November 20, 2018

1 Robust Inference for Risk Prices

1.1 Asymptotic Distribution of the Reduced-Form Parameter

Write $\omega := (\omega_1, \omega_2, \omega_3)'$, where $\omega_1 = (\rho, c, \delta) \in O_1$, $\omega_2 = (\gamma, \beta, \psi) \in O_2$, and $\omega_3 = \zeta \in O_3$. The parameter space for ω is $O = O_1 \times O_2 \times O_3 \subset R^{d_\omega}$. The true value of ω is assumed to be in the interior of the parameter space.

Below we describe the estimator $\widehat{\omega} := (\widehat{\omega}_1, \widehat{\omega}_2, \widehat{\omega}_3)'$ and provide its asymptotic distribution. We estimate these parameters separately because ω_1 only shows up in the conditional mean and variance of σ_{t+1}^2, ω_2 only shows up in the conditional mean of r_{t+1}, and ω_3 only shows up in the conditional variance of r_{t+1}.

We first estimate $\omega_1 = (\rho, c)$ based on the conditional mean and variance of σ_{t+1}^2, which can be equivalently written as

$$\begin{aligned} E[\sigma_{t+1}^2|\sigma_t^2] &= A \text{ and } E[\sigma_{t+1}^4|\sigma_t^2] = B, \text{ where} \\ A &= \rho\sigma_t^2 + c\delta \text{ and } B = A^2 + \left(2c\rho\sigma_t^2 + c^2\delta\right). \end{aligned} \tag{1.1}$$

Because the conditional mean of σ_{t+1}^2 and σ_{t+1}^4 are linear and quadratic functions, respectively, of the conditioning variable σ_t^2, without loss of efficiency, they can be transformed to the unconditional moments

$$E[h_t(\omega_{10})] = 0, \text{ where } h_t(\omega_1) = [(1, \sigma_t^2) \otimes (\sigma_{t+1}^2 - A), (1, \sigma_t^2, \sigma_t^4) \otimes (\sigma_{t+1}^4 - B)]', \tag{1.2}$$

where ω_{10} represents the true value of ω_1. The two-step GMM estimator of ω_1 is

$$\widehat{\omega}_1 = \underset{\omega_1 \in O_1}{\arg\min} \left(T^{-1} \sum_{t=1}^T h_t(\omega_1)\right)' \widehat{V}_1 \left(T^{-1} \sum_{t=1}^T h_t(\omega_1)\right), \tag{1.3}$$

where $\widehat{V}_1$ is a consistent estimator of $V_1 = \sum_{m=-\infty}^{\infty} \mathbb{C}ov[h_t(\omega_{10}), h_{t+m}(\omega_{10})]$.

We estimate ω_2 by the generalized least squares (GLS) estimator because the conditional mean of r_{t+1} is a linear function of the conditioning variable σ_t^2 and σ_{t+1}^2 and the conditional variance is proportional to σ_{t+1}^2. The GLS estimator of ω_2 is

$$\begin{aligned} \widehat{\omega}_2 &= \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t y_t, \text{ where} \\ x_t &= \sigma_{t+1}^{-1}(1, \sigma_t^2, \sigma_{t+1}^2)' \text{ and } y_t = \sigma_{t+1}^{-1} r_{t+1}. \end{aligned} \tag{1.4}$$

We estimate ω_3 by the sample variance estimator

$$\widehat{\omega}_3 = T^{-1} \sum_{t=1}^{T} (y_t - \widehat{y}_t)^2, \text{ where } \widehat{y}_t = x_t' \widehat{\omega}_2. \tag{1.5}$$

Let

$$f_t(\omega) = \begin{pmatrix} h_t(\omega_1) \\ x_t(y_t - x_t'\omega_2) \\ (y_t - x_t'\omega_2)^2 \end{pmatrix} \in R^{d_f} \text{ and } V = \sum_{m=-\infty}^{\infty} \mathbb{C}\mathbf{ov}[f_t(\omega_0), f_{t+m}(\omega_0)]. \tag{1.6}$$

The estimator $\widehat{\omega}$ defined above is based on the first moment of $f_t(\omega)$. Let $\widehat{V}$ denote a heteroskeasticity and autocorrelation consistent (HAC) estimator of V. The estimator $\widehat{V}_1$ is a submatrix of $\widehat{V}$ associate with V_1.

Let P denote the distribution of the data $\mathcal{W} = \{W_t = (r_{t+1}, \sigma_{t+1}^2) : t \geq 1\}$ and $\mathcal{P}$ denote the parameter space of P. Note that the true values of the structural parameter and the reduced-form parameters are all determined by P. We allow P to change with T. For notational simplicity, the dependence on P and T is suppressed. Let $H_t(\omega_1) = \partial h_t(\omega_1)/\partial \omega_1'$.

Assumption R. The following conditions hold uniformly over $P \in \mathcal{P}$.
(i) $V^{-1/2}\{T^{-1/2}(\sum_{t=1}^{T} f_t(\omega_0) - \mathbb{E}[f_t(\omega_0)]\} \rightarrow_d N(0, I)$ and $\widehat{V} - V \rightarrow_p 0$.
(ii) $T^{-1}\sum_{t=1}^{T}(h_t(\omega_1) - \mathbb{E}[h_t(\omega_1)) \rightarrow_p 0$ and $T^{-1}\sum_{t=1}^{T}(H_t(\omega_1) - \mathbb{E}[H_t(\omega_1)]) \rightarrow_p 0$, $\mathbb{E}[H_t(\omega_1)]$ is continuous in ω_1, all uniformly over the parameter space of ω_1.
(iii) $T^{-1}\sum_{t=1}^{T}(x_t x_t' - \mathbb{E}[x_t x_t']) \rightarrow_p 0$.
(iv) $C^{-1} \leq \lambda_{\min}(A) \leq \lambda_{\max}(A) \leq C$ for $A = V, \mathbb{E}[H_t(\omega_{1,0})' H_t(\omega_{1,0})]), \mathbb{E}[x_t x_t'], \mathbb{E}[z_t z_t']$, where $z_t = (1, \sigma_t^2, \sigma_t^4)'$.

Assumptions R(i)-(iii) are the central limit theorem and the uniform law of large numbers applied to weakly dependent time series data. Assumption R(iv)-R(v) are typical regularity conditions for the identification and $T^{1/2}$ normality of the reduced-form parameter estimator.

Let $H(\omega_1) = \mathbb{E}[H_t(\omega_1)]$ and $\overline{H}(\omega_1) = T^{-1}\sum_{t=1}^{T} H_t(\omega_1)$. Define

$$\begin{aligned} F &= diag\{[H(\omega_{10})V_1^{-1}H(\omega_{10})]^{-1}H(\omega_{10})V_1^{-1}, \mathbb{E}[x_t x_t']^{-1}, 1\}, \\ \widehat{F} &= diag\{[\overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}\overline{H}(\widehat{\omega}_1)]^{-1}\overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}, [T^{-1}\sum_{t=1}^{T} x_t x_t']^{-1}, 1\}. \end{aligned} \tag{1.7}$$

The following Lemma provides the asymptotic distribution of the reduced-form parameter and a consistent estimator of its asymptotic covariance. Note that we put the asymptotic covariance on the left side of the convergence to allow the distribution of the data to change with sample size T.

Lemma 1.1 *Suppose Assumption R holds. The following results hold uniformly over $P \in \mathcal{P}$.*
(i) $\xi_T := \Omega^{-1/2}T^{-1/2}(\widehat{\omega} - \omega_0) \to_d \xi \sim N(0, I)$, *where* $\Omega = FVF'$.
(ii) $\widehat{\Omega} - \Omega \to_p 0$, *where* $\widehat{\Omega} = \widehat{F}\widehat{V}\widehat{F}'$.

1.2 Weak Identification

The true value of the structural parameter λ and the reduced-form parameter ω satisfies the link function $g(\lambda_0, \omega_0) = 0$.In a standard problem without any identification issue, we can estimate λ_0 by the minimum distance estimator $\widehat{\lambda} = (\widehat{\theta}, \widehat{\pi}, \widehat{\phi})$ that minimizes $Q_T(\lambda) = g(\lambda, \widehat{\omega})'W_T g(\lambda, \widehat{\omega})$ for some weighting matrix W_T and construct tests and confidence sets for λ_0 based on the asymptotic normal distribution of $T^{1/2}(\widehat{\lambda} - \lambda_0)$. However, this standard method does not work in the present problem when π_0 is only weak identified. In this case, $g(\lambda, \widehat{\omega})$ is almost flat in π and the minimum distance estimator of $\widehat{\pi}$ is not even consistent. To make the problem even more complicated, the inconsistency of $\widehat{\pi}$ has a spillover effect on $\widehat{\theta}$ and $\widehat{\phi}$, making the distribution of $\widehat{\theta}$ and $\widehat{\phi}$ non-normal even in large sample.

Before presenting the robust test, we first introduce some useful quantities and provide some heuristic discussions of the identification problem and its consequence. Let $G(\lambda, \omega)$ denote the partial derivative of $g(\lambda, \omega)$ wrt ω. Let $g_0(\lambda) = g(\lambda, \omega_0)$ and $G_0(\lambda) = G(\lambda, \omega_0)$ be the link function and its derivative evaluated at ω_0 and $\widehat{g}(\lambda) = g(\lambda, \widehat{\omega})$ and $\widehat{G}(\lambda) = G(\lambda, \widehat{\omega})$ be the same quantities evaluate at the estimator $\widehat{\omega}$. The delta method gives

$$\eta_n(\lambda) := T^{1/2}\left[\widehat{g}(\lambda) - g_0(\lambda)\right] = G_0(\lambda)\Omega^{1/2} \cdot \xi_T + o_p(1), \tag{1.8}$$

where $\xi_T \to_d N(0, I)$ following Lemma 1.1. Thus, $\eta_n(\cdot)$ weakly converges to a Gaussian process $\eta(\cdot)$ with covariance function $\Sigma(\lambda_1, \lambda_2) = G_0(\lambda_1)\Omega G_0(\lambda_2)'$.

Following (1.8), we can write $T^{1/2}\widehat{g}(\lambda) = \eta_n(\lambda) + T^{1/2}g_0(\lambda)$, where $\eta_n(\lambda)$ is the noise from the reduced-form parameter estimation and $T^{1/2}g_0(\lambda)$ is the signal from the link function. Under weak identification, $g_0(\lambda)$ is almost flat in λ, modelled by the signal $T^{1/2}g_0(\lambda)$ being finite even for $\lambda \neq \lambda_0$ and $T \to \infty$. Thus, the signal and the noise are of the same order of magnitude, yielding an inconsistent minimum distance estimator $\widehat{\lambda}$. This is in contrast with the strong identification scenario, where $T^{1/2}g_0(\lambda) \to \infty$ for $\lambda \neq \lambda_0$ as $T \to \infty$ and $g_0(\lambda_0) = 0$. In this case, the signal is so strong that the minimum distance estimator is consistent.

The identification strength of λ_0 is determined by the function $T^{1/2}g_0(\lambda)$. However, this function is unknown and cannot be consistently estimated (due to $T^{1/2}$). Thus, we take the conditional inference procedure as in Andrews and Mikusheva (2016) and view $T^{1/2}g_0(\lambda)$ as an infinite dimensionalnuisance parameter for the inference for λ_0. The goal is to control robust confidence set (CS) for λ_0 that has correct size asymptotically regardless of this unknown nuisance parameter.

1.3 Conditional QLR Test

We construct a confidence set for λ by inverting the test $H_0 : \lambda = \lambda_0$ vs $H_1 : \lambda \neq \lambda_0$. The test statistic is a QLR statistic that takes the form

$$QLR(\lambda_0) = T\widehat{g}(\lambda_0)'\widehat{\Sigma}(\lambda_0,\lambda_0)^{-1}\widehat{g}(\lambda_0) - \min_{\lambda\in\Lambda} T\widehat{g}(\lambda)'\widehat{\Sigma}(\lambda,\lambda)^{-1}\widehat{g}(\lambda), \tag{1.9}$$

where $\widehat{\Sigma}(\lambda_1,\lambda_2,) = \widehat{G}(\lambda_1)\widehat{\Omega}\widehat{G}(\lambda_2)'$ and $\widehat{\Omega}$ is the consistent estimator of Ω defined above.

Andrews and Mikusheva (2016) provide the conditional QLR test in a nonlinear GMM problem, where $\widehat{g}(\lambda)$ is replaced by a sample moment. The same method can be applied to the present nonlinear minimum distance problem. Following AM, we first project $\widehat{g}(\lambda)$ onto $\widehat{g}(\lambda_0)$ and construct a residual process

$$\widehat{r}(\lambda) = \widehat{g}(\lambda) - \widehat{\Sigma}(\lambda,\lambda_0)\widehat{\Sigma}(\lambda_0,\lambda_0)^{-1}\widehat{g}(\lambda_0). \tag{1.10}$$

The limiting distribution of $\widehat{r}(\lambda)$ and $\widehat{g}(\lambda_0)$ are Gaussian and independent. Thus, conditional on $\widehat{r}(\lambda)$, the asymptotic distribution of $\widehat{g}(\lambda)$ no longer depends on the nuisance parameter $T^{1/2}g_0(\lambda)$, making the procedure robust to all identification strength.

Specifically, we obtain the $1-\alpha$ conditional quantile of the QLR statistic, denoted by $c_{1-\alpha}(r,\lambda_0)$, as follows. For $b = 1,..,B$, we take independent draws $\eta_b^* \sim N(0,\widehat{\Sigma}(\lambda_0,\lambda_0))$ and produce a simulated process

$$g_b^*(\lambda) = \widehat{r}(\lambda) + \widehat{\Sigma}(\lambda,\lambda_0)\widehat{\Sigma}(\lambda_0,\lambda_0)^{-1}\eta_b^* \tag{1.11}$$

and a simulated statistic

$$QLR_b^*(\lambda_0) = T\widehat{g}(\lambda_0)'\widehat{\Sigma}(\lambda_0,\lambda_0)^{-1}\widehat{g}(\lambda_0) - \min_{\lambda\in\Pi} Tg_b^*(\lambda)'\widehat{\Sigma}(\lambda,\lambda)^{-1}g_b^*(\lambda). \tag{1.12}$$

Let $b_0 = \lceil (1-\alpha)B \rceil$, the smallest integer no smaller than $(1-\alpha)B$. Then the critical value $c_{1-\alpha}(r, \lambda_0)$ is the b_0^{th} smallest value among $\{QLR_b^*, b = 1, ..., B\}.\Omega$

To sum up, we execute the following steps for a robust CS for λ.

(i) Estimate the reduced-form parameter $\widehat{\omega} = (\widehat{\omega}_1, \widehat{\omega}_2, \widehat{\omega}_3)'$ following the estimators defined in (1.3) and (1.4). Obtain a consistent estimator of its asymptotic covariance $\widehat{\Omega} = \widehat{F}\widehat{V}\widehat{F}'$, where $\widehat{F}$ is define in (1.7) and $\widehat{V}$ is a heteroskedastic and autocorrelation consistent (HAC) estimator of V.

For $\lambda_0 \in \Lambda$, execute steps (ii)-(iv) below.

(ii) Construct the QLR statistic $QLR(\lambda_0)$ in (1.9) using $g(\lambda, \omega)$, $G(\lambda, \omega)$, $\widehat{\omega}$, and $\widehat{\Omega}$.

(iii) Compute the residual process $\widehat{r}(\lambda)$ in (1.10).

(iv) Given $\widehat{r}(\lambda)$, compute the critical value $c_{1-\alpha}(r, \lambda_0)$ described above.

(v) Repeat steps (ii)-(iv) for different values of λ_0. Construct a confidence set by collecting the null values that are not rejected, i.e., nominal level $1-\alpha$ confidence set for λ_0 is

$$CS_T = \{\lambda_0 : QLR_T(\lambda_0) \leq c_{1-\alpha}(r, \lambda_0)\}. \tag{1.13}$$

To obtain confidence intervals for each element of λ_0, one simple solution is to project the confidence set constructed above to each axis. The resulting confidence interval also has correct coverage. An alternative solution is to first concentrate out the nuisance parameters before apply the conditional inference approach above, see Section 5 of AM. However, this concentration approach only works when the nuisance parameter is strongly identified. In the present set-up, this approach does not work for θ and ϕ because the nuisance parameter π is weakly identified.

Assumption S. The following conditions hold over $P \in \mathcal{P}$, for any λ in its parameter space, and any ω in some fixed neighborhood around its true value.

(i) $g(\lambda, \omega)$ is twice continuously differentiable in ω and the second order derivative $G_\omega(\lambda, \omega)$ satisfies $||G_\omega(\lambda, \omega)|| \leq C$.

(ii) $C^{-1} \leq \lambda_{\min}(G(\lambda, \omega)'G(\lambda, \omega)) \leq \lambda_{\max}(G(\lambda, \omega)'G(\lambda, \omega)) \leq C$.

Lemma 1.2 *Suppose Assumption R and S hold. Then,*

$$\liminf_{T\to\infty} \inf_{P\in\mathcal{P}} \Pr(\lambda_0 \in CS_T) \geq 1-\alpha.$$

This Lemma states that the confidence set constructed by the conditional QLR test has correct uniform asymptotic size. Uniformity is important for this confidence set to cover the true parameter with a probability close to $1-\alpha$ in finite-sample. Most importantly, this uniform result is established over a parameter $\mathcal{P}$ that is large enough to allow the weak identification of the structural parameter λ.

Proof of Lemma 1.1. Under the assumption that (i) $\mathbb{E}(z_t z_t')$ has the smallest eigenvalue bounded away from 0 and (ii) $c > \varepsilon$ and $\delta > \varepsilon$ for some $\varepsilon > 0$, we not only have ω_{10} as an uniquely minimizer of $||\mathbb{E}[h_t(\omega_1)]||$ but also have a uniform positive lower bound for $||E[h_t(\omega_1)]||$ for $||\omega_1 - \omega_{10}|| \geq \varepsilon$. Thus, consistency of $\widehat{\omega}_1$ follows from standard arguments for the consistency of a GMM estimator under an uniform convergence of the criterion under Assumption R(ii).

Let $\overline{h}(\omega_1) = T^{-1}\sum_{t=1}^T h_t(\omega_1)$ and $\overline{H}(\omega) = T^{-1}\sum_{t=1}^T H_t(\omega_1)$. By construction, the estimator satisfies the first order condition

$$\begin{aligned} 0 &= \begin{pmatrix} \overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}\overline{h}(\widehat{\omega}_1) \\ T^{-1}\sum_{T=1}^T x_t(y_t - x_t'\widehat{\omega}_2) \\ \widehat{\omega}_3 - T^{-1}\sum_{t=1}^T (y_t - \widehat{y}_t)^2 \end{pmatrix} \\ &= \begin{pmatrix} \overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}\overline{h}(\omega_{10}) + \overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}\overline{H}(\widetilde{\omega}_1)(\widehat{\omega}_1 - \omega_{10}) \\ T^{-1}\sum_{t=1}^T x_t(y_t - x_t'\omega_{20}) - T^{-1}\sum_{t=1}^T x_t x_t' (\widehat{\omega}_2 - \omega_{20}) \\ (\widehat{\omega}_3 - \omega_3) + \omega_3 - T^{-1}\sum_{t=1}^T (y_t - x_t\widehat{\omega}_2)^2 \end{pmatrix}, \end{aligned} \tag{1.14}$$

where the second equality follows from a mean value expansion of $\overline{h}(\widehat{\omega}_1)$ around ω_{10}, with $\widetilde{\omega}_1$ between ω_{10} and $\widehat{\omega}_1$. Let

$$\widetilde{F} = diag\{[\overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}\overline{H}(\widetilde{\omega}_1)]^{-1}\overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}, [T^{-1}\sum_{t=1}^T x_t x_t']^{-1}, 1\}. \tag{1.15}$$

Then (1.14) implies that

$$\begin{aligned} T^{1/2}(\widehat{\omega} - \omega) &= \widetilde{F} \cdot T^{-1/2}\sum_{t=1}^T \begin{pmatrix} -h_t(\omega_{10}) \\ x_t(y_t - x_t'\omega_{20}) \\ (y_t - x_t\widehat{\omega}_2)^2 - \omega_3 \end{pmatrix} \\ &= \widetilde{F} \cdot T^{-1/2}\sum_{t=1}^T \begin{pmatrix} -h_t(\omega_{10}) \\ x_t(y_t - x_t'\omega_{20}) \\ (y_t - x_t'\omega_{20})^2 - \mathbb{E}[(y_t - x_t'\omega_{20})^2] \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \varepsilon_T \end{pmatrix}, \end{aligned} \tag{1.16}$$

where the second equality uses $\omega_3 = \mathbb{E}[(y_t - x_t'\omega_{20})^2]$ by definition and

$$\begin{aligned}
\varepsilon_T &= T^{-1/2}\sum_{t=1}^{T}\left[\left(y_t - x_t'\widehat{\omega}_2\right)^2 - \left(y_t - x_t'\omega_{20}\right)^2\right] \\
&= 2T^{-1}\sum_{t=1}^{T}\left(y_t - x_t'\omega_{20}\right)x_t'\left[T^{1/2}\left(\widehat{\omega}_2 - \omega_{20}\right)\right] + o_p(1) \\
&= o_p(1) \qquad (1.17)
\end{aligned}$$

because $T^{-1}\sum_{t=1}^{T}\left(y_t - x_t'\omega_{20}\right)x_t' \to_p 0$ and $T^{1/2}(\widehat{\omega}_2 - \omega_{20}) = O_p(1)$ following Assumption R. In addition,

$$\widetilde{F} \to_p F \qquad (1.18)$$

following from the consistency of $\widehat{\omega}_1$ and Assumption R. Finally, the desirable result follows from (1.16)-(1.18) and Assumption R. The consistency of $\widehat{\Omega}$ follows from the consistency of $\widehat{F}$ and $\widehat{V}$. □

Proof of Lemma 1.2. Assumptions 1-3 of AM holds under Assumption R, S, and Lemma 1.1. This Lemma then follows from Theorem 1 of AM. □