Inference for the Price of Volatility Risk Under Weak Identification

Xu Cheng, Eric Renault, and Paul Sangrey

May 9, 2018

Two key questions at the very heart of finance are what are the risks investors face and what are the prices of those risks. Two leading risks are equity risk and volatility risk. Although the literature has shown that volatility risk clearly matters, constructing beliefs concerning the price of volatility risk from the data has proven quite difficult. What we want is a good estimator for this parameter and a strategy for credible inference regarding it.

We take the model from Khrapov and Renault (2016) and use it to estimate the relevant parameters, which we derive below. We use a finite version of spectral GMM, which forms moment conditions from the characteristic function. Out data we use are the bivariate series $\left(r_{t+1}, \sigma^2_{t+1}\right)$. r_{t+1} is the daily return on some asset, and we use its associated realized volatility for σ^2_{t+1}.

Moving forward, we sketch the model developed in Khrapov and Renault (2016) and derive the associated moment conditions. We then provide a series of sufficient conditions for valid inference.

1 The Model

Add discussion of the dynamics of the process.

We have a pricing kernel $M_{t,t+1}(\theta)$ which allows us to characterize the price P_t at time t of any payoff at time $t+1$ of a function f and information set $\mathcal{F}_t$, $f\left(r_{t+1}, \sigma^2(t+1) \,\middle|\, \mathcal{F}_t\right)$. We use $\mathbb{Q}$ to denote the risk-neutral measure.

Definition 1. The Model Setup

$$P_t = \mathbb{E}\left[M_{t,t+1}(\theta) f\left(r_{t+1}, \sigma^2_{t+1} \,\middle|\, \mathcal{F}_t\right) \,\middle|\, \mathcal{F}_t\right] = \mathbb{E}_{\mathbb{Q}}\left[f\left(r_{t+1}, \sigma^2_{t+1} \,\middle|\, \mathcal{F}_t\right) \,\middle|\, \mathcal{F}_t\right] \tag{1}$$

To make the problem tractable, we assume that the problem is Markov and that there is no Granger causality from return to volatility. This implies the conditional probability distribution of $\sigma^2_{t+1} | \mathcal{F}_t$ equals the conditional probability distribution of $\sigma^2_{t+1} | \sigma^2_t$. Consequently, we can write down our model in the risk-neutral measure using some functions, $a_{\mathbb{Q}}(u), b_{\mathbb{Q}}(u)$, and $\alpha_{\mathbb{Q}}(v), \beta_{\mathbb{Q}}(v), \& \gamma_{\mathbb{Q}}(v)$, as the following two equations in terms of the Laplace transforms of the probability distributions.

Definition 2. The Risk-Neutral Model

$$\mathbb{E}_{\mathbb{Q}}\left[\exp\left(-x\sigma_{t+1}^2\right) \,\middle|\, \sigma_t^2\right] = \exp\left(-a_{\mathbb{Q}}(x)\sigma_t^2 - b_{\mathbb{Q}}(x)\right) \tag{2}$$

$$\mathbb{E}_{\mathbb{Q}}\left[\exp\left(-xr_{t+1}\right) \,\middle|\, \sigma_t^2\sigma_{t+1}^2\right] = \exp\left(-\alpha_{\mathbb{Q}}(x)\sigma_{t+1}^2 - \beta_{\mathbb{Q}}(x)\sigma_t^2 - \gamma_{\mathbb{Q}}(x)\right) \tag{3}$$

We assume that the volatility follows an autoregressive gamma process—ARG(1), and so its physical measure dynamics are governed by following equations.

$$a_{\mathbb{P}}(x) = \frac{\rho x}{1+cx} \tag{4}$$

$$b_{\mathbb{P}}(x) = \delta \log\left(1+cx\right) \tag{5}$$

$$\rho \in [0,1), c > 0, \delta > 0 \tag{6}$$

The persistence is governed by ρ, the mean by δ, and c is a scaling factor for the volatility as can be seen in the formula for σ_{t+1}^2's conditional mean.

$$\mathbb{E}\left[\sigma_{t+1}^2 \,\middle|\, \sigma_t^2\right] = c\delta + \rho\sigma_t^2 \tag{7}$$

Assuming the measure change preserves the general structure between the risk-neutral and physical measures implies Equation (8). We also assume that $\left[\frac{\psi}{\phi}\right]^2 \approx \frac{\mathbb{E}\left[\sigma_{t+1}^2 \,\middle|\, \mathcal{F}_t\right]}{\mathbb{V}ar\left[r_{t+1} \,\middle|\, \mathcal{F}_t\right]}$, which enables our approximation of σ_{t+1}^2 by the realized volatility.

$$\mathbb{E}_{\mathbb{P}}\left[\exp\left(-xr_{t+1}\right) \,\middle|\, \sigma_t^2, \sigma_{t+1}^2\right] = \exp\left(-a(x)r_{t+1}^2 - \beta(x)\sigma_t^2 - \gamma(x)\right) \tag{8}$$

To estimate this equation, we need to know all of the relevant functions. The parametric structure of the problem and some algebra implies the following.

$$a(x) = \frac{\rho x}{1+cx} \tag{9}$$

$$b(x) = \delta \log\left(1+cx\right) \tag{10}$$

$$\alpha(x) = \psi x - \frac{1}{2}x^2(1-\phi^2) \tag{11}$$

$$\phi(x) = x\alpha_{\mathbb{Q}}\left(-\frac{\phi}{\sqrt{c[1+\rho]}}\right) \tag{12}$$

$$\gamma(x) = xb_{\mathbb{Q}}\left(-\frac{\phi}{\sqrt{c[1+\rho]}}\right) \tag{13}$$

In last two of the above equations we have the risk-neutral $\alpha_{\mathbb{Q}}(x)$ and $\beta_{\mathbb{Q}}(x)$ functions which we have not defined. To solve for them we drive the implied stochastic discount factor and make the appropriate measure change. We parameterize the SDF in terms of price of volatility risk—π—and the price of equity risk—θ. The SDF satisfies the following equation for some functions $m_0(\cdot)$ and $m_1(\cdot)$.

$$M_{t,t+1}(\theta) = \exp\left(-r_{f,t}\right)\exp\left(m_0(\theta) + m_1(\theta)\sigma_t^2 - \pi\sigma_{t+1}^2 - \theta r_{t+1}\right) \tag{14}$$

Then by the law of iterated expectations and some algebra.

$$\mathbb{E}\left[\exp\left(m_0(\theta) + m_1(\theta)\sigma_t^2 - \pi\sigma_{t+1}^2 - \theta r_{t+1}\right)\exp\left(-\alpha(\theta)\sigma_{t+1}^2 - \phi(\theta)\sigma_{t+1}^2 - \gamma(\theta)\right)\,\middle|\,\mathcal{F}_t\right] = 1 \tag{15}$$

This implies the two unspecified functions are as follows.

$$m_0(\theta) = \gamma(\theta) + b\left(\alpha\left(\theta\right) + \pi\right) \tag{16}$$

$$m_1(\theta) = \phi(\theta) + a\left(\alpha(\theta) + \pi\right) \tag{17}$$

Now we can solve for $\alpha_{\mathbb{Q}}(x)$ and $\beta_{\mathbb{Q}}(x)$.

$$a_{\mathbb{Q}}(x) = a\left(x + \pi + \alpha(\theta)\right) - a\left(\pi + \alpha(\theta)\right) \tag{18}$$

$$b_{\mathbb{Q}}(x) = b\left(x + \pi + \alpha(\theta)\right) - b\left(\pi + \alpha(\theta)\right) \tag{19}$$

We substitute them back into Equation (12) and Equation (13) eliminating $\alpha_{\mathbb{Q}}(x)$ and $\beta_{\mathbb{Q}}(x)$.

$$a(x) = \frac{\rho x}{1 + cx} \tag{20}$$

$$b(x) = \delta\log\left(1 + cx\right) \tag{21}$$

$$\alpha(x) = x\left(\frac{\phi}{\sqrt{c(1+\phi)}} + (1-\phi^2)\left(\theta - \frac{1}{2}\right)\right) - \frac{1}{2}x^2(1-\phi^2) \tag{22}$$

$$\beta(x) = x\left(a\left(-\frac{\phi}{\sqrt{c(1+\rho)}} + \pi + \alpha(\theta)\right) - a\left(\pi + \alpha(\theta)\right)\right) \tag{23}$$

$$\gamma(x) = x\left(b\left(-\frac{\phi}{\sqrt{c(1+\rho)}} + \pi + \alpha(\theta)\right) - b\left(\pi + \alpha(\theta)\right)\right) \tag{24}$$

The set of parameters we want to estimate is $\eta \coloneqq \{c, \rho, \delta, \phi, \pi, \theta\}$.

2 Finite Spectral GMM

We derive a set of moment conditions from the characteristic function above by evaluating it at a grid of points in $[0,1] \times i[0,1]$. That is we can define a function $g_t(x, \eta)$

Definition 3. Moment Conditions

$$g_t(x,\eta) \coloneqq Z_t \otimes \begin{bmatrix} \exp(-x\sigma_{t+1}^2) - \exp\left(-a(x)\sigma_t^2 - b(x)\right) \\ \exp\left(-xr_{t+1}\right) - \exp\left(-\alpha(x)\sigma_{t+1}^2 - \beta(x)\sigma_t^2 - \gamma(x)\right) \end{bmatrix} \tag{25}$$

Where the instruments are given by definition 4 for complex unit i.

Definition 4. Instruments

$$Z_t = \left[1, \exp\left(-i\sigma_{t-1}^2\right), \exp\left(-i\sigma_{t-2}^2\right)\right] \tag{26}$$

The implied unconditional moment restrictions are the following.

$$\mathbb{E}\begin{bmatrix} \mathrm{Re}(g_t(x,\eta)) \\ \mathrm{Im}(g_t(x,\eta)) \end{bmatrix} = 0 \tag{27}$$

The optimal weighting matrix has its standard form as the precision matrix of the moments as long as we choose a finite gird for x. If we use the entire continuum, handling the weights becomes more delicate. So we use only finitely many moments for now.

Lemma 1 (Identified Set)**.** *Let $\eta_0 \coloneqq (\rho_0, c_0, \phi_0, \pi_0, \theta_0)$. Let the domain H for η be defined as follows.*

$$\begin{aligned} &H \coloneqq \left\{\eta \in [0,\overline{\rho}] \otimes [0,\overline{c}] \otimes [-1,1] \otimes \left[-\underline{\theta}_1, \overline{\theta}_1\right] \otimes \left[-\underline{\theta}_2, \overline{\theta}_2\right]\right\} \\ &\quad 0 < \overline{c}, -\underline{\theta}_1, \overline{\theta}_1, -\underline{\theta}_2 < \infty, 0 < \overline{\rho} < 1 \end{aligned} \tag{28}$$

$$\textit{Given } \eta_0 \in H, \textit{ the moment conditions } g_t(x,\eta) \textit{ identify } \begin{cases} \eta_0 & \textit{if } \phi \neq 0 \\ \eta_0 \setminus \pi_0 & \textit{if } \phi = 0 \end{cases} \tag{29}$$

In addition, if $\phi = 0$, the GMM criterion function is independent of θ.

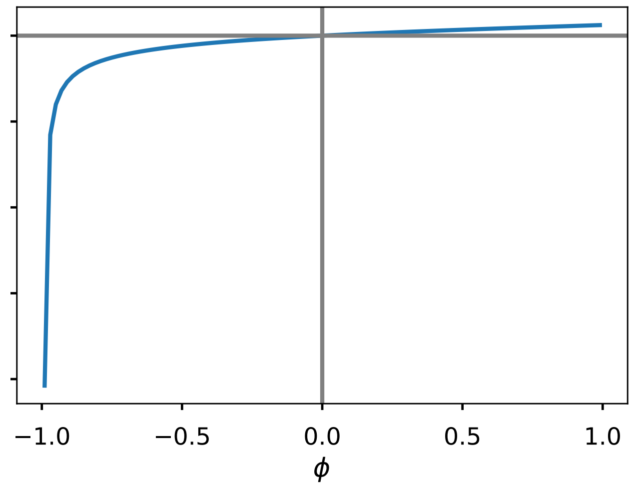
If we plug in the estimated values of the parameters from Khrapov and Renault (2016) into $\frac{\partial \phi}{\partial \pi}$ and plot it as a function of ϕ, we get the following. The scale is omitted because it is not meaningful. As can clearly be seen in Figure 1, there is a zero when $\phi = 0$.

Lemma 2 (Uniform Convergence under Strong Identification)**.** *Let H be the identified set defined by Equation (29). Further assume that $\phi_0 \neq 0$. Let $\overline{g}_T$ be the sample moment condition defined above, and $\mathcal{W}_T$ be the associated optimal weight matrix estimator. Then we have the following convergence.*

$$\sup_{\eta \in H} \|\overline{g}_T(\eta) - \mathbb{E}\left[g\left(\eta \mid \gamma_0\right)\right]\|_{Fro} \to_p 0 \tag{30}$$

$$\sup_{\eta \in H} \|\mathcal{W}_T(\eta) - \mathbb{E}\left[\mathcal{W}\left(\eta \mid \gamma_0\right)\right]\| \to_p \tag{31}$$

Figure 1: Derivative of $\gamma(x)$ with respect to θ



By the above arguments, we have a consistent estimator for η and the optimal weight matrix $\mathcal{W} := (\mathbb{E}[gg'])^{-1}$, and we will assume that the true value η_0 is in the interior of its sample space H.[1] Let $G := \mathbb{E}\left[\frac{\partial}{\partial \eta} g\right]$ Clearly, g is continuously differentiable, and its derivative G is continuous. In addition, by the identification discussion $G' W \nabla G$ is nonsingular.

Assumption 1 (Weak Dependence)**.** $z_t := \begin{pmatrix} r_{t+1} \\ \sigma^2_{t+1} \end{pmatrix}$ are α-mixing with $\alpha_t = O\left(T^{-5}\right)$

Since, $\|g_t\|$ is almost surely bounded by 1 it has all of its moments and z_t being α-mixing implies g_t is as well by the central limit theorem for strongly mixing process $\sqrt{T}\bar{g}_T(\eta^*) \to_d N\left(0, \mathbb{E}[\mathcal{W}]^{-1}\right)$ as required. Consequently, by Newey and McFadden (1994, theorem 3.2) we have convergence in distribution as well as convergence in probability.

Theorem 3 (Inference for η under Strong Identification)**.** *Assume that $\phi_0 \in (-1, 1) \setminus 0$, $\rho_0 \in [0, 1)$, and $c_0 > 0$. Further assume that the data are ergodic, stationary, and satisfy assumption 1. Then the following convergence in distribution holds.*

$$\sqrt{T}(\widehat{\eta}_T - \eta_0) \to_d N\left(0, \left(G'\mathbb{E}[\mathcal{W}]G\right)^{-1}\right) \tag{32}$$

3 Weak Identification Setup

In this section, take the model described in the previous sections and place it in the setup of Andrews and Cheng (2014) so that we ca analyze the effects of possible lack of identification in the model in a nice clean way. The goal here is to perform valid inference for π, θ even when ϕ might be zero.

From the discussion above, we can collect the parameters discussed above into a parameter vector of the following form,i.e. recall the following: $\eta = \{\rho, c, \delta, \phi, \pi, \theta\}$ To write it in the notation of Andrews and Cheng (2014), we partition η into three subsets.

1. Throughout we will use subscript 0 to denote true values for parameters.

$$\phi \coloneqq \phi \in (-1, 1) \tag{33}$$

$$\zeta \coloneqq \{\rho, c, \delta, \theta\} \in [0, 1) \times \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathbb{R} \tag{34}$$

$$\pi \coloneqq \pi \in \mathbb{R} \tag{35}$$

Let H be the set of possible η, that as defined above. It is worth noting that the parameter space has a product form, i.e. the values do not affect the valid values of the other parameters.

In this environment, π is not identified when $\phi = 0$. Both ϕ and ζ are always identified, and ζ does not affect the identification of π.

Let $Q_T(\eta)$ be the GMM criterion function, then the GMM estimator $\hat{\eta}_T$ satisfies the following.

$$\widehat{\eta}_T \in H \text{ and } Q_T(\hat{\eta}_T) = \inf_{\eta \in H} Q_T(\eta) + o\left(T^{-1}\right) \tag{36}$$

Now that we have defined the parameters, we can characterize the set of assumptions necessary for valid inference. We will work through the assumptions described in Andrews and Cheng (2014). The set of necessary assumptions is relatively complicated because we have to characterize the asymptotic distribution under several different estimation strengths simultaneously, and the assumptions required to do that differ in the various cases. In what follows, we will use

The first assumption specifies the basic identification problem. It also provides conditions that are used to determine the probability limit of the GMM estimator, when it exists, under all categories of drifting sequences of distributions. Let ξ index the part of the distribution of the data r_{t+1}, σ^2_{t+1} that is not determined by the moment equations. In general, it is a (likely infinite-dimensional) nuisance parameter that affects the distribution of the data.

We collect the parameters that we are estimating η and the nuisance parameter ξ into one parameter, γ and associated parameter space Γ. In the previous discussion we characterized the parameter spaces in a non-compact fashion, let H^* be a compact subset of H, where the true parameter values live.

Definition 5. Complete Parameter Space

$$\Gamma \coloneqq \{\gamma = (\eta, \xi) \,|\, \eta \in H, \xi \in \Xi\} \tag{37}$$

We characterize these drifting sequences of distributions by sequences of true parameters $\gamma_T \coloneqq (\eta_T, \phi_T)$.

TODO Add discussion of the limiting process. Verify that the assumptions on the parameter space hold. Discuss what happens if we lack identification and hence cannot consistently estimate the parameter.

Theorem 4 (Inference for η under Weak Identification)**.** *Let that $\phi_0 \in (\underline{\phi}_0, 1)$, for some $\underline{\phi}_0 > -1$. $\rho_0 \in [0, 1)$, and $c_0 > 0$.*

TODO Add Conclusion

References

Andrews, Donald W.K. 1991. "An Empirical Process Central Limit Theorem for Dependent Non-identically Distributed Random Variables." *Journal of Multivariate Analysis* 38 (2): 187–203.

Andrews, Donald W.K., and Xu Cheng. 2014. "GMM Estimation and Uniform Subvector Inference with Possible Identification Failure." *Econometric Theory* 30 (2): 287–333.

Khrapov, Stanislav, and Eric Renault. 2016. *Affine Option Pricing Model in Discrete Time.* Brown University.

Newey, Whitney K., and Daniel McFadden. 1994. "Large Sample Estimation and Hypothesis Testing," 4:2111–2245. Handbook of Econometrics, Chapter 36. Elsevier.

Appendix A Assumptions

In what follows, three sets of drifting sequences $\{\gamma_T\}$ are key.

Definition 6. Drifting Sequence Parameter Spaces

$$\Gamma(\gamma_0) := \{\{\gamma_T \in \Gamma\} \mid \gamma_T \to \gamma_0 \in \Gamma\} \tag{38}$$

$$\Gamma(\gamma_0, 0, b) := \left\{\{\gamma_T\} \in \Gamma(\gamma_0) \,\middle|\, \phi_0 = 0 \text{ and } \sqrt{T}\phi_T \to b \in (\mathbb{R} \cup \{\pm\infty)\right\} \tag{39}$$

$$\Gamma(\gamma_0, \infty, b_0) := \left\{\{\gamma_T\} \in \Gamma(\gamma_0) \,\middle|\, \sqrt{T}\|\phi_T\| \to \infty \text{ and } \frac{\phi_T}{\|\phi_T\|} \to b_0\right\} \tag{40}$$

These are the standard GMM regularity conditions appropriately adjusted for the lack of identification when $\phi = 0$.

Assumption 2 (GMM 1)**.** (i) If $\phi_0 = 0$, $\bar{g}_T(\eta)$ and $\mathcal{W}_T(\eta)$ do not depend on π for all $\eta \in H$, for all $T \geq 1$, and for all $\gamma^* \in \Gamma$.

(ii) If $\{\gamma_T\} \in \Gamma(\gamma_0)$, $\sup_{\eta \in H}\|\bar{g}_T(\eta) - \mathbb{E}[g(\eta \mid \gamma_0)]\| \to_p 0$ and $\sup_{\eta \in H}\|\mathcal{W}_T(\eta) - \mathbb{E}[\mathcal{W}(\eta \mid \gamma_0)]\| \to_p 0$.

(iii) When $\phi_0 = 0$, $g_0(\phi, \zeta, \pi \mid \gamma_0) = 0$ if and only if $\phi = \phi_0$ and $\zeta = \zeta_0$ for all $\pi \in \Pi$ and for all $\gamma_0 \in \Gamma$.

(iv) When $\phi_0 \neq 0$, $g_0(\eta \mid \gamma_0) = 0$ if and only if $\eta = \eta_0$ for all $\gamma_0 \in \Gamma$.

(v) $g_0(\eta \mid \gamma_0)$ is continuously differentiable in η on H with partial derivatives with respect to η and ξ denoted by $g_\eta(\theta \mid \gamma_0) \in R^{k \times d_\eta}$ and $g_\xi(\eta \mid \gamma_0) \in R^{k \times d_\xi}$, respectively.

(vi) $\mathcal{W}(\eta \mid \gamma_0)$ is continuous in η on H for all $\gamma_0 \in \Gamma$.

(vii) $0 < \lambda_{\min}(\mathcal{W}(\xi_0, \pi \mid \gamma_0)) \leq \lambda_{\max}(\mathcal{W}(\xi_0, \pi \mid \gamma_0)) < \infty$, $\forall \pi \in \Pi$, for all $\gamma_0 \in \Gamma$.

(viii) $\lambda_{\min}(g_\xi(\xi_0, \pi \mid \gamma_0)' \mathcal{W}(\xi_0, \pi \mid \gamma_0)\, g_\xi(\xi_0, \pi \mid \gamma_0)) > 0$, for all $\pi \in \Pi$, and for all $\gamma_0 \in \Gamma$ with $\phi_0 = 0$.

(ix) $\Xi(\pi)$ is compact for all $\pi \in \Pi$, and both Π and H are compact.

(x) For all $\epsilon > 0$, there exits a $\delta > 0$ such that $d_H(\Xi(\pi_1), \Xi(\pi_2)) < \epsilon$ for $\pi_1, \pi_2 \in \Pi$ with $\|\pi_1 - \pi_2\| < \delta$, where $d_H(\cdot)$ is the Hausdorff metric.

Assumption 3 (GMM 2*)**.** (i) $\bar{g}_T(\eta)$ is continuously differentiable in η for all $T \geq 1$.

(ii) If $\{\gamma_T\} \in \Gamma(\gamma_0, 0, b)$, $\sup_{\{\eta \in H \mid \|(\phi, \zeta')' - (\phi_T, \zeta_0')\| \leq \delta_T\}} \left\| \frac{\partial}{\partial(\phi, \zeta')'} \bar{g}_T(\eta) - \mathbb{E}\left[g_{(\phi,\zeta')'}(\eta) \,\middle|\, \gamma_0\right] \right\| = o_p(1)$ for all deterministic sequences $\delta_T \to 0$.

(iii) Let $H_T := \left\{\eta \in H \,\middle|\, \left\|(\phi, \zeta_) - (\phi_T, \zeta_T)\right\| \leq \delta_T\|\beta_T\| \text{ and } \|\pi - \pi_T\| \leq \delta_T\right\}$. Let δ_T be a deterministic sequence that converges to zero. If $\{\gamma_T\} \in \Gamma(\gamma_0, \infty, b_0)$, then we have the following asymptotic behavior. $\sup_{\eta \in H_T} \left\| \left(\frac{\partial}{\partial \eta'} \bar{g}_T - \mathbb{E}[g_\eta(\eta) \mid \gamma_0]\right) \operatorname{diag}\left(1_{1+d_\zeta}', (1/\phi_T)_{d_\pi}'\right) \right\| = o_p(1)$.

Once we have GMM 1 and GMM 2*, we use GMM 3 to derive the asymptotic distribution under weak and semi-strong identification. These conditions will be characterized using the expected derivative of the population moment conditions.

Definition 7.

$$K_{T,g}\left(\eta \,|\, \gamma^*\right) \coloneqq \frac{1}{T}\sum_{i=1}^{T}\frac{\partial}{\partial \phi^*}\mathbb{E}\left[g(W_T,\eta)\,|\,\gamma^*\right] \tag{41}$$

Assumption 4 (GMM 3)**.** (i) $\overline{g}_T(\eta) = \frac{1}{T}\sum_{i=1}^{T} g(W_T,\eta)$ for some function $g(W_T,\eta) : \mathbb{R}^{k\times k} \times H \to \mathbb{R}^k$.

(ii) $\mathbb{E}\left[g(W_T,\beta_0,\zeta^*,\pi)\,|\,\gamma^*\right] = 0$ for all $\pi \in \Pi$ and for all $i \geq 1$ if $\gamma^* = (0,\zeta^*,\pi^*,\xi^*) \in \Gamma$.

(iii) If $\{\gamma_T\} \in \Gamma(\gamma_0, 0, b)$, $\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\left(g(W_T,\zeta_{0,T},\pi_T) - \mathbb{E}\left[g(W_T,\zeta_{0,T},\pi_T)\,|\,\gamma_T\right]\right) \to_d N\left(0, \aleph(\gamma_0)\right)$, where $\aleph(\gamma_0)$ is a $k \times k$ matrix.

(iv)
1. $K_{T,g}\left(\eta \,|\, \gamma^*\right)$ exists for all $\{\eta,\gamma^*\} \in (H_\delta \times \Gamma_0)$ and for all $T \geq 1$.
2. $K_{T,g}\left(\phi_T,\zeta_T,\pi \,|\, \widetilde{\gamma}_T\right)$ uniformly converges to some non-stochastic matrix-valued function $K_g\left(0,\zeta_0,\pi \,|\, \gamma_0\right)$ over $\pi \in \Pi$ for all deterministic sequences $\{\phi_T,\zeta_T,\widetilde{\gamma}_T\}$ satisfying $\widetilde{\gamma}_T \in \Gamma$, $\widetilde{\gamma}_T \to \gamma_0 \coloneqq (0,\zeta_0,\pi_0,\xi)$, $\{\phi_T,\zeta_T,\pi\} \in H$ and $\{\phi_T,\zeta_T\} \to (0,\zeta_0)$.
3. $K_g\left(\phi_0,\zeta_0,\pi \,|\, \gamma_0\right)$ is continuous on Π for all $\gamma_0 \in \Gamma$ with $\phi_0 = 0$.

(v) $K\left(\phi_0,\zeta_0,\pi \,|\, \gamma_0\right) = g_{\phi,\zeta}\left(\phi_0,\pi \,|\, \gamma_0\right) x$ for some $x \in \mathbb{R}^{1+d_\zeta}$ if and only $\pi = \pi_0$.[2]

(vi) If $\{\gamma_T\} \in \Gamma(\gamma_0,0,b)$, $\frac{1}{T}\sum_{i=1}^{T}\frac{\partial}{\partial \eta}\mathbb{E}\left[g\left(W_T,\eta_T\right)\,|\,\gamma_T\right] \to g_\eta\left(\eta_0 \,|\, \gamma_0\right)$.

Definition 8. g^*

$$g^*_{\phi,\zeta}\left(\phi_0,\zeta_0,\pi_1,\pi_2 \,|\, \gamma_0\right) = \left[g_\phi\left(\phi_0,\zeta_0,\pi_1 \,|\, \gamma_0\right), g_\phi\left(\phi_0,\zeta_0,\pi_2 \,|\, \gamma_0\right), g_\zeta\left(\phi_0,\zeta_0 \,|\, \gamma_0\right)\right] \in \mathbb{R}^{k\times(d_\zeta+2)} \tag{42}$$

Assumption 5 (GMM 4)**.** (i) ϕ is a scalar.

(ii) $g^*_{\phi,\zeta}\left(\phi_0,\zeta_0,\pi_1,\pi_2 \,|\, \gamma_0\right)$ has full column rank.

(iii) $\aleph(\gamma_0)$ is positive definite for all $\gamma_0 \in \gamma$ with $\phi_0 = 0$.

Appendix B Proofs

Lemma 1 (Identified Set)**.** *Let $\eta_0 \coloneqq (\rho_0, c_0, \phi_0, \pi_0, \theta_0)$. Let the domain H for η be defined as follows.*

$$\begin{gathered} H \coloneqq \left\{\eta \in [0,\overline{\rho}] \otimes [0,\overline{c}] \otimes [-1,1] \otimes \left[-\underline{\theta}_1, \overline{\theta}_1\right] \otimes \left[-\underline{\theta}_2, \overline{\theta}_2\right]\right\} \\ 0 < \overline{c}, -\underline{\theta}_1, \overline{\theta}_1, -\underline{\theta}_2 < \infty, 0 < \overline{\rho} < 1 \end{gathered} \tag{28}$$

$$\textit{Given } \eta_0 \in H, \textit{ the moment conditions } g_t(x,\eta) \textit{ identify } \begin{cases} \eta_0 & \textit{if } \phi \neq 0 \\ \eta_0 \setminus \pi_0 & \textit{if } \phi = 0 \end{cases} \tag{29}$$

In addition, if $\phi = 0$, the GMM criterion function is independent of θ.

Proof. Since the exponential function is a strictly positive function, and we are considering a grid of x values, a sufficient condition for $\rho, \delta,$ & c to be identified is for the relevant rows of $\nabla a(x)$ and

2. Since $\dim(\phi) = 1$, we can assume without loss of generality that the ω_0 from Andrews and Cheng (2014) equals 1.

$\nabla b(x)$ to equal zero only at η_0 which are satisfied if $\rho, c, \delta > 0$. Testing if if ϕ and θ are identified is somewhat trickier. Consider $\nabla \alpha(x)$. Since we are using a grid of x's, and the gradient of α is a nonlinear function of x, the first two rows of the Equation (43) imply ϕ is identified.

$$\frac{\partial \alpha(x)}{\partial(\phi, \theta, c)'} = \begin{bmatrix} \phi x^2 + x\left(-2\beta\left(\theta_2 - \frac{1}{2}\right) - \frac{\phi}{2\sqrt{c}(\phi+1)^{\frac{3}{2}}} + \frac{1}{\sqrt{c}\sqrt{\phi+1}}\right) \\ x\left(-\phi^2 + 1\right) \\ \frac{\phi x}{2c^{\frac{3}{2}}\sqrt{\phi+1}} \end{bmatrix} \tag{43}$$

The top line of Equation (43) can be solved for θ, which would create a local lack of identification for θ. However, this creates a linear relationship between θ and the other parameters and x. However, since we are using multiple x's we can avoid this issue. Consequently, $\phi \in (-1, 1], c > 0$, are sufficient to identify all of the parameters except for π, the price of volatility risk. We can use Equation (21) to identify ρ.

If $\phi = 0$, $\beta(x) = x\left(a(\pi + \alpha(\theta)) - a(\pi + \alpha(\theta))\right)$, and $\gamma(x) = x\left(b(\pi + \alpha(\theta)) - b(\pi + \alpha(\theta))\right)$. These both identically zero, and π does not show up in anywhere else in the criterion function.

□

Lemma 2 (Uniform Convergence under Strong Identification)**.** *Let H be the identified set defined by Equation (29). Further assume that $\phi_0 \neq 0$. Let $\overline{g}_T$ be the sample moment condition defined above, and $\mathcal{W}_T$ be the associated optimal weight matrix estimator. Then we have the following convergence.*

$$\sup_{\eta \in H} \left\| \overline{g}_T(\eta) - \mathbb{E}\left[g\left(\eta \,|\, \gamma_0\right)\right] \right\|_{Fro} \to_p 0 \tag{30}$$

$$\sup_{\eta \in H} \left\| \mathcal{W}_T(\eta) - \mathbb{E}\left[\mathcal{W}\left(\eta \,|\, \gamma_0\right)\right] \right\| \to_p \tag{31}$$

Proof. In this proof we rely heavily on the continuity of the moment conditions over their domain. This can be seen from simple inspection since we assumed that $\phi_0 \geq \underline{\phi} - 1$. Furthermore since H is compact, this continuity implies uniform continuity.

For any positive definite weight-matrix by Newey and McFadden (1994, Lemma 2.3) our criterion function has a unique optimum. The data, σ^2_{t+1}, r_{t+1}, are ergodic and stationary. Since the moment conditions are not redundant the optimal (GMM) weight matrix $\mathcal{W}$ is positive definite. In addition, g is continuous at each η, given the restrictions above and properties of characteristic functions imply that g is uniformly bounded. For convenience, we assume that the space of η is compact. This should not be an issue here because the parameters are either a priori bounded, such as ϕ or we have substantial a priori knowledge on their plausible magnitudes. Hence, Newey and McFadden (1994, Theroem 2.6) implies our estimator is consistent.

However, when we allow for weak identification late on, we need this convergence to be uniform. One straightforward way to show this is to show that our criterion function is globally Lipschitz in a set of high probability.

The other issue is that we need the weight matrix to converge uniformly to its expectation. Since the moments are continuous functions over their domain as is the square function. This convergence is uniform if and only if the matrix inverse is continuous.

Since we have a finite number of non-redundant moments, the minimum eigenvalue, $\lambda_{min}\left(\mathcal{W}\left(\phi_0, \zeta_0, \pi \,|\, \gamma_0\right)\right) > 0$, and so the matrix inverse is uniformly continuous in γ_0 with respect to the Frobenius norm, which is the sum of the eigenvalues. (Recall, that the eigenvalues of the inverse are the inverse of the eigenvalues.)

□

Theorem 4 (Inference for η under Weak Identification)**.** *Let that* $\phi_0 \in (\underline{\phi}_0, 1)$*, for some* $\underline{\phi}_0 > -1$*.* $\rho_0 \in [0, 1)$*, and* $c_0 > 0$*.*

TODO Add Conclusion

Proof. We prove this result by showing that Assumptions GMM 1-4 are satisfied.

Part 4.1. In this part, we show that GMM 1 is satisfied. To do this, we break GMM 1 down into three subsections. Assumptions GMM 1(i), GMM 1(ii), GMM 1(iii), and GMM 1(iv) state that when $\phi = 0$, the moment conditions contain no information regarding π, but when $\phi \neq 0$, the model is identified. This is what we show in Lemma 1. We further showed the relevant uniform convergence to verity GMM 1(ii) in Lemma 2.

The next two assumptions (GMM 1(v) and GMM 1(vi)) are technical conditions regarding the behavior of the moment conditions and weight matrix. Since our moment conditions are derived from an infinitely-differentiable characteristic function and the weight matrix is the optimal one, they both hold trivially.

The third subsection of Assumption GMM 1 concerns the weight matrix. Since we are using the inverse covariance matrix of valid non-redundant model, assumptions GMM 1(vii) and GMM 1(viii) automatically hold.

The last two assumptions, GMM 1(ix) and GMM 1(x) require that the parameter spaces do not vary too much with the parameters and are compact. Since H is compact, GMM 1(ix) holds trivially, and since it has has a product form, GMM 1(x) holds trivially as well.

Part 4.2. In this section, we show that the derivatives of the moment conditions have the correct behavior locally to the true parameters. We have to do this for the different classes of drifting sequences. We will do this by verifitying assumption 3. This is valid since Andrews and Cheng (2014) show that this is a sufficient condition for their Assumption GMM2, which is what we actually need.

Our moment conditions are sample averages of the characteristic function, they satisfy GMM 2*(i) automatically. Since characteristic functions are uniformly bounded, by the dominated convergence theorem we can interchange the expectation and derivative operators. Hence GMM 2*(ii) and GMM 2*(iii) are equivalent to the statements in terms of the moment conditions themselves mutatis mutandis. In addition, sice the derivate is a linear operator, we can pull it outside of the norm. The reason that the uniform law of large numbers in Part 4.1 does not trivially imply this result is because we are not considering sequences $\phi_T \to \phi_0$.

We create a mean value expansions around around (ϕ_0, ζ_T, π_T) of the sample moment condition and around (ϕ_0, ζ_0, π_0) for the population moment condition. (This is not the same in both cases, not is it the true parameter for the drifting sequence in the case of the sample moment condition.) In addition, also since we are considering continuous functions of compact spaces — the δ_T ball in $\mathbb{R}^{\dim(\eta)}$ — pointwise convergence implies uniform convergence, and so we only need to show pointwise convergence below.

$$\left\|\overline{g}_T(\phi_T, \zeta_T, \pi_T) - \mathbb{E}\left[g(\phi_T, \zeta_T, \pi_T) \,|\, \gamma_0\right]\right\| \tag{44}$$

We take a mean value expansion of both functions around η_0. The point at which the derivative in the two locations is taken may not be the same.

$$= \left\| \overline{g}_T(\phi_0, \zeta_0, \pi_0) + \frac{\partial}{\partial(\phi, \zeta, \pi)} \overline{g}_T(\widetilde{\phi}^s, \widetilde{\zeta}^s, \widetilde{\pi}^s) \left((\phi_0, \zeta_0, \pi_0) - (\phi, \zeta, \pi)\right) \right. \tag{45}$$

$$\left. -\mathbb{E}\left[g(\phi_0, \zeta_0, \pi_0) \,|\, \gamma_0\right] + \frac{\partial}{\partial(\phi, \zeta, \pi)} \mathbb{E}\left[g(\widetilde{\phi}^p, \widetilde{\zeta}^p, \widetilde{\pi}^p) \,\middle|\, \gamma_0 \right] \left((\phi_0, \zeta_0, \pi_0) - (\phi, \zeta, \pi)\right) \right\| \tag{46}$$

By the triangle inequality.

$$\leq \left\|\overline{g}_T(\phi_0, \zeta_0, \pi_0) - \mathbb{E}\left[g(\phi_0, \zeta_0, \pi_0) \,|\, \gamma_0\right]\right\| \tag{47}$$

$$+ \left\| \frac{\partial}{\partial(\phi, \zeta)} \overline{g}_T(\widetilde{\phi}^s, \widetilde{\zeta}^s, \widetilde{\pi}^s) \left((\phi_0, \zeta_0) - (\phi, \zeta)\right) - \frac{\partial}{\partial(\phi, \zeta)} \mathbb{E}\left[g(\widetilde{\phi}^s, \widetilde{\zeta}^p, \widetilde{\pi}^p) \,\middle|\, \gamma_0 \right] \left((\phi_0, \zeta_0) - (\phi, \zeta)\right) \right\| \tag{48}$$

$$+ \left\| \frac{\partial}{\partial \pi} \overline{g}_T(\widetilde{\phi}^s, \widetilde{\zeta}^s, \widetilde{\pi}^s) \left(\pi_0 - \pi\right) - \frac{\partial}{\partial \pi} \mathbb{E}\left[g(\widetilde{\phi}^p, \widetilde{\zeta}^p, \widetilde{\pi}^p) \,\middle|\, \gamma_0 \right] \left(\pi_0 - \pi\right) \right\| \tag{49}$$

By the uniform law of law numbers in Part 4.1, the first equation is $o_p(1)$. For $(\phi_0, \zeta_0) - (\phi, \zeta)$ small, the middle term is bounded by the quantity below.

$$\left\| \frac{\partial}{\partial(\phi, \zeta)} \overline{g}_T(\widetilde{\phi}^s, \widetilde{\zeta}^s, \widetilde{\pi}^s) - \frac{\partial}{\partial(\phi, \zeta)} \mathbb{E}\left[g(\widetilde{\phi}^s, \widetilde{\zeta}^p, \widetilde{\pi}^p) \,\middle|\, \gamma_0 \right] \right\| \left\|(\phi_0, \zeta_0) - (\phi, \zeta)\right\| \tag{50}$$

The first term is almost surely bounded, and the second term is less than δ_T by assumption, and so the product is $o_p(1)$.

The hard part is the third expression. Like before we can bound the pull the $\pi_0 - \pi$ term out of the equation. However, this term is no longer converges to zero. We will consider the two cases, separately. Throughout, we will refer to the behavior of the following equation, which bounds Equation (49).

$$\left\| \frac{\partial}{\partial \pi} \overline{g}_T(\widetilde{\phi}^s, \widetilde{\zeta}^s, \widetilde{\pi}^s) - \frac{\partial}{\partial \pi} \mathbb{E}\left[g(\widetilde{\phi}^p, \widetilde{\zeta}^p, \widetilde{\pi}^p) \,\middle|\, \gamma_0 \right] \right\| \left\|\pi_0 - \pi\right\| \tag{51}$$

In general, $\tilde{\pi}^s$ and $\tilde{\pi}^p$ can be arbitrarily far apart. However, since $\overline{g}_T \to_p \mathbb{E}\left[g \,|\, \gamma_0\right]$, and the limiting value is independent of π, the derivative does not depend upon π asymptotically by the dominated convergence theorem. This applies that the difference between the two derivatives evaluated at different π converges to zero. Consequently, Equation (51) is $o_p(1)$ and we have shown GMM 2*(ii).

$$\left\| \left\| \frac{\partial}{\partial \pi} \overline{g}_T(\widetilde{\phi}^s, \widetilde{\zeta}^s, \widetilde{\pi}^s) - \frac{\partial}{\partial \pi} \mathbb{E}\left[g(\widetilde{\phi}^p, \widetilde{\zeta}^p, \widetilde{\pi}^p) \,\middle|\, \gamma_0 \right] \right| (1, 1, T) \right\| \|\pi_0 - \pi\| \tag{52}$$

If we consider the setup in GMM 2*(iii), $\tilde{\pi}^s$ and $\tilde{\pi}^p$ are now close together. However, we need to show that Equation (52) is $o_p(1)$. Since we are considering the limiting behavior of a function with a continuous derivative, we can assume that the derivative is uniformly bounded without loss of generality. (The constant might depend upon the true value, but not T.) By a Taylor series expansion of $\overline{g}_T$ around the true value, $\frac{\tilde{\pi}^p}{\sqrt{T}} \to 0$, and $\|\pi_0 - \pi\| \propto \delta_T$ the result follows.

Part 4.3. Assumption GMM 3(i) is trivially satisfied, and we showed that GMM 3(ii) is satisfied in Section 2.

The conceptual idea driving the reulsts this section is that moment conditions for each T minus their conditional exceptions converge to a normal random variable. In other words, we are almost in a standard triangular C.L.T. setup with weak time-dependence.

In particular, both σ_t^2 and r_t are infinitely differentiable functions of the innovations to the volatility and return processes and g are infinitely differentiable functions of r_{t+1} and r_t and the innovations are i.i.d. across time by assumption. (Note, i.i.d. implies strong mixing of any size.) Consequently, g is near epoch dependent (NED) of any size as defined in Andrews (1991). (Take $s > 2$.) By Andrews (1991, Theorem 3), we have the necessary finite-dimensional convergence in distribution to a Gaussian random variable.

We now show that GMM 3(iv) holds. Clearly, $K_{t,g}\left(\eta \,|\, \gamma^*\right)$ always exists. It uniformly converges because the derivatives of the moments are continuous functions of the data and the parameters, the process is ergodic, and the characteristic function lives on a compact set. It is also clearly continuous. By the dominated convergence theorem, we can exchange the derivative and expectations. In addition, since g does not depend upon ϕ_T, the limiting behavior is independent of the value of ϕ_T. Hence GMM 1(iv) holds.

GMM 3(v) says the derivative of the moment function does not depend the true parameter ϕ if and only if $\pi = \pi_0$. We showed this in the proof of Part 4.2. GMM 3(vi) follows directly from the compactness of the parameter space and the continuity of the cross derivatives of g by the dominated convergence theorem.

Part 4.4. GMM 4(i) holds trivially. We verified GMM 4(ii) when we showed that π and ζ are strongly identified (Lemma 1). GMM 4(iii) holds because the identification conditions do not create any singularity in the asymptotic covariance matrix.

□