November 21, 2018

1 Link Functions

So far, we have introduced the following parameters: (m_0, m_1, θ, π) in SDF, (ρ, c, δ) in the volatility dynamic, and $(\psi, \beta, \gamma, \phi)$ in the return dynamic. Next, we explore restrictions among these parameters that are consistent with this model. In other words, not all of these parameters can change freely under the structural model.

We use these restrictions to construct link functions between a set of reduced-form parameter and a set of structural parameters. These link functions play an important role on separating the regularly behaved reduced-form parameters from the structural parameters. They also are used to conduct identification robust inference for the structural parameters based on a minimum distance criterion.

All of these restrictions are also presented and imposed in the GMM estimation in Han et al. (2018). However, they estimate all parameters together with all of these restriction imposed, because they assume all parameters being estimated are strongly identified.

1.1 Pricing Equation Restrictions

We first explore the restriction implied by the pricing equation $\mathbb{E}[M_{t,t+1}\exp(r_{t+1})|\mathcal{F}_t] = 1$. We first provide a simple result stating that the constants m_0 and m_1 are normalization constants implied by all the other parameters. Thus, m_0 and m_1 are not free parameters to be estimated. Instead, they should take the value given, once other parameters are specified. These restrictions on m_0 and m_1 are obtained by applying the restriction $\mathbb{E}[M_{t,t+1}\exp(r_{t+1})|\mathcal{F}_t] = 1$ to the risk free asset. Applying the same argument to any other asset, we also obtain another set of two restrictions, which can be written in terms of coefficient β and γ under the linear form of $D(x)$ and $E(x)$.

Lemma 1.1 *Given the parameterization in the model. The pricing equation* $\mathbb{E}[M_{t,t+1}\exp(r_{t+1})|\mathcal{F}_t] = 1$ *implies that*

$$
\begin{aligned}
m_0 &= E(\theta) + B\left(\pi + C\left(\theta\right)\right), \\
m_1 &= D\left(\theta\right) + A\left(\pi + C\left(\theta\right)\right),
\end{aligned}
$$

and

$$\begin{aligned}\gamma &= B\left(\pi + C\left(\theta - 1\right)\right) - B\left(\pi + C\left(\theta\right)\right),\\ \beta &= A\left(\pi + C\left(\theta - 1\right)\right) - A\left(\pi + C\left(\theta\right)\right).\end{aligned}$$

The two equalities on β and γ link them to the market risk price θ and volatility risk π through the functions $A(\cdot), B(\cdot), C(\cdot)$, which also involve parameters $(\rho, c, \delta, \psi, \phi)$. We treat these two equalities as link functions in the minimum distance criterion specified below.

1.2 Leverage Effect Restrictions

We first show that both parameter ψ and ϕ are linked to the leverage effect. Given the variance of r_{t+1} conditional on $(\sigma_{t+1}^2, \sigma_t^2)$, specified in (**), we have

$$\phi^2 = \sigma_{t+1}^2 - Var[r_{t+1}|\sigma_{t+1}^2, \sigma_t^2].$$

This shows that ϕ is linked to the leverage effect because it measures the return volatility reduction after conditioning on the volatility path. On the other hand, given the mean of r_{t+1} conditional on $(\sigma_{t+1}^2, \sigma_t^2)$, specified in (**), we have[1]

$$E[r_{t+1}|\sigma_{t+1}^2, \sigma_t^2] - E[r_{t+1}|\sigma_t^2] = \psi\left\{\sigma_{t+1}^2 - E[\sigma_{t+1}^2|\sigma_t^2]\right\}. \tag{1.1}$$

The parameter ψ is also linked to the leverage effect because it incorporates the instantaneous relationship between change in the innovation in σ_{t+1}^2 and change in the return forecast. Besides the leverage effect, ψ also include the change in r_{t+1} through two other channels: (i) the market risk price due to the correlation between $M_{t,t+1}$ and $\exp(r_{t+1})$ and (ii) the Jensen's effect term that captures the change in the mean of $\exp(r_{t+1})$ with the volatility of r_{t+1} (with a 1/2 factor). Following Han et al (2018), these two measures of the leverage effect are restricted to

$$\psi - (1-\phi^2)\theta + \frac{1}{2}(1-\phi^2) = k\phi \tag{1.2}$$

for a constant k, where the left hand side is the leverage effect in ψ after the other two effects are removed. Han et al (2018) show that an appropriate choice of k is the value under which $corr[r_{t+1}, \sigma_{t+1}^2|\sigma_t^2] = \phi$ if this correlation is indeed time invariant. Guided by this condition, they show that $k = 1/(2c)^{1/2}$ should be used for the volatility dynamic specified in () and ().

[1] To see this result, note that he mean of $r_{t+1} - \psi\sigma_{t+1}^2$ given $(\sigma_{t+1}^2, \sigma_t^2)$ does not depend on σ_{t+1}^2.

1.3 Structural and Reduced-Form Parameters

Because ϕ is the leverage effect parameter, we group it together with market risk price θ and the volatility risk price π and call $\lambda = (\theta, \pi, \phi)'$ structural parameters. These structural parameters are estimated by restrictions from this structural model. In contrast, the other parameters in the conditional mean and variance of the return and volatility, see ()-() and ()-(), are simply estimated using these moments, without any model restriction. As such, we call them the reduced-form parameter. Because $1 - \phi^2$ shows up in the conditional variance of r_{t+1}, see (**), we define $\zeta = 1 - \phi^2$ as a reduced-form parameter and link it to the structural parameter ϕ through this relationship. To sum up, the reduced-form parameters are $\omega = (\rho, c, \delta, \psi, \beta, \gamma, \zeta)'$.

Using ζ as a reduced-form parameter has the additional benefit that the sample variance, denoted by $\widehat{\zeta}$, is a simple consistent estimator with normal distribution. Estimation of ϕ is more complicated because $-1 < \phi \leq 0$ by definition and this estimation involves boundary constraints that results in a non-standard distribution. The inference procedure below does not require estimation of ϕ and is uniform over ϕ even if its true value is on or close to the boundary 0.

The link functions between the structural parameter λ and the reduced-form parameter ω are collected together in

$$g(\lambda, \omega) = \begin{pmatrix} B\left(\pi + C\left(\theta - 1\right)\right) - B\left(\pi + C\left(\theta\right)\right) - \gamma \\ A\left(\pi + C\left(\theta - 1\right)\right) - A\left(\pi + C\left(\theta\right)\right) - \beta \\ \psi - (1 - \phi^2)\theta + \frac{1}{2}(1 - \phi^2) - 1/(2c)^{1/2}\phi \\ \zeta - \left(1 - \phi^2\right) \end{pmatrix}. \tag{1.3}$$

For the inference problem studied below, we know $g(\lambda_0, \omega_0) = 0$ when evaluated at the true value of λ and ω.

1.4 Identification

One of the important contributions of Han etal is to establish the relationship between the identification of the volatility risk price and the leverage effect. In particular, they show that when the leverage effect parameter $\phi = 0$, the volatility price π is not identified. To see this result, note that the only source of identification information on π are the first two link functions in $g(\lambda_0, \omega_0) = 0$, which comes from Lemma 1.1. In these two equalities, π is not identified if $C(\theta) = C(\theta - 1)$. Using the definition $C(\theta) = \psi\theta + (1 - \phi^2)\theta^2/2$ and the restriction (1.2), we have

$$C(\theta) - C(\theta - 1) = \psi + (1 - \phi^2)\left(\theta - \frac{1}{2}\right) = k\phi.$$

Therefore, a non-zero leverage effect i.e., $\phi \neq 0$, is required for the identification of the volatility risk price π.

The identification result above is based on the theoretical model only, without considering data uncertainty in practical applications. With a finite-sample size and different types of noise in the data, such as measurement errors and omitted variables, a much more substantial leverage effect is required to obtain a standard identification situation where the noise in the data is negligible compared to the information to identify π. However, if only a small leverage effect is documented, e.g., Ait-Sahalia et al (2014), or the magnitude of the leverage effect is completely unknown, an identification robust procedure is needed to conduct inference in this problem. We provide such a procedure now.

2 Robust Inference for Risk Prices

2.1 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^2_{t+1}, 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^2_{t+1}, which can be equivalently written as

$$\begin{aligned} E[\sigma^2_{t+1}|\sigma^2_t] &= A \text{ and } E[\sigma^4_{t+1}|\sigma^2_t] = B, \text{ where} \\ A &= \rho\sigma^2_t + c\delta \text{ and } B = A^2 + \left(2c\rho\sigma^2_t + c^2\delta\right). \end{aligned} \tag{2.1}$$

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

$$E[h_t(\omega_{10})] = 0, \text{ where } h_t(\omega_1) = [(1, \sigma^2_t) \otimes (\sigma^2_{t+1} - A), (1, \sigma^2_t, \sigma^4_t) \otimes (\sigma^4_{t+1} - B)]', \tag{2.2}$$

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

$$\widehat{\omega}_1 = \arg\min_{\omega_1 \in O_1} \left(T^{-1} \sum_{t=1}^{T} h_t(\omega_1) \right)' \widehat{V}_1 \left(T^{-1} \sum_{t=1}^{T} h_t(\omega_1) \right), \tag{2.3}$$

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

$$\begin{aligned} \widehat{\omega}_2 &= \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1} \sum_{t=1}^{T} x_t y_t, \text{ where} \\ x_t &= \sigma_{t+1}^{-1}(1, \sigma_t^2, \sigma_{t+1}^2)' \text{ and } y_t = \sigma_{t+1}^{-1} r_{t+1}. \end{aligned} \tag{2.4}$$

We estimate ω_3 by the sample variance estimator

$$\widehat{\omega}_3 = T^{-1} \sum_{t=1}^{T} (y_t - \widehat{y}_t)^2, \text{ where } \widehat{y}_t = x_t' \widehat{\omega}_2. \tag{2.5}$$

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

$$f_t(\omega) = \begin{pmatrix} h_t(\omega_1) \\ x_t(y_t - x_t'\omega_2) \\ (y_t - x_t'\omega_2)^2 \end{pmatrix} \in R^{d_f} \text{ and } V = \sum_{m=-\infty}^{\infty} \mathbb{C}\mathbf{ov}[f_t(\omega_0), f_{t+m}(\omega_0)]. \tag{2.6}$$

The estimator $\widehat{\omega}$ defined above is based on the first moment of $f_t(\omega)$. Thus, the limiting distribution of $\widehat{\omega}$ relates to the limiting distribution of $T^{-1/2} \sum_{t=1}^{T} f_t(\omega_0) - \mathbb{E}[f_t(\omega_0)$ following from the central limit theorem. Furthermore, because ω_1 is the GMM estimator based on some nonlinear moment conditions, we need uniform convergence of the the sample moments and their derivatives to show the consistency and asymptotic normality of $\widehat{\omega}_1$. These uniform convergence follows from the uniform law of large numbers. Because $\widehat{\omega}_2$ is a simple OLS estimator by regressing y_t and x_t, we need no-multicollinearity among the regressors. We make all of the these assumptions below.. All of them are easily verifiable with weakly dependent time series data.

Let $\widehat{V}$ denote a heteroskedasticity and autocorrelation consistent (HAC) estimator of V. The estimator $\widehat{V}_1$ is a submatrix of $\widehat{V}$ associate with V_1. Let $H_t(\omega_1) = \partial h_t(\omega_1)/\partial \omega_1'$.

Assumption R. The following conditions hold uniformly over $P \in \mathcal{P}$, for some fixed $0 < C < \infty$.
(i) $T^{-1}\sum_{t=1}^{T}(h_t(\omega_1) - \mathbb{E}[h_t(\omega_1)) \rightarrow_p 0$ 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.
(ii) $T^{-1}\sum_{t=1}^{T}(x_t x_t' - \mathbb{E}[x_t x_t']) \rightarrow_p 0$.
(iii) $V^{-1/2}\{T^{-1/2}(\sum_{t=1}^{T} f_t(\omega_0) - \mathbb{E}[f_t(\omega_0)]\} \rightarrow_d N(0, I)$ and $\widehat{V} - V \rightarrow_p 0$.
(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 $z_t = (1, \sigma_t^2, \sigma_t^4)'$.

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

$$\begin{aligned} \mathcal{B} &= diag\{[H(\omega_{10})V_1^{-1}H(\omega_{10})]^{-1}H(\omega_{10})V_1^{-1}, \mathbb{E}[x_t x_t']^{-1}, 1\}, \\ \widehat{\mathcal{B}} &= diag\{[\overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}\overline{H}(\widehat{\omega}_1)]^{-1}\overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}, [T^{-1}\sum_{t=1}^{T} x_t x_t']^{-1}, 1\}. \end{aligned} \tag{2.7}$$

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

2.2 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

$$\eta_T(\lambda) := T^{1/2}\left[\widehat{g}(\lambda) - g_0(\lambda)\right] = G_0(\lambda)\Omega^{1/2} \cdot \xi_T + o_p(1), \tag{2.8}$$

where $\xi_T \to_d N(0, I)$ following Lemma 2.1. Thus, $\eta_T(\cdot)$ 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 (2.8), we can write $T^{1/2}\widehat{g}(\lambda) = \eta_T(\lambda) + T^{1/2}g_0(\lambda)$, where $\eta_T(\lambda)$ 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 dimensional nuisance 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.

2.3 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

$$QLR(\lambda_0) = T\widehat{g}(\lambda_0)'\widehat{\Sigma}(\lambda_0, \lambda_0)^{-1}\widehat{g}(\lambda_0) - \min_{\lambda \in \Lambda} T\widehat{g}(\lambda)'\widehat{\Sigma}(\lambda, \lambda)^{-1}\widehat{g}(\lambda), \tag{2.9}$$

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

$$\widehat{r}(\lambda) = \widehat{g}(\lambda) - \widehat{\Sigma}(\lambda, \lambda_0)\widehat{\Sigma}(\lambda_0, \lambda_0)^{-1}\widehat{g}(\lambda_0). \tag{2.10}$$

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

$$g_b^*(\lambda) = \widehat{r}(\lambda) + \widehat{\Sigma}(\lambda, \lambda_0)\widehat{\Sigma}(\lambda_0, \lambda_0)^{-1}\eta_b^* \tag{2.11}$$

and a simulated statistic

$$QLR_b^*(\lambda_0) = T\widehat{g}(\lambda_0)'\widehat{\Sigma}(\lambda_0, \lambda_0)^{-1}\widehat{g}(\lambda_0) - \min_{\lambda\in\Pi} Tg_b^*(\lambda)'\widehat{\Sigma}(\lambda, \lambda)^{-1}g_b^*(\lambda). \tag{2.12}$$

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 (2.3) and (2.4). Obtain a consistent estimator of its asymptotic covariance $\widehat{\Omega} = \widehat{\mathcal{B}}\widehat{V}\widehat{\mathcal{B}}'$, where $\widehat{\mathcal{B}}$ is define in (2.7) and $\widehat{V}$ is a HAC estimator of V.

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

(ii) Construct the QLR statistic $QLR(\lambda_0)$ in (2.9) using $g(\lambda,\omega)$, $G(\lambda,\omega)$, $\widehat{\omega}$, and $\widehat{\Omega}$.

(iii) Compute the residual process $\widehat{r}(\lambda)$ in (2.10).

(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

$$CS_T = \{\lambda_0 : QLR_T(\lambda_0) \leq c_{1-\alpha}(r, \lambda_0)\}. \tag{2.13}$$

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, for some fixed $0 < C < \infty$.

(i) $g(\lambda,\omega)$ is partially differentiable in ω, with partial derivative $G(\lambda,\omega)$ that satisfies $||G(\lambda_1,\omega) - G(\lambda_2,\omega)|| \leq C||\lambda_1 - \lambda_2||$ and $||G(\lambda,\omega_1) - G(\lambda,\omega_2)|| \leq C||\omega_1 - \omega_2||$.

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

Lemma 2.2 *Suppose Assumption R and S hold. Then,*

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

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. *QED*

Proof of Lemma 1.1. For the risk free asset, $r_{t+1} = 0$. Therefore, we have

$$\begin{aligned}
1 &= E\left[\exp\left(m_0 + m_1\sigma_t^2 - \pi\sigma_{t+1}^2 - \theta r_{t+1}\right)|\mathcal{F}_t\right] \\
&= \exp(m_0 + m_1\sigma_t)E\left[\exp\left(-\pi\sigma_{t+1}^2\right)E\left[\exp\left(-\theta r_{t+1}\right)|\mathcal{F}_t, \sigma_{t+1}^2\right]|\mathcal{F}_t\right] \\
&= \exp(m_0 - E\left(\theta\right) + m_1\sigma_t - D\left(\theta\right)\sigma_t^2)E\left[\exp\left(-\pi\sigma_{t+1}^2 - C\left(\theta\right)\sigma_{t+1}^2\right)|\mathcal{F}_t\right] \\
&= \exp(m_0 - E\left(\theta\right) + m_1\sigma_t - D\left(\theta\right)\sigma_t^2 - A\left(\pi + C\left(\theta\right)\right)\sigma_t^2 - B\left(\pi + C\left(\theta\right)\right)),
\end{aligned}$$

where the first equality follows from the pricing equation, the second equality follows from the law of iterated expectation, the third equation uses the Laplace transform for r_{t+1} in (**), and the last equality follows from the Laplace transform for σ_{t+1}^2 in (**). For the final line to be always equal to 1, we restrict the constant term and the coefficient for σ_t^2 to be 0 and that gives the claimed result for m_0 and m_1.

Apply the same argument above to any asset r_{t+1}, we have the same result, except θ is replaced by $\theta - 1$ throughout. This implies that the two equalities for m_0 and m_1 also hold with θ replaced by $\theta - 1$. Therefore,

$$\begin{aligned}
E(\theta-1) + B\left(C\left(\theta-1\right)+\pi\right) &= E(\theta) + B\left(C\left(\theta\right)+\pi\right), \\
D\left(\theta-1\right) + A\left(C\left(\theta-1\right)+\pi\right) &= D\left(\theta\right) + A\left(C\left(\theta\right)+\pi\right).
\end{aligned}$$

The claimed results for γ and β follow from $\gamma = E(\theta) - E(\theta-1)$ and $\beta = D(\theta) - D(\theta-1)$ under the linear specification of $E(x) = \gamma x$ and $D(x) = \beta x$. *QED*.

Proof of Lemma 2.1. 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(i) and 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

$$\begin{aligned} 0 &= \begin{pmatrix} \overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}\overline{h}(\widehat{\omega}_1) \\ T^{-1}\sum_{T=1}^{T} x_t(y_t - x_t'\widehat{\omega}_2) \\ \widehat{\omega}_3 - T^{-1}\sum_{t=1}^{T}(y_t - \widehat{y}_t)^2 \end{pmatrix} \\ &= \begin{pmatrix} \overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}\overline{h}(\omega_{10}) + \overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}\overline{H}(\widetilde{\omega}_1)(\widehat{\omega}_1 - \omega_{10}) \\ T^{-1}\sum_{t=1}^{T} x_t(y_t - x_t'\omega_{20}) - T^{-1}\sum_{t=1}^{T} x_t x_t'(\widehat{\omega}_2 - \omega_{20}) \\ (\widehat{\omega}_3 - \omega_3) + \omega_3 - T^{-1}\sum_{t=1}^{T}(y_t - x_t\widehat{\omega}_2)^2 \end{pmatrix}, \end{aligned} \tag{2.14}$$

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

$$\widetilde{\mathcal{B}} = diag\{[\overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}\overline{H}(\widetilde{\omega}_1)]^{-1}\overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}, [T^{-1}\sum_{t=1}^{T} x_t x_t']^{-1}, 1\}. \tag{2.15}$$

Then (2.14) implies that

$$\begin{aligned} T^{1/2}(\widehat{\omega} - \omega) &= \widetilde{\mathcal{B}} \cdot T^{-1/2}\sum_{t=1}^{T} \begin{pmatrix} -h_t(\omega_{10}) \\ x_t(y_t - x_t'\omega_{20}) \\ (y_t - x_t\widehat{\omega}_2)^2 - \omega_3 \end{pmatrix} \\ &= \widetilde{\mathcal{B}} \cdot T^{-1/2}\sum_{t=1}^{T} \begin{pmatrix} -h_t(\omega_{10}) \\ x_t(y_t - x_t'\omega_{20}) \\ (y_t - x_t'\omega_{20})^2 - \mathbb{E}[(y_t - x_t'\omega_{20})^2] \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \varepsilon_T \end{pmatrix}, \end{aligned} \tag{2.16}$$

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

$$\begin{aligned} \varepsilon_T &= T^{-1/2}\sum_{t=1}^{T}\left[\left(y_t - x_t'\widehat{\omega}_2\right)^2 - \left(y_t - x_t'\omega_{20}\right)^2\right] \\ &= 2T^{-1}\sum_{t=1}^{T}\left(y_t - x_t'\omega_{20}\right)x_t'\left[T^{1/2}\left(\widehat{\omega}_2 - \omega_{20}\right)\right] + o_p(1) \\ &= o_p(1) \end{aligned} \tag{2.17}$$

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

$$\widetilde{\mathcal{B}} \to_p \mathcal{B} \tag{2.18}$$

following from the consistency of $\widehat{\omega}_1$ and Assumption R. Finally, the desirable result follows from (2.16)-(2.18) and Assumption R. The consistency of $\widehat{\Omega}$ follows from the consistency of $\widehat{\mathcal{B}}$ and $\widehat{V}$. $\square$

Proof of Lemma 2.2. We obtain this result by applying Theorem 1 of AM. We now verify Assumptions 1-3 in AM. To show weak convergence $\eta_T(\cdot)$ to $\eta(\cdot)$ uniformly over $\mathcal{P}$, note that by a second-order Taylor expansion,

$$\begin{aligned} \eta_T(\lambda) \quad &: \quad = T^{1/2}\left[\widehat{g}(\lambda) - g_0(\lambda)\right] = G_0(\lambda)\Omega^{1/2}\xi_T + \delta_T, \text{ where} \\ \xi_T \quad &= \quad \Omega^{-1/2}T^{1/2}\left(\widehat{\omega} - \omega_0\right), \ \delta_T = \left(G(\lambda, \widetilde{\omega}) - G(\lambda, \omega_0)\right) T^{1/2}(\widehat{\omega} - \omega_0) \end{aligned} \tag{2.19}$$

and $\widetilde{\omega}$ is between $\widehat{\omega}$ and ω_0. Because $||G(\lambda, \widetilde{\omega}) - G(\lambda, \omega_0)|| \leq C||\widetilde{\omega} - \omega_0||$, $\delta_T = o_p(1)$ uniformly over $\mathcal{P}$ following Lemma 2.1. To show $G_0(\lambda)\Omega^{1/2}\xi_T$ weakly converges to $\eta(\cdot)$, it is sufficient to show (i) the pointwise convergence

$$\begin{pmatrix} G_0(\lambda_1)\Omega^{1/2}\xi_T \\ G_0(\lambda_2)\Omega^{1/2}\xi_T \end{pmatrix} \to_d \begin{pmatrix} \eta(\lambda_1) \\ \eta(\lambda_2) \end{pmatrix}, \tag{2.20}$$

which follows from Lemma 2.1, and (ii) the stochastic equicontinuity condition, i.e., for every $\varepsilon > 0$ and $\xi > 0$, there exists a $\delta > 0$ such that

$$\limsup_{T\to\infty} \Pr\left(\sup_{P\in\mathcal{P}} \sup_{||\lambda_1-\lambda_2||\leq\delta} \left\|G_0(\lambda_1)\Omega^{1/2}\xi_T - G_0(\lambda_2)\Omega^{1/2}\xi_T\right\| > \varepsilon\right) < \xi. \tag{2.21}$$

For some $C < \infty$, we have $||G_0(\lambda_1) - G(\lambda_2)|| \leq C||\lambda_1 - \lambda_2||$ under a uniform bound for the derivative under Assumption S and we have $||\Omega^{1/2}|| \leq C$ under Assumption R because F and V both have bounded largest eigenvalue. Thus,

$$\begin{aligned} &\limsup_{T\to\infty} \Pr\left(\sup_{P\in\mathcal{P}} \sup_{||\lambda_1-\lambda_2||\leq\delta} \left\|G_0(\lambda_1)\Omega^{1/2}\xi_T - G_0(\lambda_2)\Omega^{1/2}\xi_T\right\| > \varepsilon\right) \\ \leq\; &\limsup_{T\to\infty} \Pr\left(C^2 \sup_{P\in\mathcal{P}} \|\xi_T\| > \frac{\varepsilon}{\delta}\right). \end{aligned} \tag{2.22}$$

Because $\xi_T = O_p(1)$ uniformly over $P \in \mathcal{P}$, there exists δ such that ε/δ is large enough to make the right hand side of the inequality in (2.22) smaller than ξ.

Assumption 2 and 3 of AM follows from Assumption R. QED