 (7.62)	−1.60 (−3.14)			7.51 (.023)	.01 (7.85)	−1.88 (−3.56)			7.56 (.023)
1947:01-1982:12	.01 (2.58)	1.37 (.47)			11.6 (.003)	.01 (2.47)	1.64 (.57)			11.6 (.003)
1983:01-2018:12	.01 (5.64)	−1.67 (−3.28)			1.75 (.42)	.01 (5.87)	−2.00 (−3.65)			1.75 (.42)
Panel C: $E_{t-1}\left[r_{m,t} - b_0 - b_1 \upsilon_t - b_2 \upsilon_{t-1} - b_3 r_{m,t-1}\right] = 0$										
1947:01-2018:12	.002 (.38)	8.32 (1.69)	−4.90 (−1.57)	.13 (1.77)	.16 (.92)	.001 (.12)	10.9 (1.88)	−6.76 (−1.79)	.12 (1.51)	.13 (.92)
1947:01-1982:12	.001 (.08)	8.48 (.51)	−1.41 (−.18)	.11 (.85)	12.1 (.002)	.000 (.04)	9.68 (.58)	−2.23 (−.28)	.12 (.89)	12.3 (.002)
1983:01-2018:12	.005 (.82)	3.49 (.94)	−2.04 (−1.06)	.08 (.92)	.52 (.77)	.005 (.70)	4.62 (1.21)	−2.84 (−1.41)	.07 (.82)	.34 (.77)

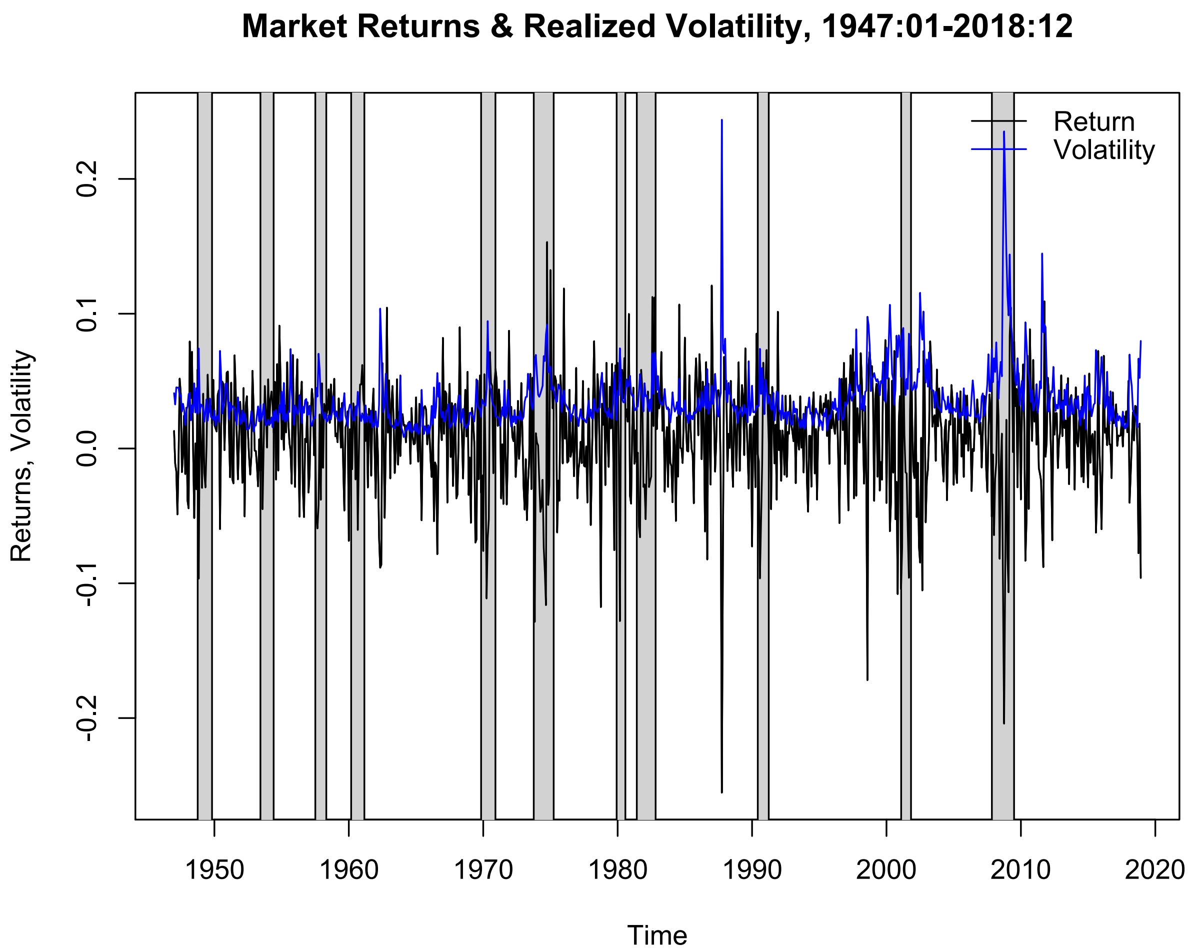
The table reports estimation results at the monthly frequency for over identified systems. The short term risk free rate and the default spread are used as instruments, in addition to the lagged integrated variance.

Table VI: Estimation Results For Subperiods

Sample	Bias-Uncorrected Estimators		Bias-Corrected Estimators	
	b_0	b_1	b_0^{bc}	b_1^{bc}
Panel A: $E_{t-1}\left[r_{m,t} - b_0 - b_1 \upsilon_t\right] = 0$				
Expansions	.02 (3.03)	−3.88 (−1.14)		
Recessions	.01 (2.07)	−1.90 (−1.90)		
Panel B: $E_{t-1}\left[r_{m,t} - b_0 - b_1 \upsilon_{t-1}\right] = 0$				
Expansions	.01 (7.18)	−1.00 (−1.78)	.01 (7.61)	−1.23 (−2.49)
Recessions	.01 (1.82)	−1.42 (−1.55)	.01 (1.93)	−1.57 (−1.73)

The table reports estimation results at the monthly frequency for exactly identified systems.

Figure 1



Notes: The figure plots the time series of the monthly stock market returns and realized volatility over the full sample period 1947:01-2018:12. The realized variance for a given month is computed as the sum of squares of the daily returns within the month.