Identification Robust Inference for Risk Prices in Structural Stochastic Volatility Models

Xu Cheng* Eric Renault† Paul Sangrey‡

Current Version
This Version: November 26, 2018

Abstract

In structural stochastic volatility asset pricing models, changes in volatility affect risk premia through two channels: (1) the investor's willingness to bear high volatility in order to get high expected returns as measured by the market risk price, and (2) the investor's direct aversion to changes in future volatility as measured by the volatility risk price. Disentangling these channels is difficult and poses a subtle identification problem that invalidates standard inference. We adopt the discrete-time exponentially affine model of Han, Khrapov, and Renault (2018), which links the identification of the volatility risk price to the leverage effect. In particular, we develop a minimum distance criterion that links the market risk price, the volatility risk price, and the leverage effect to the well-behaved reduced-form parameters governing the return and volatility's joint distribution. The link functions are almost flat if the leverage effect is close to zero, making estimating the volatility risk price difficult. We apply the conditional quasi-likelihood ratio test Andrews and Mikusheva (2016) develop in a nonlinear GMM framework to a minimum distance framework. The resulting conditional quasi-likelihood ratio test is uniformly valid. We invert this test to derive robust confidence sets that provide correct coverage for the risk prices regardless of the leverage effect's magnitude.

JEL Codes: C12, C14, C38, C58, G12

Keywords: weak identification, robust inference, stochastic volatility, leverage, market risk premium, volatility risk premium, risk price, confidence set, asymptotic size

*University of Pennsylvania, The Perelman Center for Political Science and Economics, 133 South 36th Street, Philadelphia, PA 19104, xucheng@upenn.edu

†Brown University, Department of Economics – Box B, 64 Waterman Street, Providence, RI 02912, eric_renault@brown.edu

‡University of Pennsylvania, The Perelman Center for Political Science and Economics, 133 South 36th Street, Philadelphia, PA 19104, paul@sangrey.io

1 Introduction

A fundamental question in finance is how investors optimally trade off risk and return. Economic theories predict investors demand a higher return as compensation for bearing more risk. Hence, we should expect a positive relationship between the mean and volatility of returns. Some seminal early papers proposed a static trade-off between risk and expected return, most notably the capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965). In practice, volatility varies over time. Consequently, a significant strand of the recent literature examines the dynamic tradeoff between volatility and returns, including structural stochastic volatility models (Christoffersen, Heston, and Jacobs 2013; Bansal et al. 2014; Dew-Becker et al. 2017). In nonlinear models such as these, investors care not just about how an asset's returns co-move with the volatility but also care how they co-move with changes in volatility.

In structural stochastic volatility models, changes in volatility affect risk premia through two channels: (1) the investor's willingness to tolerate high volatility in order to get high expected returns as measured by the market risk price, and (2) the investor's direct aversion to changes in future volatility as measured by the volatility risk price. We adopt the discrete-time exponentially affine model of Han, Khrapov, and Renault (2018), who represent the market risk price and the volatility risk price by two structural parameters. In this model, Han, Khrapov, and Renault (2018) establish the significant result that the identification of the volatility risk price depends on a substantial leverage effect, which is the negative contemporaneous correlation between returns and volatility.

Although the leverage effect is theoretically less than zero, it is difficult to quantify empirically, and its estimate usually is small (Aït-Sahalia, Fan, and Li 2013). When the leverage effect is small, the data only provide a limited amount of information about the volatility risk, compared to the finite-sample noise in the data. This low signal-to-noise ratio, as modeled by weak identification, invalidates standard inference based on the generalized method of moments (GMM) estimator, see Stock and Wright (2000) and Andrews and Cheng (2012).

We provide an identification-robust confidence set for the structural parameters that measure the market risk price, the volatility risk price, and the leverage effect. The robust confidence set provides correct asymptotic coverage, uniformly over a large set of models that allow for any magnitude of the leverage effect. This uniform validity is crucial for the confidence set to have good finite-sample coverage (Mikusheva 2007; Andrews and Guggenberger 2010). In contrast, standard confidence sets based on the GMM estimator and its asymptotic normality do not have uniform validity in the presence of a small leverage effect. This issue affects all of the structural parameters because they are estimated simultaneously.

We achieve robust inference in two steps. First, we establish a minimum distance criterion using link functions between the structural parameters and a set of reduced-form parameters that determine the joint distribution of the return and volatility. The structural model implies that the link functions are zero when evaluated at the true values of the structural parameters and the reduced-form parameters. Identification and estimation of these reduced form parameters are standard and are not affected by the presence of a small leverage effect. However, the link functions are almost flat in one of the structural parameters when the leverage effect is small, resulting in weak identification. Second, given this minimum distance criterion, we invert the conditional quasi-likelihood ratio (QLR) test by Andrews and Mikusheva (2016) to construct a robust confidence set. The key feature of this test is that it treats the flat link functions as an infinite dimensional nuisance parameter. The critical value is constructed by conditioning on a sufficient statistic for this nuisance parameter, and it is shown to yield a valid test regardless of the nuisance parameter. Andrews and Mikusheva (2016) develop this test in a GMM framework. We show it works in the minimum distance context here and provide conditions for its asymptotic validity. For practitioners, we provide a detailed algorithm for the construction of this simulation-based robust confidence set.

Our paper relates to the empirical analysis of the effect of volatility on risk premia. As Lettau and Ludvigson (2010) mention, the evidence here is inconclusive. Bollerslev, Engle, and Wooldridge (1988), Harvey (1989), Ghysels, Santa-Clara, and Valkanov (2005), Bali and Peng (2006), and Ludvigson and Ng (2007) find a positive relationship, while Campbell (1987), Breen, Glosten, and Jagannathan (1989), Pagan and Hong (1991), Whitelaw (1994), and Brandt and Kang (2004) find a negative relationship. Also, some papers use both a market risk factor and a variance risk factor to explain the risk premia dynamics, including Christoffersen, Heston, and Jacobs (2013), Feunou et al. (2014), and Dew-Becker et al. (2017). In related strand of the literature, Bollerslev, Law, and Tauchen (2008) and Drechsler and Yaron (2011) document a substantial positive variance risk premium. We contribute to this literature by providing the first method for making valid inference for the market risk price and the volatility risk price. This new confidence set not only allows for both effects but also takes into account the potential identification issue.

The weak identification issue studied in this paper is relevant in many economic applications, ranging from linear instrumental variable models (Staiger and Stock 1997) to nonlinear structural models (Mavroeidis, Plagborg-Møller, and Stock 2014; Andrews and Mikusheva 2015). This paper is the first one to study this issue in structural asset pricing models with stochastic volatility. Moreira (2003) introduces the conditional inference approach to the linear instrumental variable model, and Kleibergen (2005) applies it to the nonlinear GMM problem. Andrews and Mikusheva (2016) propose conditional inference for nonlinear GMM problems with an infinite-dimensional nuisance parameter. This paper applies it to a minimum distance criterion and extends the scope of its application to a new type of asset pricing model.

The rest of the paper is organized as follows. Section 2 provides the model and its parameterization. Section 3 provides model-implied restrictions and use them to derive the link function. Section 4 provides the asymptotic distribution of the reduced-form parameter and robust inference for the structural parameter. A detailed algorithm to construct the robust confidence set is given in Section 4.3. Proofs are given in the appendix.

2 The Model

This section provides a parametric structural model with stochastic volatility, following Han, Khrapov, and Renault (2018). They extend the discrete-time exponentially-affine model of Darolles, Gouriéroux, and Jasiak (2006), and their model is a natural discrete-time analog of the Heston (1993) model. We specify this model using a stochastic discount factor (SDF), also called the pricing kernel, and the physical measure, which gives the joint distribution of the return and volatility dynamics.[1] We first define the SDF and parameterize it as an exponential affine function with unknown parameters. We then provide parametric distribution for the physical measure.

Let P_t be the price of the asset under consideration. Let $r_{t+1} = \log(P_{t+1}/P_t) - r_f$ denote the log excess return minus the risk-free rate and σ^2_{t+1} denote its volatility. The observed data is $W_t = (r_t, \sigma^2_t)$ for $t = 1, \ldots, T$. Let $\mathcal{F}_t$ be the representative investor's information set at time t .

2.1 Stochastic Discount Factor and Its Parameterization

The prices of all assets satisfy the following asset pricing equation in terms of the SDF:

$$P_t = \mathbb{E}\left[M_{t,t+1} \exp\left(-r_f\right) P_{t+1} \,|\, \mathcal{F}_t\right]. \tag{1}$$

Following the definition of r_{t+1}, the pricing equation implies that for all assets

$$1 = \mathbb{E}\left[M_{t,t+1} \exp\left(r_{t+1}\right) | \, \mathcal{F}_t\right]. \tag{2}$$

We start by parameterizing the SDF by the exponential affine model. Let π be the price of volatility risk and κ be the price of market risk. They are both considered as structural parameters.

Definition 1. Parameterize the Stochastic Discount Factor

$$M_{t,t+1}(\pi, \kappa) = \exp\left(m_0 + m_1 \sigma^2_t - \pi \sigma^2_{t+1} - \kappa r_{t+1}\right). \tag{3}$$

Throughout we assume that the two risks that command nonzero prices are the market risk price and the volatility risk price. Consequently, we only use variation in the first two moments of the data to estimate these parameters.

2.2 Parameterizing the Volatility and Return Dynamics

Next, we parameterize the joint distribution of $\{W_t : t = 1, \ldots, T\}$. Following Han, Khrapov, and Renault (2018), we make the following assumptions. First, the return r_t and volatility σ^2_t are first-order Markov. Second, there is no Granger-causality from the return to the volatility. Third, returns are independent across time given the volatility. We do allow σ^2_t and r_t to be contemporaneously correlated, as they are in the data.

1. The risk-neutral measure is unobserved due to the lack of option data.

Under these assumptions, the volatility drives all of the dynamics of the process. The only relevant information in the information set $\mathcal{F}_t$ for time $t+1$-measurable variables is contained in σ_t^2. In general, σ_t^2, σ_{t+1}^2, and r_{t+1} form a sufficient statistic for $\mathcal{F}_{t+1}$.

We adopt the conditional autoregressive gamma process as in Gouriéroux and Jasiak (2006) and Han, Khrapov, and Renault (2018) for the volatility process. The model is parameterized in terms of the Laplace transform:

$$\mathbb{E}\left[\exp(-x\sigma_{t+1}^2) \,\middle|\, \mathcal{F}_t\right] = \exp\left(-A(x)\sigma_t^2 - B(x)\right) \tag{4}$$

for all $x \in \mathbb{R}$. The function $A(x)$ and $B(x)$ are parameterized as follows.

Definition 2. Parameterize the Volatility Dynamics

$$A(x) \coloneqq \frac{\rho x}{1 + cx}, \tag{5}$$

$$B(x) \coloneqq \delta \log(1 + cx), \tag{6}$$

with $\rho \in [0, 1-\epsilon]$, $c > \epsilon$, $\delta > \epsilon$ for some $\epsilon > 0$.

In this specification, ρ is a persistence parameter, c is a scale parameter, and δ is a level parameter. We can see this clearly in the following conditional mean and variance formulas for σ_{t+1}^2.

Remark 1 (Volatilty Moment Conditions)**.**

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

$$\mathbb{V}\mathrm{ar}\left[\sigma_{t+1}^2 \,\middle|\, \sigma_t^2\right] = 2c\rho\sigma_t^2 + c^2\delta. \tag{8}$$

Next, we model the return dynamics. Similar to the volatility dynamics, the distribution of r_t given σ_{t+1}^2 and σ_t^2 is specified in terms of the Laplace transform:

$$\mathbb{E}\left[\exp(-xr_{t+1}) \,\middle|\, \mathcal{F}_t, \sigma_{t+1}^2\right] = \exp\left(-C(x)\sigma_{t+1}^2 - D(x)\sigma_t^2 - E(x)\right) \tag{9}$$

for all $x \in \mathbb{R}$. The function $C(x)$, $D(x)$, and $E(x)$ are parameterized as follows such that the return has a conditional Gaussian distribution.

Definition 3. Parameterize the Return Dynamics

$$C(x) \coloneqq \psi x - \frac{1-\phi^2}{2}x^2, \tag{10}$$

$$D(x) \coloneqq \beta x, \tag{11}$$

$$E(x) \coloneqq \gamma x, \tag{12}$$

with $\phi \in [-1+\epsilon, 0]$ for some $\epsilon > 0$.

Under this specification, we have the following representation of the conditional mean and variance for r_{t+1}.

Remark 2 (Return Moment Conditions)**.**

$$\mathbb{E}\left[r_{t+1} \,\middle|\, \sigma_t^2, \sigma_{t+1}^2\right] = \psi\sigma_{t+1}^2 + \beta\sigma_t^2 + \gamma, \tag{13}$$

$$\mathbb{V}\mathrm{ar}\left[r_{t+1} \,\middle|\, \sigma_t^2, \sigma_{t+1}^2\right] = (1-\phi^2)\sigma_{t+1}^2. \tag{14}$$

The parameter ϕ represents the leverage effect because it measures the return volatility reduction after conditioning on the volatility path.

3 Link Functions

So far, we have introduced the following parameters: (m_0, m_1, κ, π) in SDF, (ρ, c, δ) for the volatility dynamics, and $(\psi, \beta, \gamma, \phi)$ for the return dynamics. 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 parameters 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 imposed in the GMM estimation in Han, Khrapov, and Renault (2018). However, because the volatility risk price is weakly identified, they calibrate it instead of estimating it. Given this calibrated value, they proceed to estimate all other parameters with GMM.

3.1 Pricing Equation Restrictions

We first explore restrictions implied by the pricing equation $\mathbb{E}[M_{t,t+1}\exp(r_{t+1}) \,|\, \mathcal{F}_t] = 1$. We start with 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 below, 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 the coefficients β and γ under the linear form of $D(x)$ and $E(x)$.

Lemma 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*

$$m_0 = E(\kappa) + B\left(\pi + C\left(\kappa\right)\right),$$
$$m_1 = D\left(\kappa\right) + A\left(\pi + C\left(\kappa\right)\right),$$

and

$$\gamma = B\left(\pi + C\left(\kappa - 1\right)\right) - B\left(\pi + C\left(\kappa\right)\right),$$
$$\beta = A\left(\pi + C\left(\kappa - 1\right)\right) - A\left(\pi + C\left(\kappa\right)\right).$$

The two equalities on β and γ link them to the market risk price, κ, and the volatility risk price, 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.

3.2 Leverage Effect Restrictions

Following Han, Khrapov, and Renault (2018), we parameterize ψ as

$$\psi = \frac{\phi}{\sqrt{2c}} - \frac{1-\phi^2}{2}\sigma_{t+1}^2 + (1-\phi^2)\kappa\sigma_{t+1}^2. \tag{15}$$

The first part $\phi/\sqrt{2c}$ measures the leverage effect arising from the instantaneous correlation between r_{t+1} and σ_{t+1}^2. The second part is the traditional Jensen effect term that arises from taking expectation of a log-Gaussian random variable. The third term arises from risk-aversion, which is why it is proportional to κ.

3.3 Structural and Reduced-Form Parameters

Because ϕ is the leverage effect parameter, we group it together with market risk price, κ, and the volatility risk price, π, and call $\theta := (\kappa, \pi, \phi)'$ 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 Remark 1 and Remark 2, are simply estimated using these moments, without any model restrictions. As such, we call them the reduced-form parameters. Because $1-\phi^2$ shows up in the conditional variance of r_{t+1}, see (14), 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 of avoiding estimating ϕ directly. Estimating ϕ when its true value is close to 0 results in an estimator with a non-standard asymptotic distribution due to the boundary constraint. 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 θ and the reduced-form parameter ω are collected together in

$$g(\theta, \omega) = \begin{pmatrix} \gamma - \left[B\left(\pi + C\left(\kappa - 1\right)\right) - B\left(\pi + C\left(\kappa\right)\right)\right] \\ \beta - \left[A\left(\pi + C\left(\kappa - 1\right)\right) - A\left(\pi + C\left(\kappa\right)\right)\right] \\ \psi - (1-\phi^2)\kappa + \frac{1}{2}(1-\phi^2) - 1/(2c)^{1/2}\phi \\ \zeta - \left(1-\phi^2\right) \end{pmatrix}. \tag{16}$$

For the inference problem studied below, we know $g(\theta_0, \omega_0) = 0$ when evaluated at the true value of θ and ω.

3.4 Identification

One of the important contributions of Han, Khrapov, and Renault (2018) 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 risk 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 $g(\theta_0, \omega_0) = 0$, which come from Lemma 1. Clearly, these two equations are independent of π if $C(\kappa) = C(\kappa - 1)$. Using the definition of $C(\cdot)$ and (15), we have

$$C(\kappa) - C(\kappa - 1) = \psi - (1 - \phi^2)\left(\kappa - \frac{1}{2}\right) = \frac{\phi}{\sqrt{2c}}. \tag{17}$$

Clearly, the strength of identification is governed by the strength of the leverage effect. In other words, we need $\phi \neq 0$ to identify the volatility risk price π.

Even if $\phi \neq 0$, we do not know it. In practice, with a finite-sample size and different types of noise in the data, such as measurement errors and omitted variables, a 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 found, as in Bandi and Renò (2012) and Aït-Sahalia, Fan, and Li (2013), 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.

4 Robust Inference for Risk Prices

4.1 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

$$\begin{aligned} &E[\sigma^2_{t+1} | \sigma^2_t] = A \text{ and } E[\sigma^4_{t+1} | \sigma^2_t] = B, \text{ where} \\ &\qquad A = \rho\sigma^2_t + c\delta \text{ and } B = A^2 + \left(2c\rho\sigma^2_t + c^2\delta\right). \end{aligned} \tag{18}$$

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

$$E[h_t(\omega_{10})] = 0, \text{ where } h_t(\omega_1) = [(1, \sigma_t^2) \otimes (\sigma_{t+1}^2 - A), (1, \sigma_t^2, \sigma_t^4) \otimes (\sigma_{t+1}^4 - B)]', \tag{19}$$

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

$$\widehat{\omega}_1 = \arg\min_{\omega_1 \in O_1} \left(T^{-1} \sum_{t=1}^{T} h_t(\omega_1) \right)' \widehat{V}_1 \left(T^{-1} \sum_{t=1}^{T} h_t(\omega_1) \right), \tag{20}$$

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

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

$$\begin{aligned} \widehat{\omega}_2 &= \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1} \sum_{t=1}^{T} x_t y_t, \text{ where} \\ x_t &= \sigma_{t+1}^{-1} (1, \sigma_t^2, \sigma_{t+1}^2)' \text{ and } y_t = \sigma_{t+1}^{-1} r_{t+1}. \end{aligned} \tag{21}$$

We estimate ω_3 by the sample variance estimator

$$\widehat{\omega}_3 = T^{-1} \sum_{t=1}^{T} (y_t - \widehat{y}_t)^2, \text{ where } \widehat{y}_t = x_t' \widehat{\omega}_2. \tag{22}$$

Let P denote the distribution of the data $\{W_t = (r_{t+1}, \sigma_{t+1}^2) : t \geq 1\}$ 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

$$f_t(\omega) = \begin{pmatrix} h_t(\omega_1) \\ x_t(y_t - x_t'\omega_2) \\ (y_t - x_t'\omega_2)^2 \end{pmatrix} \in R^{d_f} \text{ and } V = \sum_{m=-\infty}^{\infty} \mathbb{C}\text{ov}\left[f_t(\omega_0), f_{t+m}(\omega_0)\right]. \tag{23}$$

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 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 the regressors to not exhibit multicollinearity. We make the necessary 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 $0 < C < \infty$.

1. $T^{-1}\sum_{t=1}^{T}(h_t(\omega_1) - \mathbb{E}[h_t(\omega_1)) \to_p 0$ and $T^{-1}\sum_{t=1}^{T}(H_t(\omega_1) - \mathbb{E}[H_t(\omega_1)]) \to_p 0$, $\mathbb{E}[H_t(\omega_1)]$ is continuous in ω_1, all uniformly over the parameter space of ω_1.

2. $T^{-1}\sum_{t=1}^{T}(x_t x_t' - \mathbb{E}[x_t x_t']) \to_p 0$.

3. $V^{-1/2}\{T^{-1/2}(\sum_{t=1}^{T} f_t(\omega_0) - \mathbb{E}[f_t(\omega_0)]\} \to_d N(0, I)$ and $\widehat{V} - V \to_p 0$.

4. $C^{-1} \leq \lambda_{\min}(A) \leq \lambda_{\max}(A) \leq C$ 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

$$\begin{aligned} \mathcal{B} &= \operatorname{diag}\{[H(\omega_{10})V_1^{-1}H(\omega_{10})]^{-1}H(\omega_{10})V_1^{-1}, \mathbb{E}[x_t x_t']^{-1}, 1\}, \\ \widehat{\mathcal{B}} &= \operatorname{diag}\{[\overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}\overline{H}(\widehat{\omega}_1)]^{-1}\overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}, [T^{-1}\sum_{t=1}^{T} x_t x_t']^{-1}, 1\}. \end{aligned} \tag{24}$$

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. *Suppose Assumption R holds. The following results hold uniformly over $P \in \mathcal{P}$.*

$$\xi_T := \Omega^{-1/2}T^{-1/2}(\widehat{\omega} - \omega_0) \to_d \xi \sim N(0, I), \text{ where } \Omega = \mathcal{B}V\mathcal{B}',$$

and

$$\widehat{\Omega} - \Omega \to_p 0, \text{ where } \widehat{\Omega} = \widehat{\mathcal{B}}\widehat{V}\widehat{\mathcal{B}}'.$$

4.2 Weak Identification

The true value of the structural parameter θ and the reduced-form parameter ω satisfy the link function $g(\theta_0, \omega_0) = 0$. In a standard problem without any identification issues, we can estimate θ_0 by the minimum distance estimator $\widehat{\theta} = (\widehat{\kappa}, \widehat{\pi}, \widehat{\phi})'$, which minimizes $Q_T(\theta) = g(\theta, \widehat{\omega})'W_T g(\theta, \widehat{\omega})$ for some weighting matrix W_T, and construct tests and confidence sets for θ_0 using an asymptotically normal approximation for $T^{1/2}(\widehat{\theta} - \theta_0)$. However, this standard method does not work in the present problem when π_0 is only weakly identified. In this case, $g(\theta, \widehat{\omega})$ 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 $\widehat{\kappa}$ and $\widehat{\phi}$, making the distribution of $\widehat{\kappa}$ and $\widehat{\phi}$ non-normal even in large samples.

Before presenting the robust confidence set, we first introduce some useful quantities and provide a heuristic discussion of the identification problem and its consequences. Let $G(\theta, \omega)$ denote the partial derivative of $g(\theta, \omega)$ with respect to (w.r.t.) ω. Let $g_0(\theta) = g(\theta, \omega_0)$ and $G_0(\theta) = G(\theta, \omega_0)$ be the link function and its derivative evaluated at ω_0 and $\widehat{g}(\theta) = g(\theta, \widehat{\omega})$ and $\widehat{G}(\theta) = G(\theta, \widehat{\omega})$ be the same quantities evaluated at the estimator $\widehat{\omega}$. The delta method gives

$$\eta_T(\theta) := T^{1/2}\left[\widehat{g}(\theta) - g_0(\theta)\right] = G_0(\theta)\Omega^{1/2} \cdot \xi_T + o_p(1), \tag{25}$$

where $\xi_T \rightarrow_d N(0, I)$ following Lemma 2. Thus, $\eta_T(\cdot)$ weakly converges to a Gaussian process $\eta(\cdot)$ with covariance function $\Sigma(\theta_1, \theta_2) = G_0(\theta_1)\Omega G_0(\theta_2)'$.

Following (25), we can write $T^{1/2}\widehat{g}(\theta) = \eta_T(\theta) + T^{1/2}g_0(\theta)$, where $\eta_T(\theta)$ is the noise from the reduced-form parameter estimation and $T^{1/2}g_0(\theta)$ is the signal from the link function. Under weak identification, $g_0(\theta)$ is almost flat in θ, modeled as the signal $T^{1/2}g_0(\theta)$ being finite even for $\theta \neq \theta_0$ and $T \rightarrow \infty$. Thus, the signal and the noise are of the same order of magnitude, yielding an inconsistent minimum distance estimator $\widehat{\theta}$. This is in contrast with the strong identification scenario, where $T^{1/2}g_0(\theta) \rightarrow \infty$ for $\theta \neq \theta_0$ as $T \rightarrow \infty$ and $g_0(\theta_0) = 0$. In this case, the signal is strong enough that the minimum distance estimator is consistent.

The identification strength of θ_0 is determined by the function $T^{1/2}g_0(\theta)$. 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 $T^{1/2}g_0(\theta)$ as an infinite dimensional nuisance parameter for the inference of θ_0. The goal is to construct robust confidence set for θ_0 that has correct size asymptotically regardless of this unknown nuisance parameter.

4.3 Conditional QLR Test

We construct a confidence set for $\theta \in \Theta := [0, M_1] \times [-M_2, 0] \times [1 - \epsilon, 0]$ by inverting the test $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$, where M_1 and M_2 are large positive constants and ϵ is a small positive constant. The test statistic is a QLR statistic that takes the form

$$QLR(\theta_0) := T\widehat{g}(\theta_0)'\widehat{\Sigma}(\theta_0, \theta_0)^{-1}\widehat{g}(\theta_0) - \min_{\theta \in \Theta} T\widehat{g}(\theta)'\widehat{\Sigma}(\theta, \theta)^{-1}\widehat{g}(\theta), \tag{26}$$

where $\widehat{\Sigma}(\theta_1, \theta_2,) = \widehat{G}(\theta_1)\widehat{\Omega}\widehat{G}(\theta_2)'$ 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 $\widehat{g}(\theta)$ is replaced by a sample moment. The same method can be applied to the present nonlinear minimum distance problem. Following Andrews and Mikusheva (2016), we first project $\widehat{g}(\theta)$ onto $\widehat{g}(\theta_0)$ and construct a residual process

$$\widehat{r}(\theta) = \widehat{g}(\theta) - \widehat{\Sigma}(\theta, \theta_0)\widehat{\Sigma}(\theta_0, \theta_0)^{-1}\widehat{g}(\theta_0). \tag{27}$$

The limiting distributions of $\widehat{r}(\theta)$ and $\widehat{g}(\theta_0)$ are Gaussian and independent. Thus, conditional on $\widehat{r}(\theta)$, the asymptotic distribution of $\widehat{g}(\theta)$ no longer depends on the nuisance parameter, $T^{1/2}g_0(\theta)$, making the procedure robust to any identification strength.

Specifically, we obtain the $1-\alpha$ conditional quantile of the QLR statistic, denoted by $c_{1-\alpha}(r, \theta_0)$, as follows. For $b = 1, \ldots, B$, we take independent draws $\eta_b^* \sim N(0, \widehat{\Sigma}(\theta_0, \theta_0))$ and produce a simulated process,

$$g_b^*(\theta) := \widehat{r}(\theta) + \widehat{\Sigma}(\theta, \theta_0)\widehat{\Sigma}(\theta_0, \theta_0)^{-1}\eta_b^*, \tag{28}$$

and a simulated statistic,

$$QLR_b^*(\theta_0) := T\widehat{g}(\theta_0)'\widehat{\Sigma}(\theta_0, \theta_0)^{-1}\widehat{g}(\theta_0) - \min_{\theta \in \Pi} T g_b^*(\theta)'\widehat{\Sigma}(\theta, \theta)^{-1} g_b^*(\theta). \tag{29}$$

Let $b_0 = \lceil (1-\alpha)B \rceil$, the smallest integer greater than or equal to $(1-\alpha)B$. Then the critical value $c_{1-\alpha}(r, \theta_0)$ is the b_0^{th} smallest value among $\{QLR_b^*, b = 1, \ldots, B\}$. We execute the steps reported in Algorithm 1 to form a robust confidence set for θ.

Algorithm 1 Construing the Confidence Set

1. Estimate the reduced-form parameter $\widehat{\omega} = (\widehat{\omega}_1, \widehat{\omega}_2, \widehat{\omega}_3)'$ following the estimators defined in (20), (21), and (22).
2. Obtain a consistent estimator of $\widehat{\omega}$'s asymptotic covariance $\widehat{\Omega} = \widehat{\mathcal{B}}\widehat{V}\widehat{\mathcal{B}}'$, where $\widehat{\mathcal{B}}$ is defined in (24) and $\widehat{V}$ is a HAC estimator of V.
3. For $\theta_0 \in \Theta$,
   (a) Construct the QLR statistic $QLR(\theta_0)$ in (26) using $g(\theta, \omega)$, $G(\theta, \omega)$, $\widehat{\omega}$, and $\widehat{\Omega}$.
   (b) Compute the residual process $\widehat{r}(\theta)$ in (27).
   (c) Given $\widehat{r}(\theta)$, compute the critical value $c_{1-\alpha}(r, \theta_0)$ described above.
4. Repeat these steps for different values of θ_0. Construct a confidence set by collecting the null values that are not rejected, i.e., the nominal level $1-\alpha$ confidence set for θ_0 is

$$CS_T = \{\theta_0 : QLR_T(\theta_0) \leq c_{1-\alpha}(r, \theta_0)\}. \tag{30}$$

To obtain confidence intervals for each element of θ_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 applying the conditional inference approach above, see Andrews and Mikusheva (2016, Section 5). However, this concentration approach only works when the nuisance parameter is strongly identified. In the present set-up, this approach does not work for κ and ϕ because the nuisance parameter π is weakly identified.

Assumption S. The following conditions hold over $P \in \mathcal{P}$, for any θ in its parameter space, and any ω in some fixed neighborhood around its true value, for some fixed $0 < C < \infty$.

1. $g(\theta, \omega)$ is partially differentiable in ω, with partial derivative $G(\theta, \omega)$ that satisfies $||G(\theta_1, \omega) -$

$G(\theta_2, \omega)|| \leq C||\theta_1 - \theta_2||$ and $||G(\theta, \omega_1) - G(\theta, \omega_2)|| \leq C||\omega_1 - \omega_2||$.

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

Theorem 3. *Suppose Assumption R and Assumption S hold. Then,*

$$\liminf_{T \to \infty} \inf_{P \in \mathcal{P}} \Pr\left(\theta_0 \in CS_T\right) \geq 1 - \alpha.$$

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

5 Simulations and Empirical Results

To be added.

References

Aït-Sahalia, Yacine, Jianqing Fan, and Yingying Li. 2013. "The Leverage Effect Puzzle: Disentangling Sources of Bias at High Frequency." *Journal of Financial Economics* 109 (1): 224–249.

Andrews, Donald W. K., and Xu Cheng. 2012. "Estimation and Inference With Weak, Semi-Strong, and Strong Identification." *Econometrica* 80 (5): 2153–2211.

Andrews, Donald W.K., and Patrik Guggenberger. 2010. "Asymptotic Size and a Problem with Subsampling and with the M out of N Bootstrap." *Econometric Theory* 26 (2): 426–468.

Andrews, Isaiah, and Anna Mikusheva. 2015. "Maximum Likelihood Inference in Weakly Identified Dynamic Stochastic General Equilibrium Models." *Quantitative Economics* 6 (1): 123–152.

———. 2016. "Conditional Inference With a Functional Nuisance Parameter." *Econometrica* 84 (4): 1571–1612.

Bali, Turan G., and Lin Peng. 2006. "Is There a Risk–Return Trade-off? Evidence from High-Frequency Data." *Journal of Applied Econometrics* 21 (8): 1169–1198.

Bandi, Federico M., and Roberto Renò. 2012. "Time-varying Leverage Effects." *Journal of Econometrics* 169 (1): 94–113.

Bansal, Ravi, Dana Kiku, Ivan Shaliastovich, and Amir Yaron. 2014. "Volatility, the Macroeconomy, and Asset Prices." *The Journal of Finance* 69 (6): 2471–2511.

Bollerslev, Tim, Robert F. Engle, and Jeffrey M. Wooldridge. 1988. "A Capital-Asset Pricing Model with Time-varying Covariances." *Journal of Political Economy* 96 (1): 116–131.

Bollerslev, Tim, Tzuo Hann Law, and George Tauchen. 2008. “Risk, Jumps and Diversification.” *Journal of Econometrics* 144 (1): 234–256.

Brandt, Michael W., and Qiang Kang. 2004. “On the Relationship Between the Conditional Mean and Volatility of Stock Returns: A Latent VAR Approach.” *Journal of Financial Economics* 72 (2): 217–257.

Breen, William, Lawrence R. Glosten, and Ravi Jagannathan. 1989. “Economic Significance of Predictable Variations in Stock Index Returns.” *The Journal of Finance* 44 (5): 1177–1189.

Campbell, John Y. 1987. “Stock Returns and the Term Structure.” *Journal of Financial Economics* 18 (2): 373–399.

Christoffersen, Peter, Steven Heston, and Kris Jacobs. 2013. “Capturing Option Anomalies with a Variance-Dependent Pricing Kernel.” *The Review of Financial Studies* 26 (8): 1963–2006.

Darolles, Serge, Christian Gouriéroux, and Joann Jasiak. 2006. “Structural Laplace Transform and Compound Autoregressive Models.” *Journal of Time Series Analysis* 27 (4): 477–503.

Dew-Becker, Ian, Stefano Giglio, Anh Le, and Marius Rodriguez. 2017. “The Price of Variance Risk.” *Journal of Financial Economics* 123 (2): 225–250.

Drechsler, Itamar, and Amir Yaron. 2011. “What’s Vol Got to Do with It.” *The Review of Financial Studies* 24 (1): 1–45.

Feunou, Bruno, Jean-Sébastien Fontaine, Abderrahim Taamouti, and Roméo Tédongap. 2014. “Risk Premium, Variance Premium, and the Maturity Structure of Uncertainty.” *Review of Finance* 18 (1): 219–269.

Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov. 2005. “There is a Risk-Return Trade-off after All.” *Journal of Financial Economics* 76 (3): 509–548.

Gouriéroux, Christian, and Joann Jasiak. 2006. “Autoregressive Gamma Processes.” *Journal of Forecasting* 25 (2): 129–152.

Han, Hyojin, Stanislav Khrapov, and Eric Renault. 2018. *The Leverage Effect Puzzle Revisited: Identification in Discrete Time.* Working Paper. Brown University.

Harvey, Campbell R. 1989. “Time-Varying Conditional Covariances in Tests of Asset Pricing Models.” *Journal of Financial Economics* 24 (2): 289–317.

Heston, Steven L. 1993. “A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options.” *The Review of Financial Studies* 6 (2): 327–343.

Kleibergen, Frank. 2005. “Testing Parameters in GMM Without Assuming That They Are Identified.” *Econometrica* 73 (4): 1103–1123.

Lettau, Martin, and Sydney C. Ludvigson. 2010. “Measuring and Modeling Variation in the Risk-Return Tradeoff.” Chap. 11, edited by Yacine Aït-Sahalia and Lars Peter Hansen, 1:617–690. Handbook of Financial Econometrics. Elsevier.

Lintner, John. 1965. “Security Prices, Risk, and Maximal Gains From Diversification.” *The Journal of Finance* 20 (4): 587–615.

Ludvigson, Sydney C., and Serena Ng. 2007. “The Empirical Risk–Return Relation: A Factor Analysis Approach.” *Journal of Financial Economics* 83 (1): 171–222.

Mavroeidis, Sophocles, Mikkel Plagborg-Møller, and James H. Stock. 2014. “Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve.” *Journal of Economic Literature* 52, no. 1 (March): 124–88.

Mikusheva, Anna. 2007. “Uniform Inference in Autoregressive Models.” *Econometrica* 75 (5): 1411–1452.

Moreira, Marcelo J. 2003. “A Conditional Likelihood Ratio Test for Structural Models.” *Econometrica* 71 (4): 1027–1048.

Pagan, Adrian R., and Y.S. Hong. 1991. “Nonparametric Estimation and the Risk Premium.” Chap. 2 in *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics,* edited by William A. Barnett, James Powell, and George Tauchen, 51–75. Cambridge University Press.

Sharpe, William F. 1964. “Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk.” *The Journal of Finance* 19 (3): 425–442.

Staiger, Douglas, and James H. Stock. 1997. “Instrumental Variables Regression with Weak Instruments.” *Econometrica* 65 (3): 557–586.

Stock, James H., and Jonathan H. Wright. 2000. “GMM with Weak Identification.” *Econometrica* 68 (5): 1055–1096.

Whitelaw, Robert F. 1994. “Time Variations and Covariations in the Expectation and Volatility of Stock Market Returns.” *The Journal of Finance* 49 (2): 515–541.

Appendix A Proofs

A.1 Proof of Lemma 1

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

$$\begin{aligned}
1 &= E\left[\exp\left(m_0 + m_1\sigma_t^2 - \pi\sigma_{t+1}^2 - \kappa r_{t+1}\right) \middle| \mathcal{F}_t\right] \\
&= \exp(m_0 + m_1\sigma_t)E\left[\exp\left(-\pi\sigma_{t+1}^2\right) E\left[\exp\left(-\kappa r_{t+1}\right) \middle| \mathcal{F}_t, \sigma_{t+1}^2\right] \middle| \mathcal{F}_t\right] \\
&= \exp(m_0 - E\left(\kappa\right) + m_1\sigma_t - D\left(\kappa\right)\sigma_t^2)E\left[\exp\left(-\pi\sigma_{t+1}^2 - C\left(\kappa\right)\sigma_{t+1}^2\right) \middle| \mathcal{F}_t\right] \\
&= \exp(m_0 - E\left(\kappa\right) + m_1\sigma_t - D\left(\kappa\right)\sigma_t^2 - A\left(\pi + C\left(\kappa\right)\right)\sigma_t^2 - B\left(\pi + C\left(\kappa\right)\right)), \qquad (31)
\end{aligned}$$

where the first equality follows from the pricing equation, the second equality follows from the law of iterated expectations, the third equation uses the Laplace transform for r_{t+1} in (9), and the last equality follows from the Laplace transform for σ_{t+1}^2 in (4). Since $M_{t,t+1}$ must integrate to 1, the constant term and coefficient for σ_t^2 must equal 0, which gives the claimed result for m_0 and m_1.

We can apply the same argument above to any asset r_{t+1}. This gives the same result, except κ is replaced by $\kappa - 1$ throughout. This implies that the two equalities for m_0 and m_1 also hold with κ replaced by $\kappa - 1$. Therefore,

$$\begin{aligned}
E(\kappa - 1) + B\left(C\left(\kappa - 1\right) + \pi\right) &= E(\kappa) + B\left(C\left(\kappa\right) + \pi\right), \\
D\left(\kappa - 1\right) + A\left(C\left(\kappa - 1\right) + \pi\right) &= D\left(\kappa\right) + A\left(C\left(\kappa\right) + \pi\right). \qquad (32)
\end{aligned}$$

The claimed results for γ and β follow from $\gamma = \mathbb{E}(\kappa) - E(\kappa - 1)$ and $\beta = D(\kappa) - D(\kappa - 1)$ under the linear specification of $E(x) = \gamma x$ and $D(x) = \beta x$.

□

A.2 Proof of Lemma 2

Proof. 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 $\|\mathbb{E}[h_t(\omega_1)]\|$ 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 (1) and (2).

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

$$
\begin{aligned}
0 &= \begin{pmatrix} \overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}\overline{h}(\widehat{\omega}_1) \\ T^{-1}\sum_{T=1}^{T} x_t(y_t - x_t'\widehat{\omega}_2) \\ \widehat{\omega}_3 - T^{-1}\sum_{t=1}^{T}\left(y_t - \widehat{y}_t\right)^2 \end{pmatrix} \\
&= \begin{pmatrix} \overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}\overline{h}(\omega_{10}) + \overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}\overline{H}(\widetilde{\omega}_1)(\widehat{\omega}_1 - \omega_{10}) \\ T^{-1}\sum_{t=1}^{T} x_t(y_t - x_t'\omega_{20}) - T^{-1}\sum_{t=1}^{T} x_t x_t'\left(\widehat{\omega}_2 - \omega_{20}\right) \\ (\widehat{\omega}_3 - \omega_3) + \omega_3 - T^{-1}\sum_{t=1}^{T}\left(y_t - x_t\widehat{\omega}_2\right)^2 \end{pmatrix}, \qquad (33)
\end{aligned}
$$

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

$$
\widetilde{\mathcal{B}} = \operatorname{diag}\left\{[\overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}\overline{H}(\widetilde{\omega}_1)]^{-1}\overline{H}(\widehat{\omega}_1)'\widehat{V}_1^{-1}, [T^{-1}\sum_{t=1}^{T} x_t x_t']^{-1}, 1\right\}. \qquad (34)
$$

Then (33) implies that

$$
\begin{aligned}
T^{1/2}\left(\widehat{\omega} - \omega\right) &= \widetilde{\mathcal{B}} \cdot T^{-1/2}\sum_{t=1}^{T}\begin{pmatrix} -h_t(\omega_{10}) \\ x_t(y_t - x_t'\omega_{20}) \\ \left(y_t - x_t\widehat{\omega}_2\right)^2 - \omega_3 \end{pmatrix} \\
&= \widetilde{\mathcal{B}} \cdot T^{-1/2}\sum_{t=1}^{T}\begin{pmatrix} -h_t(\omega_{10}) \\ x_t(y_t - x_t'\omega_{20}) \\ \left(y_t - x_t'\omega_{20}\right)^2 - \mathbb{E}\left[\left(y_t - x_t'\omega_{20}\right)^2\right] \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \varepsilon_T \end{pmatrix}, \qquad (35)
\end{aligned}
$$

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

$$
\begin{aligned}
\varepsilon_T &= T^{-1/2}\sum_{t=1}^{T}\left[\left(y_t - x_t'\widehat{\omega}_2\right)^2 - \left(y_t - x_t'\omega_{20}\right)^2\right] \\
&= 2T^{-1}\sum_{t=1}^{T}\left(y_t - x_t'\omega_{20}\right)x_t'\left[T^{1/2}\left(\widehat{\omega}_2 - \omega_{20}\right)\right] + o_p(1) \\
&= o_p(1) \qquad (36)
\end{aligned}
$$

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,

$$
\widetilde{\mathcal{B}} \to_p \mathcal{B} \qquad (37)
$$

following from the consistency of $\widehat{\omega}_1$ and Assumption R. Finally, the desirable result follows from (35)–(37) and Assumption R. The consistency of $\widehat{\Omega}$ follows from the consistency of $\widehat{\mathcal{B}}$ and $\widehat{V}$.

□

A.3 Proof of Theorem 3

Proof. We obtain this result by applying Andrews and Mikusheva (2016, Theorem 1). We now verify Assumptions 1-3 in Andrews and Mikusheva (2016). To show weak convergence $\eta_T(\cdot)$ to $\eta(\cdot)$ uniformly over $\mathcal{P}$, note that by a second-order Taylor expansion,

$$\begin{aligned}\eta_T(\theta) &\coloneqq T^{1/2}\left[\widehat{g}(\theta) - g_0(\theta)\right] = G_0(\theta)\Omega^{1/2}\xi_T + \delta_T, \text{ where} \\ \xi_T &= \Omega^{-1/2}T^{1/2}\left(\widehat{\omega} - \omega_0\right), \delta_T = \left(G(\theta,\widetilde{\omega}) - G(\theta,\omega_0)\right)T^{1/2}(\widehat{\omega} - \omega_0),\end{aligned} \tag{38}$$

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

$$\begin{pmatrix} G_0(\theta_1)\Omega^{1/2}\xi_T \\ G_0(\theta_2)\Omega^{1/2}\xi_T \end{pmatrix} \to_d \begin{pmatrix} \eta(\theta_1) \\ \eta(\theta_2) \end{pmatrix}, \tag{39}$$

which follows from Lemma 2, and (ii) the stochastic equicontinuity condition, i.e., for every $\varepsilon > 0$ and $\chi > 0$, there exists a $\delta > 0$ such that

$$\limsup_{T\to\infty} \Pr\left(\sup_{P\in\mathcal{P}} \sup_{\|\theta_1-\theta_2\|\leq\delta} \left\| G_0(\theta_1)\Omega^{1/2}\xi_T - G_0(\theta_2)\Omega^{1/2}\xi_T \right\| > \varepsilon\right) < \chi. \tag{40}$$

For some $C < \infty$, we have $\|G_0(\theta_1) - G(\theta_2)\| \leq C\|\theta_1 - \theta_2\|$ by Assumption S, and we have $\|\Omega^{1/2}\| \leq C$ under Assumption R because F and V both have bounded largest eigenvalues. Thus,

$$\begin{aligned}&\limsup_{T\to\infty} \Pr\left(\sup_{P\in\mathcal{P}} \sup_{\|\theta_1-\theta_2\|\leq\delta} \left\| G_0(\theta_1)\Omega^{1/2}\xi_T - G_0(\theta_2)\Omega^{1/2}\xi_T \right\| > \varepsilon\right) \\ &\leq \limsup_{T\to\infty} \Pr\left(C^2 \sup_{P\in\mathcal{P}} \|\xi_T\| > \frac{\varepsilon}{\delta}\right).\end{aligned} \tag{41}$$

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 (41) smaller than χ.

Assumptions 2 and 3 of Andrews and Mikusheva (2016, Theorem 1) follow from Assumption R.

□