Macro-Finance Decoupling: Robust Evaluations of Macro Asset Pricing Models*

Xu Cheng†, Winston Wei Dou‡, Zhipeng Liao§

August 23, 2021

Abstract

This paper shows that robust inference under weak identification is important to the evaluation of many influential macro asset pricing models, including (time-varying) rare-disaster risk models and long-run risk models. Building on recent developments in the conditional inference literature, we provide a novel conditional specification test by simulating the critical value conditional on a sufficient statistic. This sufficient statistic can be intuitively interpreted as a measure capturing the macroeconomic information decoupled from the underlying content of asset pricing theories. Macro-finance decoupling is an effective way to improve the power of the specification test when asset pricing theories are difficult to refute because of a severe imbalance in the information content about the key model parameters between macroeconomic moment restrictions and asset pricing cross-equation restrictions. We apply the proposed conditional specification test to the evaluation of a time-varying rare-disaster risk model and the construction of robust model uncertainty sets.

Keywords: Conditional inference, Information imbalance, Long-run risk, Rare disasters, Structural asset pricing, Weak identification.
JEL Classification: C12, C32, C52, G12.

---

*We thank Donald Andrews, Isaiah Andrews, Xiaohong Chen, Frank Diebold, Lars Peter Hansen, Ken Judd, Frank Kleibergen, Jia Li, Sophocles Mavroeidis, Adam McCloskey, Anna Mikusheva, Marcelo Moreira, Markus Pelger, Mikkel Plagborg-Møller, Tom Sargent, Frank Schorfheide, Ivan Shaliastovich, Rob Stambaugh, Jessica Wachter, Ole Wilms, Jonathan Wright, Amir Yaron, Motohiro Yogo, Chu Zhang, and seminar and conference participants at Cambridge, Cornell, Duke, Federal Reserve Bank at Boston, Harvard-MIT, HKUST, McGill, Oxford, Penn, Pittsburgh, Princeton, Rutgers, Simon Fraser, Standford SITE, Syracuse, UC Riverside, Wharton, Cowels Foundation conference, Chamberlain Online Seminar, Dongbei econometrics conference, Econometric Society North American and Asian meetings, and the Society of Financial Econometrics annual conference for their comments. We are very grateful to Di Tian for his superb research assistance. Cheng and Dou acknowledge financial supports from the Jacobs Levy equity management center for quantitative financial research. Liao's research was partially supported by the National Science Foundation [Grant SES-1628889].

†Department of Economics, University of Pennsylvania; Email: xucheng@econ.upenn.edu.

‡Finance Department, The Wharton School, University of Pennsylvania; Email: wdou@wharton.upenn.edu.

§Department of Economics, University of California, Los Angeles; Email: zhipeng.liao@econ.ucla.edu.

1 Introduction

Many influential macro asset pricing models have been developed and widely used by researchers, practitioners, and monetary authorities.[1] The econometric evaluation of their specifications is important for policy analysis and forecasting, but conventional methods face severe challenges due to ubiquitous information imbalance in these models; see, among others, Campbell (2018) and Chen, Dou, and Kogan (2020) for structural models, as well as Kan and Zhang (1999), Kleibergen (2009), and Gospodinov, Kan, and Robotti (2017) for linear models with spurious factors. The econometric question of how to construct a robust and efficient evaluation of macro asset pricing models, especially nonlinear structural models, is of great importance for the asset pricing literature.

To address this crucial question, this paper proposes a specification test robust to information imbalance in a unified weak identification framework. In the generalized method of moments (GMM) setup, we show that the rare-disaster risk model (Rietz, 1988; Barro, 2006), the time-varying rare-disaster risk model (Gabaix, 2012; Wachter, 2013), and the long-run risk model (Bansal and Yaron, 2004) all fit in a general framework — there exists a set of baseline moments that are valid regardless of the asset pricing theory but only provide weak identification of the key model parameters, characterized by the near flatness of the baseline moments (Stock and Wright, 2000).[2] The remainder of the moments are asset pricing moments implied by a specific macro asset pricing theory. These asset pricing moments, by design, provide tighter cross-equation restrictions and thus strongly identify the key model parameters. We propose a new conditional specification test that evaluates the validity of the asset pricing moments by effectively exploiting the noisy information embedded in the baseline moments. This approach yields substantial power improvement compared to the widely used J test (Hansen, 1982) for models with information imbalance.

This new conditional specification test builds on the approach of conditional inference with a functional nuisance parameter by Andrews and Mikusheva (2016a). In our analysis, the object of interest shifts from structural parameters to model specifications. The test statistic is an incremental J statistic as in the C test of Eichenbaum, Hansen, and Singleton (1988). However, the critical value is simulation-based and is conditional on a sufficient statistic that captures the macroeconomic information decoupled from the underlying content of asset pricing theories. It has correct asymptotic size uniformly over the identification strength in the baseline moments. In contrast, the C test, with a critical value based on the chi-square distribution, may under- or over-reject the null hypothesis because estimators for weakly identified parameters have distributions that are

[1] See, e.g., Campbell (2003), Brunnermeier and Sannikov (2016), Cochrane (2017), He and Krishnamurthy (2018), and Dou, Fang, Lo, and Uhlig (2020), for reviews on macro-finance models.

[2] See, e.g., Stock, Wright, and Yogo (2002), Andrews and Stock (2007), Mikusheva (2013), and Andrews, Stock, and Sun (2019), for reviews on weak instruments and weak identification.

poorly approximated by a normal distribution even in large samples (e.g., Stock and Wright, 2000; Andrews and Cheng, 2012). Our conditional specification test becomes equivalent to the optimal test in the classical scenario without weak identification (Newey, 1985). In contrast, the J test based on all moment conditions neglects the valid information in the baseline moments.

This paper contributes to the connection between finance and econometrics, and makes the following methodological and empirical contributions. First, we study the identification issue in (time-varying) rare-disaster risk and long-run risk models. Our study shows how to coherently fit a structural macro asset pricing model featuring information imbalance into a formal econometric framework with both weak and strong identification, which paves the way for uniformly valid inference. Each example demonstrates how to conduct a model-specific reparameterization that exploits the major implications of the macro asset pricing model. To the best of our knowledge, the identification and robust testing problems we study here have never been explored for these macro asset pricing structural models under a formal econometric framework. Moreover, the empirical application analyzes a full-blown time-varying rare-disaster risk model. There has been little formal econometric study of the validity of the rare-disaster risk mechanisms in the asset pricing literature, with a few exceptions (e.g., Julliard and Ghosh, 2012). Our empirical study also fills this gap in the literature.

Second, we develop a new robust specification test and study its asymptotic properties. We consider a general GMM framework, although the applications focus on asset pricing models. The proposed conditional specification test builds on the work of Andrews and Mikusheva (2016a) and differs from it in two main ways. First, our hypothesis test is on a model rather than the value of some parameters. We estimate the model parameters instead of plugging in their true values under the null because the latter is not available for the purpose of model evaluation. Estimation of unknown parameters introduces additional technical complications for uniformly valid asymptotic analysis. Second, we explore the differences in both validity and information content across two sets of moments instead of studying inference based on one set of near-flat moments. Such imbalance in the information content is the key feature of the models we study.

Related Literature. We contribute to improving upon existing specification tests and broadening the scope of weak-identification robust inference methods. Conditional inference has been successfully applied in constructing confidence sets for weakly identified parameters, following the pioneering work of Moreira (2003) for linear instrumental variable (IV) models. Kleibergen (2005) extends its application to nonlinear GMM models. Furthermore, Andrews and Mikusheva (2016a) provide a new perspective – viewing the near-flat population moment function as a functional nui-

sance parameter. Standing on their shoulders, we apply the conditional inference approach to the evaluation of nonlinear structural models, a new setting with many important applications. Hahn, Ham, and Moon (2011) study a generalized Hausman test robust to weak instruments in linear models and discuss several micro-econometric applications where our conditional specification test also applies.

A growing body of literature is concerned with the efficacy of conventional inference methods for macro asset pricing models and the development of robust methods. Using identification-robust inference methods, Stock and Wright (2000) study the preference parameters through nonlinear Euler equations, Yogo (2004) investigates the elasticity of intertemporal substitution (EIS) parameter through linearized cross-equation restrictions, and Ascari, Magnusson, and Mavroeidis (2019) study the preference parameters under different structural models featuring habits, hand-to-mouth consumers, or recursive preferences. Recently, a large number of papers have studied robust inference methods for linear asset pricing models (e.g., Kan and Zhang, 1999; Gospodinov, Kan, and Robotti, 2017; Kleibergen and Zhan, 2020; Anatolyev and Mikusheva, 2020), Further, in predictive models of stock returns with highly persistent predictors, standard asymptotic inference can largely fail (e.g., Elliott and Stock, 1994; Stambaugh, 1999), and new valid and efficient procedures have been developed (e.g., Campbell and Yogo, 2006; Elliott, Müller, and Watson, 2015). Nevertheless, the existing literature lacks reliable and powerful model evaluation methods in the presence of information imbalance.

The rest of the paper is organized as follows. Section 2 provides the general GMM setup with information imbalance and two motivating examples. Section 3 describes the conditional specification test, provides the algorithm, and illustrates its finite-sample performance through Monte Carlo simulations based on the two motivating examples. Section 4 establishes the theoretical results on the size of the proposed test and its uniform validity. Section 5 contains an empirical application of the proposed test to time-varying rare-disaster risk models. Section 6 concludes. The appendix contains proofs of the theoretical results on the size of the test. The supplemental appendix contains proofs of the auxiliary lemmas in the appendix, theoretical and simulation results on the power of the test, and details for the empirical application. Moreover, a note on additional materials Cheng, Dou, and Liao (2021) is available on the authors' personal websites.

2 Information Imbalance: General Setup and Examples

General Setup. Our objective is to statistically assess the validity of a macro asset pricing model by applying specification tests to a set of model-implied cross-equation restrictions. The

specification test can be formulated as below:

$$H_0 : \mathbb{E}[\bar{g}_1(\theta_0)] = 0_{k_1\times 1} \quad \text{versus} \quad H_1 : \mathbb{E}[\bar{g}_1(\theta_0)] \neq 0_{k_1\times 1}, \tag{2.1}$$

where $\bar{g}_1(\theta) \equiv n^{-1}\sum_{i=1}^n g_{1,t}(\theta)$ and $g_{1,t}(\theta) \equiv g_1(Y_t,\theta) \in \mathbb{R}^{k_1}$ depends on the data Y_t and the $d_\theta \times 1$ dimensional parameter θ, whose true value is denoted by θ_0. Some additional baseline moments are always valid under both the null and the alternative,

$$\mathbb{E}[\bar{g}_0(\theta_0)] = 0_{k_0\times 1}, \tag{2.2}$$

where $\bar{g}_0(\theta) \equiv n^{-1}\sum_{i=1}^n g_{0,t}(\theta)$ and $g_{0,t}(\theta) \equiv g_0(Y_t,\theta) \in \mathbb{R}^{k_0}$.[3] Under the null, the full moments are $\mathbb{E}\left[\bar{g}(\theta_0)\right] = 0_{k\times 1}$, where $\bar{g}(\theta) \equiv [\bar{g}_0(\theta)',\ \bar{g}_1(\theta)']'$ is $k \equiv k_0 + k_1$ dimensional and differentiable in θ almost surely. We allow the baseline moments to depend only on a subvector of θ.

The models we consider have two key features. First, the baseline moments $\mathbb{E}[\bar{g}_0(\theta_0)] = 0_{k_0\times 1}$ may only weakly identify θ_0, i.e., $\mathbb{E}[\bar{g}_0(\theta)]$ is nearly flat in θ. Second, once the asset pricing theory is imposed, the full moments $\mathbb{E}[\bar{g}(\theta_0)] = 0_{k\times 1}$ strongly identify θ_0 under the null, i.e., the singular values of the associated Jacobian matrix $Q \equiv \mathbb{E}[\partial\bar{g}(\theta_0)/\partial\theta']$ are bounded away from zero. For moments created with IVs, the baseline moments $\mathbb{E}[\bar{g}_0(\theta_0)] = 0_{k_0\times 1}$ and the additional moments $\mathbb{E}[\bar{g}_1(\theta_0)] = 0_{k_1\times 1}$ are created with weak IVs and strong IVs, respectively.

Motivating Example 1: Disaster Risk Model for Equity Premium. We consider a simple variant of Rietz (1988) and Barro (2006, 2009) with macroeconomic disasters characterized by extremely large consumption declines. We assume that real consumption growth follows[4]

$$\Delta c_t \equiv \ln\left(\frac{C_t}{C_{t-1}}\right) = \sigma\varepsilon_t - \zeta_t, \tag{2.1}$$

where C_t is real consumption per capita, the consumption shock ε_t follows a standard normal distribution, and ζ_t is a disaster variable. Particularly, the disaster variable ζ_t is characterized by

$$\zeta_t \equiv x_t(\underline{v} + J_t), \tag{2.2}$$

where the variable $x_t \sim \text{Bernoulli}(p)$ captures the occurrence of disasters, the constant $\underline{v}$ is the lower bound of the disaster size, and the variable $J_t \sim \text{Exp}(\alpha)$ is a disaster shock. The shocks $(\varepsilon_t, J_t, x_t)$ are independently and identically distributed (i.i.d.) over t and mutually independent with each other. Specification of Δc_t in (2.1) provides baseline moment conditions: $\mathbb{E}[\bar{m}_0(\alpha)] = 0$, where $\bar{m}_0(\alpha) \equiv n^{-1}\sum_{t=1}^n m_{0,t}(\alpha)$ with

$$m_{0,t}(\alpha) \equiv \begin{bmatrix} \Delta c_t + p\mu_1(\alpha) \\ \Delta c_t^2 - \sigma^2 - p\mu_2(\alpha) \end{bmatrix} \quad \text{and} \quad \mu_j(\alpha) \equiv \mathbb{E}[(\underline{v} + J_t)^j] > 0 \text{ for } j = 1, 2. \tag{2.3}$$

[3] Throughout the paper, we suppress the dependence of $\bar{g}_0(\theta)$ and $\bar{g}_1(\theta)$ on n for notational simplicity.

[4] We ignore the intercept in Δc_t to maintain simplicity since it plays little role in explaining equity premia.

Specifically, $\mu_1(\alpha) = \underline{v} + 1/\alpha$ and $\mu_2(\alpha) = \underline{v}^2 + 2\underline{v}/\alpha + 2/\alpha^2$. For illustrative purposes, we assume that the econometrician knows all parameters, except α, a parameter that can be only weakly identified by the moments based on Δc_t if p is close to 0.

The representative agent maximizes his lifetime expected utility:

$$U_0 \equiv \mathbb{E}\left[\sum_{t=0}^{\infty} e^{-\delta t} \frac{C_t^{1-\gamma}}{1-\gamma}\right], \tag{2.4}$$

where δ is the subjective discount rate and γ is the relative risk aversion coefficient. The Euler equation for the utility maximization problem gives the following moment condition for excess log return of equity $r^e_{m,t}$: $\mathbb{E}[\bar{m}_1(\alpha)] = 0$, where $\bar{m}_1(\alpha) \equiv n^{-1}\sum_{t=1}^{n} m_{1,t}(\alpha)$ with

$$\begin{aligned} m_{1,t}(\alpha) &\equiv r^e_{m,t} - \gamma\sigma^2 + \frac{1}{2}\sigma^2 + p\mu_1(\alpha) - \frac{p}{\alpha-\gamma}h(\alpha), \quad \text{and} \\ h(\alpha) &\equiv \alpha\left[e^{\gamma\underline{v}} - \frac{\alpha-\gamma}{\alpha-\gamma+1}e^{(\gamma-1)\underline{v}}\right]. \end{aligned} \tag{2.5}$$

We call this the asset pricing moment. The model and the equilibrium condition require that $\underline{v} > 0$ and $\alpha > \gamma > 1$, ensuring that the function $h(\alpha)$ is positive and finite.

The asset pricing moment (2.5) clearly demonstrates the key idea of the disaster risk model: when p and $\alpha - \gamma$ are both close to 0, the rare yet large disaster can generate a substantial equity premium as long as their ratio is a sizable loading in front of $h(\alpha)$ to match the moment of $r^e_{m,t}$. This parameter restriction on p and $\alpha - \gamma$ ensures that the disaster risk is a meaningful economic mechanism for explaining the equity premium even if p is small. To utilize this key insight, we transform the parameter α to θ with

$$\theta \equiv \frac{p}{\alpha-\gamma} \quad \text{and} \quad \theta \in \Theta \equiv [\underline{c}, \overline{c}], \tag{2.6}$$

for constants $0 < \underline{c} < \overline{c}$. Our analysis allows p and $\alpha - \gamma$ to be both arbitrarily close to 0, while keeping the ratio θ bounded from above and away from zero.

To parameterize all the moments in θ, we write

$$\bar{g}_0(\theta) \equiv \bar{m}_0(\theta^{-1}p + \gamma) \quad \text{and} \quad \bar{g}_1(\theta) \equiv \bar{m}_1(\theta^{-1}p + \gamma). \tag{2.7}$$

Let $\mu_j^{(1)}(\alpha) \equiv (d/d\alpha)\mu_j(\alpha)$, where $\mu_j(\alpha)$ for $j = 1, 2$ are defined in (2.3). For the baseline moments $\bar{g}_0(\theta)$, simple calculations give

$$\mathbb{E}\left[\frac{d}{d\theta}\bar{g}_0(\theta)\right] = -(\theta^{-1}p)^2\left[\mu_1^{(1)}(\theta^{-1}p+\gamma),\ -\mu_2^{(1)}(\theta^{-1}p+\gamma)\right]', \tag{2.8}$$

where $\mu_j^{(1)}(\theta^{-1}p + \gamma)$ for $j = 1, 2$ are positive and bounded. Because θ is bounded, we have

$$\lim_{p\to 0}\mathbb{E}\left[\frac{d}{d\theta}\bar{g}_0(\theta)\right] = [0,\ 0]' \quad \text{and} \quad \lim_{p\to 0}\mathbb{E}\left[\frac{d}{d\theta}\bar{g}_1(\theta)\right] = -\gamma e^{\gamma\underline{v}} \neq 0. \tag{2.9}$$

The baseline moments weakly identify θ when p is close to 0, whereas the asset pricing moment always strongly identifies θ.

Motivating Example 2: Long-Run Risk Model for Equity Premium. We consider a simple variant of the baseline model of Bansal and Yaron (2004). As shown in the literature (e.g., Müller and Watson, 2008, 2018), the time series of U.S. real output growth exhibit a long-run (low-frequency) component, denoted by x_t. However, economists debate whether U.S. real consumption growth and U.S. real stock return are significantly loaded on the long-run component, x_t, of the real output growth (e.g., Beeler and Campbell, 2012; Bansal, Kiku, and Yaron, 2012).

The long-run component of real output growth, x_t, is latent and obeys the following autoregressive process of order 1 (i.e., AR(1) process):

$$x_t = \rho x_{t-1} + \varepsilon_{x,t}. \tag{2.10}$$

The representative agent's consumption has the following log growth process:

$$\Delta c_t \equiv \ln\left(\frac{C_t}{C_{t-1}}\right) = \phi x_{t-1} + \sigma_c \varepsilon_{c,t}, \tag{2.11}$$

where C_t is real consumption per capita. The shocks $(\varepsilon_{x,t}, \varepsilon_{c,t})$ follow a standard bivariate normal distribution and are i.i.d. over t. By introducing parameter ϕ in (2.11), we allow the expected consumption growth to be weakly dependent on or independent of the long-run component x_t as in many macro asset pricing models. Specifically, when $\phi = 0$, the consumption growth process is exactly i.i.d. as in Campbell and Cochrane (1999). When $\phi > 0$, the time series of U.S. real consumption growth share the same long-run (low-frequency) component x_t, as suggested by Kandel and Stambaugh (1991), Hansen, Heaton, and Li (2008), and Schorfheide, Song, and Yaron (2018). When ϕ is positive yet near zero, the consumption growth process is nearly i.i.d., as argued by Beeler and Campbell (2012). In the model of Bansal and Yaron (2004, Table I), the loading parameter ϕ is effectively 0.034% in the monthly frequency. The specification of Δc_t in (2.11) implies the baseline moment conditions: $\mathbb{E}\left[\bar{m}_0(\rho)\right] = 0$, where $\bar{m}_0(\rho) \equiv n^{-1}\sum_{t=1}^{n} m_{0,t}(\rho)$ with

$$m_{0,t}(\rho) \equiv \begin{bmatrix} \Delta c_{t-1}\left(\Delta c_{t+1} - \rho \Delta c_t\right) \\ \Delta c_t\left(\Delta c_{t+1} - \rho \Delta c_t\right) + \rho \sigma_c^2 \end{bmatrix}. \tag{2.12}$$

For illustrative purposes, we assume that the econometrician knows all parameters except ρ, a parameter that can be only weakly identified by the moments based on Δc_t if ϕ is close to 0.

The representative agent has recursive preferences and the agent maximizes the following lifetime utility:

$$V_t = \left[(1-\delta)C_t^{1-1/\psi} + \delta\left(\mathbb{E}_t\left[V_{t+1}^{1-\gamma}\right]\right)^{\frac{1-1/\psi}{1-\gamma}}\right]^{\frac{1}{1-1/\psi}}, \tag{2.13}$$

where δ is the rate of time preference, γ is the coefficient of risk aversion for timeless gambles, ψ is the EIS under certainty, and $\mathbb{E}_t[\cdot]$ is the conditional expectation given the information up to the end of period t. The Euler equation as the first-order condition for the utility maximization problem requires that the equilibrium excess log return $r^e_{m,t}$ satisfies $\mathbb{E}\left[\bar{m}_1(\rho)\right] = 0$, where $\bar{m}_1(\rho) \equiv n^{-1}\sum_{t=1}^{n} m_{1,t}(\rho)$ with

$$m_{1,t}(\rho) \equiv r^e_{m,t} - \gamma\sigma_c^2 + \frac{1}{2}\sigma_c^2 - \frac{1}{2}\left(2\gamma - \psi^{-1} - 1\right)\left(1 - \psi^{-1}\right)\frac{\phi^2}{\left(\delta^{-1} - \rho\right)^2}. \tag{2.14}$$

We call this the asset pricing moment.

The key insight of the long-run risk model can be clearly seen from (2.14): when $\gamma > 1 > \psi^{-1}$, which implies that the agent has a preference for early resolution of uncertainty and the intertemporal substitution effect dominates the income effect, the equity premium is sizable if the cash flows load on the long-run component (i.e., ϕ is positive), the long-run component is persistent (i.e., ρ is close to unity), and the representative agent's rate of time preference is close to unity (i.e., δ is close to unity). This insight summarizes the central idea of the parameter calibrations in the works of Bansal and Yaron (2004) and Bansal, Kiku, and Yaron (2012), in which $\phi = 0.034\%$, $\rho = 0.975$, and $\delta = 0.9989$ in the monthly frequency. To ensure that the long-run risk is a meaningful economic mechanism for explaining the sizeable equity premium, $\phi/(\delta^{-1} - \rho)$ must be a positive component that is bounded away from zero and from above in order to match the moment of $r^e_{m,t}$. To utilize this insight, we transform ρ to θ with

$$\theta \equiv \frac{\phi}{\delta^{-1} - \rho} \quad \text{and} \quad \theta \in \Theta \equiv \left\{\theta \in [\underline{c}, \overline{c}] \text{ and } \delta^{-1} - \theta^{-1}\phi \in [0, 1)\right\}, \tag{2.15}$$

for some constants $0 < \underline{c} < \overline{c}$. Our analysis focuses on $\rho < 1$, $0 < \delta < 1$, and $\phi > 0$. It allows ρ and δ to be both arbitrarily close to 1 and ϕ to be arbitrarily close to 0, while keeping the ratio between any pair of $\delta^{-1} - 1$, $1 - \rho$, and ϕ bounded from above and away from zero.

To parameterize all the moments in θ, plugging in $\rho = \delta^{-1} - \theta^{-1}\phi$, we obtain

$$\bar{g}_0(\theta) \equiv \bar{m}_0(\delta^{-1} - \theta^{-1}\phi) \quad \text{and} \quad \bar{g}_1(\theta) \equiv \bar{m}_1(\delta^{-1} - \theta^{-1}\phi). \tag{2.16}$$

The Jacobian matrix for the baseline moment conditions is

$$\mathbb{E}\left[\frac{d}{d\theta}\bar{g}_0(\theta)\right] = -\frac{\phi^2}{[(1+\delta^{-1})\theta - \phi][1 + \phi^{-1}(1-\delta^{-1})\theta]}\left[\delta^{-1} - \theta^{-1}\phi,\ 1\right]'. \tag{2.17}$$

Thus, the baseline moment restrictions are nearly flat in θ because

$$\lim_{\phi \to 0} \mathbb{E}\left[\frac{d}{d\theta}\bar{g}_0(\theta)\right] = [0,\ 0]'. \tag{2.18}$$

However, under the reasonable calibrations in the literature (Bansal and Yaron, 2004; Beeler and Campbell, 2012; Bansal, Kiku, and Yaron, 2012), the preference parameters γ and ψ are well above 1, and thus the asset pricing moment condition has the unknown parameter θ well identified:

$$\lim_{\phi \to 0} \mathbb{E}\left[\frac{d}{d\theta}\bar{g}_1(\theta)\right] = -(2\gamma - \psi^{-1} - 1)(1 - \psi^{-1})\theta \neq 0. \tag{2.19}$$

3 Conditional Specification Test

Let $\Theta \in \mathbb{R}^{d_\theta}$ denote the parameter space that includes θ_0 as an interior point. We consider the incremental J statistic:

$$\mathcal{T} \equiv J - J_0, \quad \text{where } J \equiv \min_{\theta \in \Theta} g(\theta)'(\hat{\Omega}(\theta))^{-1} g(\theta) \quad \text{and} \quad J_0 \equiv \min_{\theta \in \Theta} g_0(\theta)'(\hat{\Omega}_0(\theta))^{-1} g_0(\theta), \tag{3.1}$$

with $g_0(\theta) \equiv n^{-1/2}\sum_{t=1}^n g_{0,t}(\theta) \in \mathbb{R}^{k_0}$, $g(\theta) \equiv n^{-1/2}\sum_{t=1}^n g_t(\theta) \in \mathbb{R}^k$, $g_t(\theta) \equiv [g_{0,t}(\theta)', g_{1,t}(\theta)']'$, $\hat{\Omega}(\theta) \equiv \hat{\Omega}(\theta, \theta)$, where $\hat{\Omega}(\theta, \tilde{\theta})$ is an estimator of $\Omega(\theta, \tilde{\theta}) \equiv \lim_{n\to\infty} \mathrm{Cov}(g(\theta), g(\tilde{\theta}))$ for any $\theta, \tilde{\theta} \in \Theta$ and $\hat{\Omega}_0(\theta)$ is the leading $k_0 \times k_0$ submatrix of $\hat{\Omega}(\theta)$. Note that by definitions, $g_0(\theta)$ and $g(\theta)$ rescale the sample averages $\bar{g}_0(\theta)$ and $\bar{g}(\theta)$ by $n^{1/2}$, respectively.

If the baseline moments provide strong identification of θ_0, $\mathcal{T} \to_d \chi^2_{k_1}$ and a critical value from this chi-square distribution yields the C test (incremental J test) of Eichenbaum, Hansen, and Singleton (1988). This test is more powerful than the standard over-identification test based on the J statistic because it exploits the validity of the baseline moments. When the baseline moments only provide weak identification, the chi-square distribution is no longer a good approximation for the distribution of $\mathcal{T}$.

We propose an alternative critical value based on the conditional inference approach. Following Andrews and Mikusheva (2016a), we view the rescaled baseline moment function $\mathbb{E}[g_0(\theta)] = n^{1/2}\mathbb{E}[\bar{g}_0(\theta)]$ indexed by θ as a functional nuisance parameter and obtain a simulation-based critical value by conditioning on a sufficient statistic for $\mathbb{E}[g_0(\theta)]$. Below, we describe the steps of constructing this critical value in detail.

Step 1. First, estimate θ_0 by the continuously updated estimator (CUE) as follows:

$$\hat{\theta} \equiv \arg\min_{\theta \in \Theta} g(\theta)'(\hat{\Omega}(\theta))^{-1} g(\theta). \tag{3.2}$$

This estimator is consistent under the null, following the standard arguments (e.g., Newey and Smith, 2004). Second, let $\hat{\Omega}(\theta, \tilde{\theta})$ be a consistent estimator of the covariance function $\Omega(\theta, \tilde{\theta})$ for any $\theta, \tilde{\theta} \in \Theta$. Specifically, $\hat{\Omega}(\theta, \tilde{\theta})$ can be a sample analog for i.i.d. data or be a heteroskedasticity and autocorrelation consistent (HAC) estimator for time series data. Given this covariance estimator, compute the test statistic $\mathcal{T}$ following (3.1). Third, let $\hat{\Omega} \equiv \hat{\Omega}(\hat{\theta}, \hat{\theta})$ and $\hat{Q} \equiv \partial\bar{g}(\hat{\theta})/\partial\theta'$.

Step 2. Construct a sufficient statistic $m(\theta)$ for the rescaled baseline moments $\mathbb{E}[g_0(\theta)]$. To this end, conduct the following decomposition:

$$g_0(\theta) = m(\theta) + V(\theta)g(\hat{\theta}), \quad \text{where } m(\theta) \equiv g_0(\theta) - V(\theta)g(\hat{\theta}) \tag{3.3}$$

is the residual process obtained by projecting $g_0(\theta)$ onto $g(\hat{\theta})$, and $V(\theta) \equiv S_0\Omega(\theta, \theta_0)\Omega^{-1}$ is the projection coefficient with $S_0 \equiv [I_{k_0}, 0_{k_0 \times k_1}]$ and $\Omega \equiv \Omega(\theta_0, \theta_0)$. Importantly, $m(\theta)$ is orthogonal to $g(\hat{\theta})$ by construction. In practice, we replace the unknown function $V(\theta)$ in the formula above with its estimator $\hat{V}(\theta) \equiv S_0\hat{\Omega}(\theta, \hat{\theta})\hat{\Omega}^{-1}$.

Here is a remark on why the residual process $m(\theta)$ is a sufficient statistic for the functional nuisance parameter $\mathbb{E}[g_0(\theta)]$ under the null hypothesis. Under the null, $g(\hat{\theta})$, the full moment function evaluated at $\hat{\theta}$, is approximately a linear function of $g(\theta_0)$:

$$g(\hat{\theta}) = \Omega^{1/2}M\upsilon + \varepsilon_n, \quad \text{where } \upsilon \equiv \Omega^{-1/2}g(\theta_0) \to_d N(0, I_k), \tag{3.4}$$

$M \equiv I_k - \Omega^{-1/2}Q(Q'\Omega^{-1}Q)^{-1}Q'\Omega^{-1/2}$ is a projection matrix, and ε_n is an error term that is either zero if $g_t(\theta)$ is linear in θ or negligible asymptotically if $g_t(\theta)$ is nonlinear. Because of the asymptotic normality of $g_0(\theta)$ and $g(\hat{\theta})$, the residual process $m(\theta)$ is a Gaussian process independent of $g(\hat{\theta})$ asymptotically. Conditioning on $m(\theta)$, the distribution of $g_0(\theta)$ and the test statistic $\mathcal{T}$ do not depend on the functional nuisance parameter $\mathbb{E}[g_0(\theta)]$, meaning that $m(\theta)$ is a sufficient statistic for $\mathbb{E}[g_0(\theta)]$. The sufficient statistic $m(\theta)$ captures the identification information of the baseline moments that is independent of randomness of the full moments in large samples. Intuitively, this is the information of the baseline moments about θ_0 decoupled from the underlying content of the asset pricing theory.

Step 3. Conditioning on the sufficient statistic $m(\theta)$ obtained in step 2, we approximate the conditional distribution of the test statistic $\mathcal{T}$ by replacing $g_0(\cdot)$ with the decomposition in (3.3) and replacing $g(\hat{\theta})$ with its asymptotic approximation $\Omega^{1/2}M\upsilon$ in (3.4), where υ is drawn from the standard k-dimensional multivariate normal distribution. Specifically, we define

$$L(\upsilon; d_0) \equiv \upsilon' M \upsilon - \min_{\theta \in \Theta} \left[m(\theta) + V(\theta)\Omega^{1/2}M\upsilon\right]' (\Omega_0(\theta))^{-1} \left[m(\theta) + V(\theta)\Omega^{1/2}M\upsilon\right], \tag{3.5}$$

where $d_0 \equiv (m(\cdot)', \text{vec}(V(\cdot))', \text{vech}(\Omega)', \text{vech}(\Omega_0(\cdot))', \text{vech}(M)')'$, and $\Omega_0(\theta)$ denotes the leading $k_0 \times k_0$ submatrix of $\Omega(\theta) \equiv \Omega(\theta, \theta)$. To simulate this conditional distribution, for $b = 1, \ldots, B$, take independent draws $\upsilon_b^* \sim N(0, I_k)$ and calculate $\mathcal{T}_b^* = L(\upsilon_b^*; \hat{d})$, Replacing d_0 in (3.5) by $\hat{d}$ means that $m(\cdot)$ is the residual process obtained in step 2, V is replaced by $\hat{V}(\theta)$ in step 2, Ω and $\Omega_0(\theta)$ are replaced by $\hat{\Omega}$ and $\hat{\Omega}_0(\theta) = S_0\hat{\Omega}(\theta)S_0'$ in step 1, and M is replaced by $\hat{M} \equiv I_k - \hat{\Omega}^{-1/2}\hat{Q}(\hat{Q}'\hat{\Omega}^{-1}\hat{Q})^{-1}\hat{Q}'\hat{\Omega}^{-1/2}$.

Step 4. Let α be the nominal size of the test and $b_0 \equiv \lceil (1-\alpha)B \rceil$ be the smallest integer larger than or equal to $(1-\alpha)B$. The critical value $c_{B,\alpha}(\hat{d})$ is the b_0^{th} smallest value among $\{\mathcal{T}_b^* : b = 1, \ldots, B\}$. Reject the null hypothesis in (2.1) if the statistic $\mathcal{T}$ in step 1 is larger than the critical value $c_{B,\alpha}(\hat{d})$. This completes the algorithm.

Besides θ, the moments may also depend on an unknown parameter ψ, which, unlike θ, cannot be strongly identified by the asset pricing moments. In this case, we can apply the algorithm to the joint hypothesis $H_0 : \mathbb{E}[\bar{g}_1(\theta_0, \psi_0)] = 0_{k_1 \times 1}$ and $\psi_0 = \psi_c$, with ψ_c imposed in all the moments as if it were known. The original null hypothesis $H_0 : \mathbb{E}[\bar{g}_1(\theta_0, \psi_0)] = 0_{k_1 \times 1}$ is rejected if the joint hypothesis is rejected for all null values ψ_c in its parameter space. This is a projection-based subvector inference method with the weakly identified nuisance parameter ψ. In practice, we may consider different values ψ_c in the range of calibrations used in the literature and treat ψ_c as part of the model's functional-form specification being tested.

Monte Carlo Simulation. Here we conduct Monte Carlo simulations to compare the finite-sample size and power of the proposed conditional specification test, the J test, and the C test in the two motivating examples in Section 2. The models, moments, parameters, and notations are as described in Section 2.

We first conduct a simulation study for the disaster risk model. We generate Δc_t following (2.1) and (2.2), and we generate r_t^e according to the following data-generating process:

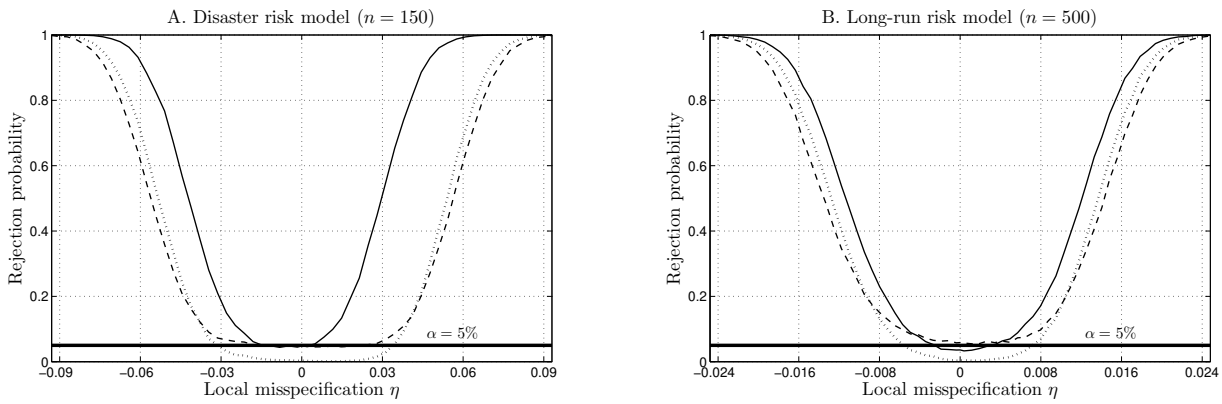
$$r_t^e = \eta + \gamma\sigma^2 - \frac{1}{2}\sigma^2 - p\mu_1(\alpha) + \frac{p}{\alpha - \gamma} h(\alpha) + \varepsilon_{d,t}^e, \tag{3.6}$$

where the error term $\varepsilon_{d,t}^e \equiv \sigma\varepsilon_t - [x_t(\underline{\nu} + J_t) - p\mu_1(\alpha)] + \sigma_d \varepsilon_{d,t}$, and the functions $\mu_1(\cdot)$ and $h(\cdot)$ are defined in (2.3) and (2.5), respectively.[5] Here, $\varepsilon_{d,t}$ is an i.i.d. standard normal variable capturing the measurement error and is independent of the other shocks. Under the null, $\eta = 0$, and under the alternative, we consider various values of η to compare the powers. The misspecification term η in (3.6) can be attributed to the missing risk factors or economic mechanisms if expected returns are not driven only by the disaster risk mechanism, and it can also be attributed to misspecified functional forms or restrictive parametric assumptions.

The parameters are set in the yearly frequency, similar to those set by Rietz (1988), Longstaff and Piazzesi (2004), and Wachter (2013), as follows: $n = 150$, $\sigma = 2\%$, $\underline{\nu} = 7\%$, $\gamma = 4$, $p = 0.5\%$, and $\sigma_d = 15\%$. In accordance with the observed equity premium and its sampling uncertainty in the data, we set $\theta = p/(\alpha - \gamma) = 0.0138$ as the calibrated "true" value in the simulation experiment to match the annual equity premium of 6%, and we estimate θ using the simulated data by searching

[5] See Section A of the note on additional materials for the derivation of (3.6) and (3.7).

Figure 1: A comparison of tests for the rare-disaster risk and long-run risk model.



Note: Panels A and B plot the rejection probabilities of three different specification tests for the disaster risk model and the long-run risk model, respectively, based on the simulated data. In both panels, the solid curve represents the rejection probability of the conditional specification test, with the test statistic $\mathcal{T}$ defined in (3.1) and the conditional critical value defined in (4.2); the dashed curve represents the rejection probability of the J test (Hansen, 1982); the dotted curve represents the rejection probability of the C test (Eichenbaum, Hansen, and Singleton, 1988); and the bold solid horizontal line represents the 5% nominal size for all three specification tests.

in the interval $[\underline{c}, \overline{c}]$, where the bounds $\underline{c}$ and $\overline{c}$ are calibrated to match the annual equity premium of 3% and 9%, respectively.

We next conduct a simulation study for the long-run risk model. We generate Δc_t following (2.10) and (2.11), and we generate r_t^e according to the following data-generating process:

$$r_t^e = \eta + \gamma\sigma_c^2 - \frac{1}{2}\sigma_c^2 + \frac{1}{2}\left(2\gamma - \psi^{-1} - 1\right)\left(1 - \psi^{-1}\right)\frac{\phi^2}{\left(\delta^{-1} - \rho\right)^2} + \varepsilon_{\ell,t}^e, \tag{3.7}$$

where the error term $\varepsilon_{\ell,t}^e \equiv \sigma_c\varepsilon_{c,t} + (1 - \psi^{-1})(\delta^{-1} - \rho)^{-1}\phi\varepsilon_{x,t} + \sigma_\ell\varepsilon_{\ell,t}$. Here, $\varepsilon_{\ell,t}$ is an i.i.d. standard normal variable capturing the measurement error and is independent of the other shocks. The misspecification term η in (3.7) has a similar interpretation to that of the disaster risk model example. The parameters are set in the quarterly frequency, close to those set by Bansal, Kiku, and Yaron (2012), as follows: $n = 500$, $\sigma_c = 0.0072 \times \sqrt{3}$, $\delta = 0.9989^3$, $\gamma = 10$, $\psi = 1.5$, $\rho = 0.975^3$, and $\sigma_\ell = 7.5\%$. In accordance with the observed equity premium and its sampling uncertainty in the data, we set $\theta = \phi/(\delta^{-1} - \rho) = 0.0665$ as the calibrated "true" value in the simulation experiment to match the annual equity premium of 6%, and we estimate θ using the simulated data by searching in the interval $[\underline{c}, \overline{c}]$, where the bounds $\underline{c}$ and $\overline{c}$ are calibrated to match the annual equity premium of 3% and 9%, respectively.

Lastly, we discuss the simulation results for both examples, which are based on 10000 simulation repetitions and $B = 2500$ random draws for the critical value of the conditional specification test in each repetition. Figure 1 reports the finite-sample rejection probabilities of the proposed conditional specification test, the J test, and the C test in the disaster risk model (panel A) and the long-run

risk model (panel B). The simulation confirms that (i) the size of the conditional specification test is at the nominal level, (ii) the power of the conditional specification test is higher than that of the J test, and (iii) the size of the C test deviates from the nominal level, severely under-rejecting the null hypothesis. In the note on additional materials, we show that these patterns are robust to a wider parameter space of θ for both models. It is worth pointing out that the C test may over-reject for a different data generating process. Figure 1 demonstrates that the proposed conditional specification test offers substantial improvement over the existing tests for these two prominent macro asset pricing models.

4 Theoretical Properties

4.1 Finite-Sample Size Control in Linear Gaussian Models

For the test statistic and the critical value in the algorithm, there are three types of approximation errors between the finite- and large-sample distributions: (i) the linear approximation error ε_n in (3.4), (ii) the Gaussian approximation error for the distribution of $m(\cdot)$ and υ, and (iii) the estimation errors in the consistent estimators of θ_0, $\Omega(\cdot,\cdot)$, $V(\cdot)$, and M. To abstract from these approximation errors, which all vanish asymptotically under the null, below we first consider a linear Gaussian statistical experiment where all types of errors are exactly zero even in finite samples. Let $\upsilon^* \equiv \Omega^{-1/2}\psi(\theta_0)$ and $m^*(\cdot) \equiv \mathbb{E}[g_0(\cdot)] + S_0\psi(\cdot) - V(\cdot)\,\Omega^{1/2}M\Omega^{-1/2}\psi(\theta_0)$ denote the Gaussian counterparts of υ and $m(\cdot)$, respectively, where $\psi(\cdot)$ is a Gaussian process with the covariance function $\Omega(\cdot,\cdot)$. In this linear Gaussian experiment, the test statistic $\mathcal{T}$ is exactly $L(\upsilon^*; d^*)$, where d^* is the same as d_0 except that $m(\cdot)$ in d_0 is replaced by $m^*(\cdot)$. Define its conditional $1-\alpha$ quantile as

$$c_\alpha^*(d^*) \equiv \inf\left\{c \in \mathbb{R} : P\left(L(\upsilon^*; d^*) > c \middle| d^*\right) \leq \alpha\right\}, \tag{4.1}$$

for nominal size α, where $P(\cdot|\, d^*)$ denotes the conditional distribution of $L(\upsilon^*; d^*)$ given d^*.

Lemma 1. *In a linear Gaussian experiment we have the following results under the null hypothesis:*
(i) $m^*(\cdot)$ *and* $M\upsilon^*$ *are independent;*
(ii) $P\left(L(\upsilon^*; d^*) > c_\alpha^*(d^*)\right) \leq \alpha$*;*
(iii) *If the conditional distribution of* $L(\upsilon^*; d^*)$ *given* d^* *is continuous at its* $1-\alpha$ *quantile almost surely, the size of the test equals the nominal level:* $P\left(L(\upsilon^*; d^*) > c_\alpha^*(d^*)\right) = \alpha$.

The critical value $c_\alpha^*(d^*)$ can be simulated using the marginal distribution of υ^* because of the following three reasons. First, υ^* enters $L(\upsilon^*; d^*)$ through $M\upsilon^*$. Second, $M\upsilon^*$ and $m^*(\cdot)$ are independent. Finally, $m^*(\cdot)$ is the only random component in d^*. In large samples, the simulated

critical value $c_{B,\alpha}(\hat{d})$ obtained in step 4 of the algorithm given in Section 3 approximates $c^*_\alpha(d^*)$ with high accuracy when B is a large number.

4.2 Asymptotic Uniform Validity for Nonlinear Models

We first state the assumptions that are used to derive the asymptotic size of the test. Let $\mathbb{P}$ denote the distribution of the data $\{Y_t\}_{t=1}^n$. We allow $\mathbb{P}$ to change with the sample size n but suppress this dependence for notational simplicity. We also suppress the dependence of $\mathbb{E}[\cdot]$ and θ_0 on $\mathbb{P}$. Let $\mathcal{P}$ denote a family of distributions for which the baseline moments are valid. Let $\mathcal{P}_0$ denote a subset of $\mathcal{P}$ consistent with the null hypothesis. Both $\mathcal{P}$ and $\mathcal{P}_0$ are allowed to change with n. Let $q(\theta) \equiv \partial\bar{g}(\theta)/\partial\theta'$, then $Q(\theta) = \mathbb{E}[q(\theta)]$. For $j = 1, \ldots, d_\theta$, let $Q_j(\theta)$ denote the jth column of $Q(\theta)$ and θ_j denote the jth component in θ. Let $\lambda_{min}(A)$ denote the minimal eigenvalue of a symmetric real matrix A, and $\|\cdot\|$ denote the matrix Frobenius norm.

Assumption 1. *The following conditions hold uniformly over $\mathbb{P} \in \mathcal{P}$:*
(i) $g(\cdot) - \mathbb{E}[g(\cdot)]$ *weakly converges to a mean-zero Gaussian process $\psi(\cdot)$ with covariance $\Omega(\cdot,\cdot)$;*
(ii) $\sup_{\theta\in\Theta} \|q(\theta) - Q(\theta)\| \to_p 0$ *and $Q(\theta)$ is continuous;*
(iii) $\sup_{\theta\in\Theta} [||\mathbb{E}[\bar{g}(\theta)]|| + ||Q(\theta)|| + \sum_{j=1}^{d_\theta} ||\partial Q_j(\theta)/\partial\theta'||] \leq C_m$ *for some finite constant C_m.*

Assumption 2. *The following conditions hold uniformly over $\mathbb{P} \in \mathcal{P}$:*
(i) *There exists an estimator $\hat{\Omega}(\cdot,\cdot)$ of $\Omega(\cdot,\cdot)$ such that $\sup_{\theta,\tilde{\theta}\in\Theta} ||\hat{\Omega}(\theta,\tilde{\theta}) - \Omega(\theta,\tilde{\theta})|| = o_p(1)$;*
(ii) *$\Omega(\theta,\tilde{\theta})$ is continuous uniformly over $(\theta,\tilde{\theta}) \in \Theta \times \Theta$;*
(iii) $\sup_{\theta,\tilde{\theta}\in\Theta} ||\partial\hat{\Omega}(\theta,\tilde{\theta})/\partial\theta_j - \partial\Omega(\theta,\tilde{\theta})/\partial\theta_j|| = o_p(1)$ *for $j = 1, \ldots, d_\theta$;*
(iv) $\sup_{\theta,\tilde{\theta}\in\Theta}[||\Omega(\theta,\tilde{\theta})|| + \sum_{j=1}^{d_\theta} ||\partial\Omega(\theta,\tilde{\theta})/\partial\theta_j||] \leq C_\Omega$ *for some finite constant C_Ω.*

Assumption 3. *The following conditions hold uniformly over $\mathbb{P} \in \mathcal{P}_0$:*
(i) *There exists $\theta_0 \in \Theta$ such that $\mathbb{E}[\bar{g}(\theta_0)] = 0_{k\times 1}$;*
(ii) *for any $\varepsilon > 0$, there exists a constant $\delta_\varepsilon > 0$ such that $\inf_{\theta\in B^c_\varepsilon(\theta_0)} ||\mathbb{E}[\bar{g}(\theta)]|| > \delta_\varepsilon$, where $B^c_\varepsilon(\theta) \equiv \{\tilde{\theta} \in \Theta : ||\tilde{\theta} - \theta|| \geq \varepsilon\}$;*
(iii) $\lambda_{\min}(Q'Q) \geq c_\lambda$ *and* $\inf_{\theta\in\Theta} \lambda_{\min}(\Omega(\theta)) \geq c_\lambda$ *for some positive constant c_λ.*

Assumption 1(i) requires that the moment is well approximated by a Gaussian limit. Its verification relies on a uniform central limit theorem, as discussed by Andrews and Mikusheva (2016a).[6]

[6] In our long-run risk example, the assumption of Gaussian approximation is innocuous even if the root of the latent autoregressive process could be arbitrarily close to unity, different from the classical near unit root analysis (e.g., Phillips, 1987; Mikusheva, 2007). In this example, the stationary component always dominates the latent non-stationary component because the loading on the latent process ϕ shrinks to 0 proportionally as $1 - \rho$ shrinks to 0 (see Section B of the note on additional materials for details). In the disaster risk model, Assumption 1(i) holds with a non-singular covariance matrix even though the disaster occurs with a small probability because the variance of the normally distributed regular shock dominates that of the disaster shock.

Assumption 1(ii) follows from the uniform law of large numbers. Assumption 1(iii) are standard regularity conditions on uniformly bounded moment functions and their derivatives.

Assumptions 2(i) and 2(iii) require that we have uniformly consistent estimators of the covariance function $\Omega(\cdot,\cdot)$ and its partial derivatives. Uniform consistency can be obtained by strengthening a pointwise consistent covariance matrix estimator with standard smoothness conditions. Assumptions 2(ii) and 2(iv) impose continuity and uniform upper bounds on the covariance function $\Omega(\cdot,\cdot)$ and its partial derivatives. Both Assumptions 1 and 2 are imposed on $\mathcal{P}$, not only on $\mathcal{P}_0$, because they are useful for both the size and power analyses of the proposed conditional test.

Assumption 3 is used to show consistency and asymptotic normality of $\hat{\theta}$ under the null hypothesis. Assumptions 3(i) and 3(ii) provide the identification uniqueness condition of the unknown parameter θ_0 using all valid moments under the null hypothesis. Assumption 3(iii) includes standard full rank conditions for the Jacobian matrix and the covariance matrix of the full moments $\bar{g}(\theta)$.

Let $\hat{d}$ be the analog of d, with $m(\cdot)$, $V(\cdot)$, Ω, $\Omega(\cdot,\cdot)$, and M all replaced by their consistent estimators, as in the practical algorithm. Given $\hat{d}$, we simulate independent draws $\upsilon^* \sim N(0, I_k)$ and obtain the critical value

$$c_\alpha(\hat{d}) \equiv \inf\left\{c \in \mathbb{R} : P^*(\upsilon^* \colon L(\upsilon^*; \hat{d}) > c) \leq \alpha\right\}, \tag{4.2}$$

where $P^*(\cdot)$ denotes the distribution of υ^*.

Theorem 1. *Suppose Assumptions 1, 2, and 3 hold. The test has correct asymptotic size, in the sense that, for any $\varepsilon > 0$,*

$$\limsup_{n\to\infty} \sup_{\mathbb{P}\in\mathcal{P}_0} \mathbb{P}\left(\mathcal{T} > c_\alpha(\hat{d}) + \varepsilon\right) \leq \alpha.$$

Theorem 1 implies that the conditional specification test controls the asymptotic size no matter whether the unknown parameter θ_0 (or its subvector) is strongly identified, weakly identified, or not identified by the baseline moments $\mathbb{E}\left[\bar{g}_0(\theta_0)\right] = 0_{k_0\times 1}$.[7]

Next, we consider the behavior of the test statistic $\mathcal{T}$ and the conditional critical value $c_\alpha(\hat{d})$ when $\mathbb{E}\left[\bar{g}_0(\theta_0)\right] = 0_{k_0\times 1}$ strongly identifies θ_0.

Assumption 4. *The following conditions hold uniformly over $\mathbb{P} \in \mathcal{P}_{00} \subset \mathcal{P}$: (i) for any $\varepsilon > 0$, there exists a constant $\delta_\varepsilon > 0$ such that $\inf_{\theta\in B^c_\varepsilon(\theta_0)} \|\mathbb{E}\left[\bar{g}_0(\theta)\right]\| > \delta_\varepsilon$; and (ii) $\lambda_{\min}\left(Q_0' Q_0\right) \geq c_\lambda$ where $Q_0 \equiv \mathbb{E}\left[\partial\bar{g}_0(\theta)/\partial\theta'\right]$.*

[7] Under additional regularity conditions on the continuity of the distribution function of the test statistic and the critical value, the test is also asymptotically similar, as discussed in Andrews and Mikusheva (2016a); see, e.g., Andrews, Cheng, and Guggenberger (2020) for discussions on asymptotic similarity.

Assumption 4 is similar to Assumption 3 and is imposed on $\mathbb{E}[g_t(\theta)]$ for the strong identification of θ_0 using all moments. This assumption is needed to show that the test statistic $\mathcal{T}$ converges to a chi-square distribution and the critical value $c_\alpha(\hat{d})$ converges to the $1-\alpha$ quantile of this chi-square distribution under strong identification.

Theorem 2. *Suppose Assumptions 1, 2, 3, and 4 hold. The following results hold uniformly over* $\mathcal{P}_0 \cap \mathcal{P}_{00}$*:* (i) $\mathcal{T} \to_d \chi^2_{k_1}$*; and* (ii) $c_\alpha(\hat{d}) \to_p q_{1-\alpha}(\chi^2_{k_1})$*, where* $q_{1-\alpha}(\chi^2_{k_1})$ *denotes the* $1-\alpha$ *quantile of a* $\chi^2_{k_1}$ *distribution.*

Theorem 2 shows that the conditional specification test is equivalent to the C test under the null when the baseline moments $\mathbb{E}[\bar{g}_0(\theta_0)] = 0_{k_0\times 1}$ provide strong identification of θ_0.

If the baseline moment conditions $\mathbb{E}[\bar{g}_0(\theta_0)] = 0_{k_0\times 1}$ only depend on a subvector $\theta_{c,0}$ of θ_0 with dimension d_c and strongly identify $\theta_{c,0}$, arguments analogous to those used to prove Theorem 2 also give

$$\mathcal{T} \to_d \chi^2_{k_1+d_c-d_\theta} \quad \text{and} \quad c_\alpha(\hat{d}) \to_p q_{1-\alpha}(\chi^2_{k_1+d_c-d_\theta}) \tag{4.3}$$

uniformly over $\mathcal{P}_0 \cap \mathcal{P}_{00}$. In this case, $k_1 \geq d_\theta - d_c$ in (4.3) because the asset pricing moments of dimension k_1 must strongly identify all the parameters not in the baseline moments.

In the presence of some additional parameter ψ that is only weakly-identified by the asset pricing moments, we can test the joint hypothesis $H_0 : \mathbb{E}[\bar{g}_1(\theta_0, \psi_0)] = 0_{k_1\times 1}$ and $\psi_0 = \psi_c$ for some null value ψ_c as discussed in Section 3. As long as Assumptions 1 to 4 hold with ψ_0 fixed at ψ_c, the results of Theorems 1 and 2 apply to the joint test. The projection-based subvector test $H_0 : \mathbb{E}[\bar{g}_1(\theta_0, \psi_0)] = 0_{k_1\times 1}$ also has correct asymptotic size. Since the projection-based test could be conservative, more efficient subvector tests are desirable.[8] Developing more efficient subvector tests for the present problem is beyond the scope of this paper.

In the supplemental appendix, we investigate the power of the conditional specification test. We prove that (i) the test is consistent when the asset pricing moments are globally misspecified regardless of the identification strength in the baseline moments, and (ii) the conditional test shares the power optimality of the C test in standard scenarios where the baseline moments provide strong identification. The optimal test in the presence of weakly identified baseline moments is beyond the scope of the paper. However, the literature has provided several encouraging power results for various conditional tests against the null hypothesis $H_0 : \theta = \theta_0$ (e.g., Andrews, Moreira, and Stock, 2006; Andrews and Mikusheva, 2016a, 2020) and we expect the conditional specification test to inherit these good properties. One may also apply the generic methods of Elliott, Müller, and

[8] See e.g., Guggenberger, Kleibergen, Mavroeidis, and Chen (2012); Andrews and Mikusheva (2016b); Guggenberger, Kleibergen, and Mavroeidis (2019); Kleibergen (2021) for recent research along this line.

Watson (2015) to evaluate the efficiency of an ad hoc test with correct size. In the supplemental appendix, we derive some power envelopes in a Gaussian experiment as in Section 4.1 based on the observations of $g(\hat{\theta})$, under various restrictions on the alternative and the test. These power envelopes are akin to those in Section 3.4 of Andrews and Mikusheva (2016a). Simulation studies in the supplemental appendix show that the power of the conditional specification test is close to that of the infeasible uniformly most powerful unbiased test in many cases, particularly when the number of baseline moments is large or when the baseline moments are strongly correlated with the asset pricing moments.

5 Empirical Application

In this section, we consider a full-blown time-varying rare-disaster risk model similar to that of Wachter (2013), a significant extension of the static rare-disaster risk model in Section 2. The model is extended in a few crucial aspects: (i) the probability of rare disasters is time varying; (ii) the representative agent has a recursive Epstein-Zin-Weil preference; (iii) the government bill is defaultable; and (iv) the corporate dividend is modeled as the levered consumption. The model is able to generate a sizeable equity premium, a low interest rate, a high volatility of equity returns, a low volatility of government bill returns, and predictable excess equity returns in different prediction horizons, without generating excessive volatility for the aggregate consumption and dividend.

We consider the time-varying disaster risk model for a real-data application because it has been one of the most influential macro-finance frameworks in the literature. For instance, the time-varying disaster risk mechanism has been used to explain important empirical patterns in macroeconomic quantities (e.g., Gourio, 2012), exchange rates and international capital flows (e.g., Martin, 2013; Dou and Verdelhan, 2017; Lewis and Liu, 2017), volatile unemployment flows (e.g., Kilic and Wachter, 2018), and prices of derivatives (e.g., Gabaix, 2012).

Model. We first describe the model. The log growth rate of consumption, $\Delta c_t \equiv \ln(C_t/C_{t-1})$, evolves as follows:

$$\Delta c_t = g_c + \sigma_c \varepsilon_{c,t} - \zeta_t, \tag{5.1}$$

where C_t is real consumption per capita at time t, ζ_t is the disaster variable, g_c is the average growth rate conditional on no disaster in the next period (i.e., $\zeta_t = 0$), and $\varepsilon_{c,t}$ is the normal consumption shock that follows a standard Gaussian distribution. The disaster variable ζ_t is characterized by

$$\zeta_t = x_t(\underline{v} + J_t), \tag{5.2}$$

where the constant $\underline{v}$ is the lower bound of the disaster size, the variable $J_t \sim \text{Exp}(\alpha)$ captures the disaster shock, and the variable $x_t = x_t^+ - x_t^-$ captures the occurrence of a jump, with x_t^+ to be a Bernoulli variable capturing the occurrence of a rare disaster with probability $\max(p_{t-1}, 0)$ and x_t^- to be a Bernoulli variable capturing the occurrence of a rare boom with probability $\max(-p_{t-1}, 0)$. Under the definition of x_t, the expectation of x_t is p_{t-1} conditioning on the information up to the end of period t. This jump probability index p_{t-1} evolves according to an AR(1) process:

$$\hat{p}_t = (1-\rho) + \rho\hat{p}_{t-1} + \sigma_p \varepsilon_{p,t}, \quad \text{with} \;\; \hat{p}_{t-1} \equiv p_{t-1}/p, \;\; \text{and } \rho, p \in (0,1). \tag{5.3}$$

Here, the shock $\varepsilon_{p,t}$ follows a standard Gaussian distribution. Thus, x_t follows a hidden Markovian process with the latent state variable p_{t-1}. The evolution in (5.3) says that the long-term average jump probability p_t is $\mathbb{E}[p_t] = p$ and the unconditional standard deviation of jump probability index p_{t-1} is $\mathbb{V}\text{ol}(p_{t-1}) = p\sqrt{\sigma_p^2/(1-\rho^2)}$. Similar in spirit to our specification of (5.3), Gourio (2012) assumes that the log transformation of the normalized jump probability index $\hat{p}_{t-1} = p_{t-1}/p$ evolves as an AR(1) process.[9]

We model the real dividend per capita D_t as the levered consumption with the log dividend growth $\Delta d_t \equiv \ln(D_t/D_{t-1})$ evolving as follows:

$$\Delta d_t = g_d + \phi\sigma_c\varepsilon_{c,t} - \phi\zeta_t, \tag{5.4}$$

similar in spirit to the works of Abel (1999) and Campbell (2003). Here, the constant g_d is the average growth rate conditional on no disaster in the next period (i.e., $\zeta_t = 0$). The shocks $(\varepsilon_{c,t}, \varepsilon_{p,t}, J_t)$ are mutually independent with each other and i.i.d. over t. The Bernoulli variables (x_t^+, x_t^-) are independent of the contemporaneous jump probability shock $\varepsilon_{p,t}$ and its leads in the time series, but (x_t^+, x_t^-) and the lags of $\varepsilon_{p,t}$ are dependent through the jump probability index p_{t-1}. The two processes (x_t^+, x_t^-) and $(\varepsilon_{c,t}, J_t)$ are mutually independent.

Consider the government bill with a one-period maturity. Like in the works of Barro (2006) and Wachter (2013), we assume that the government bill may default only when a disaster occurs. The return on the defaultable government bill can be expressed as

$$r_{b,t} = y_{b,t-1} - x_{b,t}(\underline{v} + J_t), \tag{5.5}$$

where the variable $y_{b,t-1}$ is the yield of the government bill, and the Bernoulli variable $x_{b,t} \in \{0,1\}$ characterizes the occurrence of a government bill default. The yield $y_{b,t-1}$ is observable in the data.

[9] The value of p_{t-1} can go outside the interval $[-1, 1]$ with a negligible chance under the relevant calibrations. A similar situation is encountered by Gourio (2012). We set this jump probability p_{t-1} to the nearest boundary once it goes outside $[-1, 1]$. We stick to the literature by assuming Gaussian shocks in (5.3) to ensure the (approximate) exponential-affine solution, summarized in (5.7) – (5.10).

The government bill defaults in period t if and only if $x_{b,t} = 1$. We assume that, in the event of disaster ($x_t = 1$), there will be a default on government liabilities with probability q. That is, $\mathbb{P}(x_{b,t} = 1 | x_t = 0 \text{ or } x_t = -1) = 0$ and $\mathbb{P}(x_{b,t} = 1 | x_t = 1) = q$. As reflected in (5.5), we follow Barro (2006) and Wachter (2013) in assuming that the percentage loss of the government bill is equal to the percentage decline in consumption in the event of default.

The representative agent has recursive preferences with unit EIS, and maximizes her utility V_t as follows:

$$\ln V_t = (1-\delta)\ln C_t + \delta(1-\gamma)^{-1}\ln \mathbb{E}_t\left[V_{t+1}^{1-\gamma}\right], \tag{5.6}$$

where δ is the rate of time preference, γ is the coefficient of risk aversion for timeless gambles.

Equilibrium. We solve the model using the Campbell-Shiller log-linearization approximation around the steady state, where p_{t-1} is close to $p > 0$. The equilibrium log return of the government bill, denoted by $r_{b,t}$, is

$$\begin{aligned} r_{b,t} - \mathbb{E}_{t-1}[r_{b,t}] &= -\left[x_{b,t}(\underline{v} + J_t) - q p_{t-1}\mu_1(\alpha)\right], \\ \text{with } \ \mathbb{E}_{t-1}[r_{b,t}] &= \omega_1(\vartheta) - q\mu_1(\alpha)(p_{t-1} - p) - (1-q)h_1(\alpha,\gamma)\frac{p_{t-1}-p}{\alpha-\gamma}, \end{aligned} \tag{5.7}$$

where $\mu_1(\alpha)$ and $\omega_1(\vartheta)$ are defined in (2.3) and (5.13), and $h_1(\alpha,\gamma) \equiv \alpha\left[e^{\underline{v}\gamma} - e^{\underline{v}(\gamma-1)}\frac{\alpha-\gamma}{\alpha-\gamma+1}\right]$.

The equilibrium excess log return of the equity over the government bill, denoted by $r^e_{m,t}$, is

$$\begin{aligned} r^e_{m,t} - \mathbb{E}_{t-1}\left[r^e_{m,t}\right] &= \phi\sigma_c\varepsilon_{c,t} + \beta_p\sigma_p\varepsilon_{p,t} - \left[(\phi x_t - x_{b,t})(\underline{v} + J_t) - (\phi - q)p_{t-1}\mu_1(\alpha)\right], \\ \text{with } \ \mathbb{E}_{t-1}\left[r^e_{m,t}\right] &= \omega_3(\vartheta) + h_3(\alpha,\gamma)\frac{p_{t-1}-p}{\alpha-\gamma}, \end{aligned} \tag{5.8}$$

where $\beta_p \equiv \frac{p\bar{\delta}}{1-\rho\bar{\delta}}h_2(\alpha,\gamma)$ with $\bar{\delta} \equiv \delta e^{(g_d - g_c)}$ and $\lambda_p \equiv -\frac{p}{\delta^{-1}-\rho}\left[e^{\underline{v}(\gamma-1)}\frac{\alpha}{\alpha-\gamma+1} - 1\right]$, and $\omega_3(\vartheta)$ is defined in (5.13). Here, $h_2(\alpha,\gamma)$ and $h_3(\alpha,\gamma)$ are defined as follows:

$$\begin{aligned} h_2(\alpha,\gamma) &\equiv \alpha\left[e^{\underline{v}(\gamma-\phi)}\frac{1}{\alpha-\gamma+\phi} - e^{\underline{v}(\gamma-1)}\frac{1}{\alpha-\gamma+1}\right], \\ h_3(\alpha,\gamma) &\equiv \alpha\left[(1-q)e^{\underline{v}\gamma} - e^{\underline{v}(\gamma-\phi)}\frac{\alpha-\gamma}{\alpha-\gamma+\phi} + q e^{\underline{v}(\gamma-1)}\frac{\alpha-\gamma}{\alpha-\gamma+1}\right] - (\alpha-\gamma)(\phi-q)\mu_1(\alpha). \end{aligned} \tag{5.9}$$

The equilibrium log price-dividend ratio, denoted by $z_{m,t}$, is

$$z_{m,t} = \overline{z}_m + h_2(\alpha,\gamma)\frac{p_t - p}{1-\rho\bar{\delta}}, \quad \text{where } \overline{z}_m \equiv \ln\left[\frac{\bar{\delta}}{1-\bar{\delta}}\right]. \tag{5.10}$$

Here, $\overline{z}_m$ is the long-run average log price-dividend ratio. To ensure the existence of the equilibrium, we require that $\bar{\delta} < 1$. As done by Barro (2009), we interpret $\bar{\delta}$ as the effective rate of time preference of the representative agent and require the effective rate of time preference to be less than 1. The detailed derivation of the equilibrium is relegated to the note on additional materials.

Moments. We consider a set of baseline moment conditions that summarize the key dynamic features of Δc_t and Δd_t specified in equations (5.1) – (5.4) as follows: $\mathbb{E}[\bar{m}_0(\vartheta)] = 0_{8\times 1}$, where $\bar{m}_0(\vartheta) = n^{-1}\sum_{t=1}^{n} m_{0,t}(\vartheta)$ with

$$m_{0,t}(\vartheta) \equiv \begin{bmatrix} (\Delta c_t - g_c) + p\mu_1(\alpha) \\ (\Delta d_t - g_d) + \phi p\mu_1(\alpha) \\ (\Delta c_t - g_c)^2 - \sigma_c^2 - p\mu_2(\alpha) \\ (\Delta d_t - g_d)^2 - \phi^2\sigma_c^2 - p\phi^2\mu_2(\alpha) \\ \Delta c_{t-1}\left[\Delta c_{t+1} - \rho\Delta c_t + (1-\rho)(p\mu_1(\alpha) - g_c)\right] \\ \Delta d_{t-1}\left[\Delta d_{t+1} - \rho\Delta d_t + (1-\rho)(\phi p\mu_1(\alpha) - g_d)\right] \\ \Delta d_{t-1}\left[\Delta c_{t+1} - \rho\Delta c_t + (1-\rho)(p\mu_1(\alpha) - g_c)\right] \\ \Delta c_{t-1}\left[\Delta d_{t+1} - \rho\Delta d_t + (1-\rho)(\phi p\mu_1(\alpha) - g_d)\right] \end{bmatrix}, \tag{5.11}$$

where $\mu_j(\alpha)$ is defined in (2.3). The baseline moments depend on (α, ρ), but not on (σ_p^2, γ).

We next consider the 6 asset pricing moment conditions targeted by Wachter (2013). The asset pricing moment conditions are $\mathbb{E}[\bar{m}_1(\vartheta)] = 0_{6\times 1}$, where $\bar{m}_1(\vartheta) = n^{-1}\sum_{t=1}^{n} m_{1,t}(\vartheta)$ with

$$m_{1,t}(\vartheta) \equiv \begin{bmatrix} r_{b,t} - \omega_1(\vartheta) \\ \left[r_{b,t} - \omega_1(\vartheta)\right]^2 - \omega_2(\vartheta) \\ r_{m,t}^e - \omega_3(\vartheta) \\ \left[r_{m,t}^e - \omega_3(\vartheta)\right]^2 - \omega_4(\vartheta) \\ r_{m,t-1}^e\left[r_{m,t}^e - \omega_5(\vartheta)(z_{m,t-1} - \overline{z}_m) - \omega_3(\vartheta)\right] \\ r_{m,t-1}^e\left[r_{m,t+1}^e - \omega_6(\vartheta)(z_{m,t-1} - \overline{z}_m) - \omega_3(\vartheta)\right] \end{bmatrix}, \tag{5.12}$$

and the deterministic functions $\omega_i(\vartheta)$ are described in detail as follows:

$$
\begin{aligned}
\omega_1(\vartheta) &\equiv -\ln\delta + g_c - \frac{1}{2}(2\gamma-1)\sigma_c^2 - qp\mu_1(\alpha) - (1-q)h_1(\alpha,\gamma)\frac{p}{\alpha-\gamma},\\
\omega_2(\vartheta) &\equiv qp\mu_2(\alpha) - q^2p^2\mu_1(\alpha)^2 + \left[2qp\mu_1(\alpha) + (1-q)h_1(\alpha,\gamma)\frac{p}{\alpha-\gamma}\right]\frac{(1-q)h_1(\alpha,\gamma)\sigma_p^2 p}{(1-\rho^2)(\alpha-\gamma)},\\
\omega_3(\vartheta) &\equiv \phi\gamma\sigma_c^2 + \beta_p\lambda_p\sigma_p^2 - \frac{1}{2}\left(\phi^2\sigma_c^2 + \beta_p^2\sigma_p^2\right) + h_3(\alpha,\gamma)\frac{p}{\alpha-\gamma},\\
\omega_4(\vartheta) &\equiv \phi^2\sigma_c^2 + \beta_p^2\sigma_p^2 + (q-2\phi q+\phi^2)p\mu_2(\alpha) - (q-\phi)^2\mu_1(\alpha)^2\left(p^2 + \frac{\sigma_p^2}{1-\rho^2}\right)\frac{h_3(\alpha,\gamma)^2\sigma_p^2p^2}{(1-\rho^2)(\alpha-\gamma)^2},\\
\omega_5(\vartheta) &\equiv \frac{(1-\rho\bar{\delta})}{\alpha-\gamma}h_2(\alpha,\gamma)^{-1}h_3(\alpha,\gamma), \quad \text{and} \quad \omega_6(\vartheta) \equiv \rho\omega_5(\vartheta).
\end{aligned}
\tag{5.13}
$$

The first two moments are about the low mean and low volatility of government bill returns, the third and fourth moments are about the high mean and high volatility of excess equity returns, and the last two moments are about the one- and two-period-ahead predictability of excess equity returns using lagged log price-dividend ratios. As demonstrated in many studies (e.g., Keim and Stambaugh, 1986; Campbell and Shiller, 1988), high price-dividend ratios predict low excess returns across various horizons. Importantly, Campbell and Yogo (2006) show that conventional tests of the predictability of stock returns can be invalid and lack power when the predictor variable is persistent and its innovations are highly correlated with returns.

We assume that the econometrician knows all parameters except $\vartheta = (\alpha, \rho, \sigma_p^2, \gamma)$, which governs the dynamics of time-varying rare-disaster risk and the agent's risk aversion. Other parameters $(g_c, g_d, \sigma_c^2, \phi, \underline{v}, p, q, \delta)$ are externally calibrated, and the key asset pricing implications are not sensitive to the local perturbations in these parameters (see Wachter, 2013; Chen, Dou, and Kogan, 2020, for sensitivity analysis). Specifically, we set $g_c = g_d = 0.02$, $\sigma_c^2 = 0.02^2$, $\phi = 3.5$, $\underline{v} = 0.07$, $q = 0.4$, and $\delta = 0.97$, which lie within the ballpark of the calibrations used in the literature (e.g., Bansal and Yaron, 2004; Longstaff and Piazzesi, 2004; Wachter, 2013). We consider multiple values of $p \in \{0.3\%, 0.5\%, 0.7\%, 0.9\%, 1.1\%\}$ to focus on rare disasters following the calibrations adopted by Rietz (1988) and Longstaff and Piazzesi (2004), which are consistent with the structural estimation result from the observed equity index option prices in Backus, Chernov, and Martin (2011). The calibrated parameter values are effectively part of the functional form of the model under the examination of the specification tests, similar to Julliard and Ghosh (2012) who test the rare events hypothesis using the generalized empirical likelihood methods.

Reparametrization. The asset pricing moments in (5.12) clearly demonstrate the key idea of the time-varying disaster risk model to simultaneously explain the sizeable equity risk premium and high equity volatility: when p, $\alpha - \gamma$, σ_p^2, and $1 - \rho^2$ are all close to 0, the rare yet severe

disaster can generate a substantial equity premium and large equity volatility as long as the two ratios $\frac{p}{\alpha-\gamma}$ and $\frac{\sigma_p^2}{1-\rho^2}$ are sizable to match the equity premium and the volatility of equity excess returns. This ensures that the time-varying disaster risk is a meaningful economic mechanism for explaining the equity premium and volatility even if p is very small. To utilize this key insight, we transform the parameters α and ρ to θ_1 and θ_2, respectively, with

$$\theta_1 \equiv \frac{p}{\alpha-\gamma}, \quad \theta_2 \equiv \frac{\sigma_p^2}{1-\rho^2}, \quad \theta_3 \equiv \rho, \quad \text{and} \quad \theta_4 \equiv \gamma, \tag{5.14}$$

with the stacked parameter vector $\theta \in \Theta \equiv \prod_{i=1}^4 \Theta_i$. Our analysis allows p, $\alpha-\gamma$, σ_p^2, and $1-\rho^2$ to be all close to 0, while keeping the ratios θ_1 and θ_2 bounded from above and below. We refer to θ_1 and θ_2 as the adjusted disaster size parameter and the adjusted disaster probability volatility parameter, respectively. To reparameterize all the moments from ϑ into θ, write

$$\bar{g}_i(\theta) \equiv \bar{m}_i(\theta_1^{-1}p + \theta_4, \theta_2(1-\theta_3^2), \theta_3, \theta_4), \quad \text{with } i \in \{0,1\}. \tag{5.15}$$

The asset pricing moments provide an intuitive identification structure of θ. The first and third moments on the average of (excess) returns mainly identify the adjusted disaster size parameter θ_1. The second and fourth moments on the variance of (excess) returns mainly identify the adjusted disaster probability volatility parameter θ_2 and the risk aversion parameter θ_4. The last two moments on the predictability of excess equity returns identify the persistence parameter of time-varying disaster probability θ_3.

U.S. Data, Robust Evaluations, and Model Uncertainty Sets. Based on the annual U.S. data of consumptions, dividends, government bill returns, equity returns, and log price-dividend ratios, we compare the J test and the proposed conditional specification test, then contrast the model uncertainty sets constructed based on the two specification tests.

Ideally, a reliable empirical analysis of the time-varying rare-disaster risk model should be based on the longest possible sample. As a result, we construct a set of long time series (1871 - 2019), with the data obtained from various sources. To construct the time series of log real consumption growth rates, we use the Barro-Ursua Macroeconomic Data for 1871 - 2009 and the per-capita real personal consumption expenditure on services and nondurable goods from the National Income and Product Accounts (NIPA) for 2010 - 2019. To construct the time series of log price-dividend ratios, real log dividend growth rates, and log market returns, we obtain the data from Campbell (2003) and Robert Shiller's website for 1871 - 2012, and the Center for Research in Security Prices (CRSP) S&P Index data for 2013 - 2019. For log real returns of treasury bills, we obtain the data from Campbell (2003) and Robert Shiller's website for 1871 - 2012, and the CPI-deflated 1-year treasury bill rates from the federal reserve data program (H15) for 2013 - 2019.

Table 1: Specification test results for time-varying rare-disaster risk models with different calibrated long-run average disaster probability p.

	P-values of tests		Point estimates			
Calibrated p	J test	Cond. test	$\underline{v} + \alpha^{-1}$	$p \times \sigma_p$	ρ	γ
$p = 0.3\%$	0.120	0.012	0.321	0.205%	0.979	3.610
$p = 0.5\%$	0.400	0.088	0.279	0.140%	0.994	4.230
$p = 0.7\%$	0.434	0.102	0.242	0.115%	0.998	4.950
$p = 0.9\%$	0.411	0.089	0.224	0.101%	0.999	5.350
$p = 1.1\%$	0.254	0.030	0.212	0.116%	0.999	5.670

Note: We report the point estimates of $\underline{v}+\alpha^{-1}$ and $p\times\sigma_p$, instead of α and σ_p^2, because of the direct economic interpretations: $\underline{v}+\alpha^{-1}$ is the average disaster size in (5.1) and $p \times \sigma_p$ is the volatility of the jump probability index p_t in (5.3). The GMM estimators are computed using the continuous-updating estimator proposed by Hansen, Heaton, and Yaron (1996).

Table 1 presents the results on the J test, the proposed conditional specification test, and the point estimation. [10] In the benchmark calibration with $p = 0.7\%$, the time-varying rare-disaster risk model can easily fit the U.S. data, closely matching both the baseline and asset pricing moment conditions through the lens of the J test (with the p-value equal to 0.434). The proposed conditional specification test can improve the test power by efficiently utilizing the limited information of the baseline moment conditions, with the p-value dropping substantially from 0.434 to 0.102. Yet, it is interesting and reassuring to see that the time-varying rare-disaster risk model remains statistically consistent with the U.S. data at the 5% level, even under the stringent and robust examination of the proposed conditional specification test. When increasing (or decreasing) the long-run average disaster probability p from 0.7% to 0.9% (or 0.5%), the test and estimation results remain nearly unchanged. By contrast, when disasters occur very rarely with $p = 0.3\%$ (or fairly frequently with $p = 1.1\%$), the time-varying rare-disaster risk model is statistically rejected at the 5% level by the proposed conditional specification test based on the U.S. data with a p-value of 0.012 (or a p-value of 0.030), although it is still largely accepted according to the J test with a p-value of 0.120 (or a p-value of 0.254). This result manifests the importance of robust procedures in evaluating macro asset pricing models, and our econometric analysis here addresses the irrefutability concern of economic mechanisms relying on extremely rare disasters (e.g., Campbell, 2018; Chen, Dou,

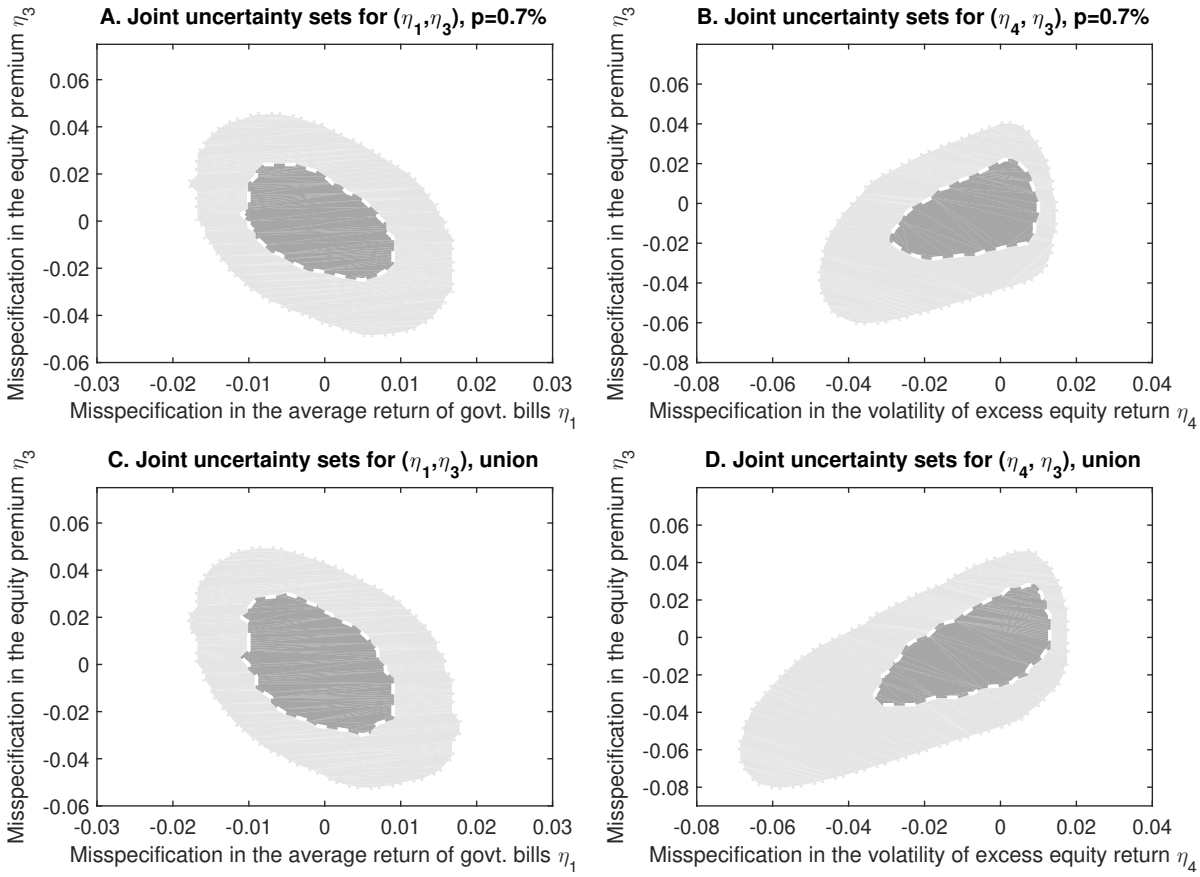
[10] For the results reported in Table 1, we set $\Theta_1 \equiv [0.001, 0.020]$, $\Theta_2 \equiv [5, 12]$, $\Theta_3 \equiv [0.950, 0.999]$ and $\Theta_4 \equiv [3, 6]$, and conduct parameter estimation by grid search with the step size 0.001, 0.01, 0.001 and 0.01, respectively. For the results reported in Figure 2 below, which is more computationally intensive, we consider a smaller parameter space $\Theta_2 \equiv [5, 8]$, which still covers the CUE estimator of θ_2 in all of the relevant cases considered in Figure 2, a larger step size 0.1 of the grid points in Θ_2 and Θ_4, and maintain the rest of the parameter space specification. For a summary of the empirical implementation details, see Section SF of the supplemental appendix.

and Kogan, 2020). Moreover, this result also echoes the quantitative study of Wachter (2013), showing that, unless conditioning on no disaster (i.e., the U.S. has been very lucky over the past centuries), the simulated data based on a time-varying disaster risk model with a fairly high disaster probability has a difficult time simultaneously matching all the baseline and asset pricing moment conditions in (5.11) and (5.12). Last but not least, our econometric analysis in Table 1 shows that the time-varying rare-disaster risk models indeed provide a potential explanation for the prominent asset pricing puzzles because, for each of the three non-rejected case, the estimated risk aversion parameter γ is less than 10, the upper bound of the "reasonable" range for γ in the macroeconomics and asset pricing literature (e.g., Campbell, 2003). There has been little formal econometric analysis on (time-varying) rare-disaster risk mechanisms in the asset pricing literature. As a contribution, the test and estimation results in Table 1 fill this gap.

Figure 2 shows that the robust specification test can serve as a powerful tool for constructing the model uncertainty sets. The model uncertainty set consists of the moment misspecification parameter η such that $H_0 : \mathbb{E}\left[\bar{g}_1(\theta_0)\right] = \eta$ cannot be rejected by a given specification test, with the asset pricing moments $\bar{g}_1(\theta)$ defined in (5.15). In fact, the specification tests in Table 1 correspond to the null of $\eta = 0$. Similar to the intuitive interpretation of Hansen and Sargent (2001), the estimated model is viewed as an approximation of the true model, lying within a collection of alternative probabilistic models whose fit of the moment conditions is statistically close to the estimated model.

For computational feasibility and economic interpretability, Figure 2 reports pairwise joint model uncertainty sets where only two elements of the vector η deviate from 0 for each model uncertainty set construction. Panel A shows that the joint model uncertainty set shrinks substantially when using the proposed conditional specification test. Specifically, the model uncertainty set shrinks by about 2 and 4 percentage points along the dimensions of the average government bill return and the equity risk premium, respectively, which are comparable to the the average interest rate and equity premium themselves in the data. Moreover, panel B shows that the model uncertainty set shrinks by about 3 percentage points along the dimension of the variance of equity excess returns, and the magnitude of the change is comparable to that of the variance of equity excess return itself in the data. In terms of the volume of the model uncertainty sets, the one constructed by the conditional specification test is about 40% of that by the J test. To further account for the uncertainty in the probability of rare disasters p, panels C and D display the unions of uncertainty sets for $p = 0.5\%,\ 0.7\%$, and 0.9%, the three cases where the asset pricing moments with $\eta = 0$ are not rejected in Table 1. The model uncertainty sets naturally become larger when accounting for uncertainty of p. Nevertheless, in both cases, the uncertainty sets based on the conditional specifi-

Figure 2: Joint model uncertainty sets based on the time-varying rare-disaster risk models.



Note: Panel A plots the joint model uncertainty set for the average return of government bills (η_1) and the equity premium (η_3) by focusing on $\eta = [\eta_1, 0, \eta_3, 0, 0, 0]'$. Panel B plots the joint model uncertainty set for the volatility of excess equity return (η_4) and the equity premium (η_3) by focusing on $\eta = [0, 0, \eta_3, \eta_4, 0, 0]'$. Panels C and D are analogous to A and B, respectively, except that each uncertainty set is the union of those obtained under $p = 0.5\%,\ 0.7\%,\ 0.9\%$. The bigger "nearly-ellipses-shaped" areas are the joint model uncertainty sets constructed using the J test, while the smaller darker "nearly-ellipses-shaped" areas are those constructed using the proposed conditional specification test. All the sets are calculated under the 95% confidence level. We focus on these two joint model uncertainty sets because interest rates, excess equity returns, and equity return volatilities are the most important quantities in macro asset pricing theories.

cation test remain substantially smaller than those based on the J test. Crucially, the data-driven joint model uncertainty sets on the mean and the variance of asset returns displayed in Figure 2 play a pivotal role in robust mean-variance portfolio analysis (e.g., Garlappi, Uppal, and Wang, 2007).

6 Conclusion

This paper provides a robust and powerful test to evaluate macro asset pricing models. The newly proposed conditional specification test gains power by exploiting valid but noisy information in

weakly identified baseline moments. To achieve robustness under weak identification, the conditional specification test decouples the useful macroeconomic information embedded in the baseline moment conditions from the additional asset pricing moment conditions. Our novel approach is particularly useful when the standard over-identification tests suffer from distorted size or poor power due to information imbalance. It can help researchers, practitioners, and monetary authorities to better understand the economic mechanisms behind the influential macro-finance models and conduct robust econometric analysis accounting for model uncertainty.

Appendix: Proofs

Throughout the proofs, we use K to denote a positive constant that may change from line to line. For any $x \in \mathbb{R}^{k_0}$ and any $k_0 \times k_0$ symmetric positive definite matrix A, $\|x\|_A \equiv (x'A^{-1}x)^{1/2}$. Let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and the largest eigenvalues of a real symmetric matrix A, respectively. The proofs of all auxiliary Lemmas, i.e., Lemmas A1 – A7 below, are given in the supplemental appendix.

Proof of Lemma 1. Since $M\upsilon^*$ and $m^*(\cdot)$ are mean-zero Gaussian, part (i) follows from $E\left[m^*(\cdot)(M\upsilon^*)'\right] = 0$. For part (ii), by the law of iterated expectation and the definition of $c^*_\alpha(d^*)$,

$$P\left(L(\upsilon^*, d^*) > c^*_\alpha(d^*)\right) = E\left[P\left(L(\upsilon^*, d^*) > c^*_\alpha(d^*)|d^*\right)\right] \leq \alpha. \tag{A.1}$$

Part (iii) follows from $P\left(L(\upsilon^*, d^*) > c^*_\alpha(d^*)|d^*\right) = \alpha$ under the specified continuity condition. *Q.E.D.*

The following results hold for the CUE $\hat{\theta}$ in (3.2) and $g(\hat{\theta})$, $\hat{\Omega}$, $\hat{M}$, and $\hat{V}(\theta)$ based on $\hat{\theta}$, regardless of the identification strength in the baseline moments.

Lemma A1. *Under Assumptions 1, 2 and 3, the following results hold uniformly over $\mathbb{P} \in \mathcal{P}_0$:*
(a) $n^{1/2}(\hat{\theta} - \theta_0) = -\left(Q'\Omega^{-1}Q\right)^{-1} Q'\Omega^{-1}g(\theta_0) + o_p(1) = O_p(1)$;
(b) $g(\hat{\theta}) = \Omega^{1/2}M\Omega^{-1/2}g(\theta_0) + o_p(1) = O_p(1)$;
(c) $\hat{\Omega} = \Omega + o_p(1)$ *where* $\hat{\Omega} \equiv \hat{\Omega}(\hat{\theta})$;
(d) $\hat{M} = M + o_p(1)$;
(e) $\sup_{\theta\in\Theta} ||\hat{V}(\theta) - V(\theta)|| = o_p(1)$ *where* $\sup_{\theta\in\Theta} \|V(\theta)\| \leq c_\lambda^{-1}C_\Omega$.

We next present a few lemmas used in the proof of Theorem 1. For any $x \in \mathbb{R}^k$, continuous vector function $m_d : \Theta \mapsto \mathbb{R}^{k_0}$, continuous matrix function $V_d : \Theta \mapsto \mathbb{R}^{k_0\times k}$, $k \times k$ symmetric

positive definite matrix Ω_d, symmetric and continuous matrix function $\Omega_{0,d}(\cdot) : \Theta \mapsto \mathbb{R}^{k_0 \times k_0}$ which is positive definite for any $\theta \in \Theta$, and $k \times k$ symmetric idempotent matrix M_d, let

$$\xi \equiv (x', d')', \text{ where } d \equiv (m_d(\cdot)', \text{vec}(V_d(\cdot))', \text{vech}(\Omega_d)', \text{vech}(\Omega_{0,d}(\cdot))', \text{vech}(M_d)')'.$$

Define

$$R(\xi) \equiv \|x\|^2_{\Omega_d} - \min_{\theta \in \Theta} \|m_d(\theta) + V_d(\theta)x\|^2_{\Omega_{0,d}(\theta)}, \tag{A.2}$$

and

$$L(\upsilon; d) \equiv \upsilon' M_d \upsilon - \min_{\theta \in \Theta} \left\| m_d(\theta) + V_d(\theta)\Omega_d^{1/2} M_d \upsilon \right\|^2_{\Omega_{0,d}(\theta)}. \tag{A.3}$$

The test statistic $\mathcal{T}$ in (3.1) can be written as

$$\mathcal{T} = R(\hat{\xi}), \text{ where } \hat{\xi} \equiv (g(\hat{\theta})', \hat{d}')' \text{ and}$$
$$\hat{d} \equiv (\hat{m}(\cdot)', \text{vec}(\hat{V}(\cdot))', \text{vech}(\hat{\Omega})', \text{vech}(\hat{\Omega}_0(\cdot))', \text{vech}(\hat{M})')'. \tag{A.4}$$

Given $\hat{d}$, the critical value $c_\alpha(\hat{d})$ is simulated using $L(\upsilon^*; \hat{d})$ with independent draws of $\upsilon^* \sim N(0, I_k)$. To show the bounded Lipschitz properties of functionals of ξ, we use the metric

$$\|\xi\|_s = \|x\| + \sup_{\theta \in \Theta} \|m_d(\theta)\| + \sup_{\theta \in \Theta} \|V_d(\theta)\| + \|\Omega_d\| + \sup_{\theta \in \Theta} \|\Omega_{0,d}(\theta)\| + \|M_d\|. \tag{A.5}$$

Lemma A2. *Under Assumptions 1, 2 and 3,*

$$\lim_{n \to \infty} \sup_{\mathbb{P} \in \mathcal{P}_0} \sup_{f \in BL_1} \left\| \mathbb{E}[f(\hat{\xi})] - E\left[f(\xi^*)\right] \right\| = 0,$$

where $\xi^* \equiv ((\Omega^{1/2} M \upsilon^*)', d^{*\prime})'$ and BL_1 denotes the set of functionals with Lipschitz constant and supremum norm bounded above by 1.

To use the weak convergence of $\hat{\xi}$ for studying the statistic $\mathcal{T}$, we follow Andrews and Mikusheva (2016a) and define a truncated version of $R(\xi)$ as

$$R_C(\xi) \equiv R(\xi) t_C(x' \Omega^{-1} x) \tag{A.6}$$

where $t_C(u) \equiv I\{u < C\} + (2C - u)C^{-1} I\{C \le u < 2C\}$ for any $u \in \mathbb{R}$ and some $C \ge 1$. Similarly, to study the critical value $c_\alpha(\hat{d})$, we define a truncated version of $L(\upsilon; d)$ as

$$\overline{L}_C(\upsilon; d) \equiv L_C(\upsilon; d) I\{\|\upsilon\|^2 \le C\}, \text{ where } L_C(\upsilon; d) \equiv L(\upsilon; d) t_C(\upsilon' M_d \upsilon). \tag{A.7}$$

Compared with $R_C(\xi)$, the truncation in $\overline{L}_C(\upsilon; d)$ has an extra term $I\{\|\upsilon\|^2 \le C\}$, which is needed to show that $\overline{L}_C(\upsilon; d)$ is Lipschitz in M_d. Since M_d may not have full rank, the truncation with $t_C(\upsilon' M_d \upsilon)$ is insufficient to bound $\|\upsilon\|$. Thus, truncation with $\|\upsilon\|^2 \le C$ is added in $\overline{L}_C(\upsilon; d)$.

Lemma A3. *Suppose that $\hat{\Omega}$ is symmetric and positive definite and $\hat{\Omega}_0$ is the leading $k_0 \times k_0$ submatrix of $\hat{\Omega}$. Then $R(\hat{\xi}) \geq 0$. Moreover, if $\hat{Q}'\hat{Q}$ is nonsingular, $L(\upsilon; \hat{d}) \geq 0$ for any $\upsilon \in \mathbb{R}^k$.*

Lemma A4. *Given $R(\xi) \geq 0$, the functional $R_C(\xi)$ is bounded and Lipschitz in ξ.*

Lemma A5. *Let $c_{\alpha,C}(d) \equiv \inf\left\{c : P^*\left(\upsilon^* \colon \overline{L}_C(\upsilon^*; d) > c\right) \leq \alpha\right\}$. Given $L(\upsilon; d) \geq 0$, $c_{\alpha,C}(d)$ is bounded and Lipschitz in d.*

The extra truncation $\|\upsilon\|^2 \leq C$ in $\overline{L}_C(\upsilon; d)$ causes a discrepancy between $c_{\alpha,C}(d^*)$ and the conditional $1-\alpha$ quantile of $R_C(\xi^*)$ given d^*. Lemma A6 below shows that we can choose C large enough such that the discrepancy is negligible, which is one of the key elements to show the uniform size control of the conditional specification test.

Lemma A6. *For any $\varepsilon \in (0,1)$ and any $\delta > 0$, there is a finite constant C_δ such that for any $C \geq C_\delta$: $P\left(R_C(\xi^*) > c_{\alpha,C}(d^*) + \varepsilon\right) \leq \alpha + \delta/4$.*

Proof of Theorem 1. The proof strategy follows from that for Theorem 1 of Andrews and Mikusheva (2016a). The major differences are as follows. (i) The test statistic and the critical value are defined with different functions, $R(\hat{\xi})$ and $L(\upsilon^*; \hat{d})$, respectively. These two functions have to be truncated differently too, as in (A.6) and (A.7), respectively, to yield the bounded Lipschitz property. (ii) The additional truncation to $L(\upsilon^*; \hat{d})$ causes a discrepancy between $c_{\alpha,C}(d^*)$ and the conditional $1-\alpha$ quantile of $R_C(\xi^*)$ given d^*. Lemma A6 is used to address these problems.

For notational simplicity, we assume that $\inf_{\theta\in\Theta} \lambda_{\min}(\hat{\Omega}(\theta)) \geq K^{-1}$, $\lambda_{\min}(\hat{Q}'\hat{Q}) \geq K^{-1}$ and $\sup_{\theta\in\Theta} \lambda_{\max}(\hat{\Omega}(\theta)) \leq K$ in the proof. This assumption is innocuous since the above properties hold with probability approaching 1 (wpa1) in view of Assumptions 1(ii), 2(i, iv) and 3(iii), and the consistency of $\hat{\theta}$ under the null. Suppose that the claim of the theorem does not hold. Then

$$\lim_{n\to\infty} \sup_{\mathbb{P}\in\mathcal{P}_0} \mathbb{P}\left(R(\hat{\xi}) > c_\alpha(\hat{d}) + \varepsilon\right) > \alpha, \tag{A.8}$$

which implies that there exists $\delta > 0$ and a divergent sequence n_i (indexed by i) such that

$$\mathbb{P}_{n_i}\left(R(\hat{\xi}) > c_\alpha(\hat{d}) + \varepsilon\right) > \alpha + \delta \text{ for all } i. \tag{A.9}$$

For any $u \in \mathbb{R}$ and any i, by the union bound of probability,

$$\mathbb{P}_{n_i}\left(R(\hat{\xi}) > u\right) \leq \mathbb{P}_{n_i}\left(R(\hat{\xi}) > u,\, g(\hat{\theta})'(\hat{\Omega}^{-1})g(\hat{\theta}) \leq C\right) + \mathbb{P}_{n_i}\left(g(\hat{\theta})'(\hat{\Omega}^{-1})g(\hat{\theta}) > C\right). \tag{A.10}$$

By the definition of $\hat{\theta}$, $g(\hat{\theta})'(\hat{\Omega}^{-1})g(\hat{\theta}) \leq g(\theta_0)'(\hat{\Omega}(\theta_0))^{-1}g(\theta_0)$ which together with Assumptions 1(i), 2(i) and 3 implies that $g(\hat{\theta})'(\hat{\Omega}^{-1})g(\hat{\theta}) = O_p(1)$ uniformly over $\mathbb{P} \in \mathcal{P}_0$. Therefore, there exists a large constant $C_{1,\delta}$ such that for all large n_i,

$$\mathbb{P}_{n_i}\left(g(\hat{\theta})'(\hat{\Omega}^{-1})g(\hat{\theta}) > C_{1,\delta}\right) \leq \delta/4, \tag{A.11}$$

which together with (A.9) and (A.10) implies that

$$\mathbb{P}_{n_i}\left(R(\hat{\xi}) > c_\alpha(\hat{d}) + \varepsilon,\, g(\hat{\theta})'(\hat{\Omega}^{-1})g(\hat{\theta}) \le C\right) > \alpha + 3\delta/4, \tag{A.12}$$

for any $C \ge C_{1,\delta}$. By definition,

$$I\left\{R_C(\hat{\xi}) > u\right\} \ge I\left\{R(\hat{\xi}) > u\right\} I\left\{g(\hat{\theta})'(\hat{\Omega}^{-1})g(\hat{\theta}) \le C\right\} \text{ for any } u \in \mathbb{R}, \tag{A.13}$$

where $R_C(\hat{\xi}) \equiv R(\hat{\xi})t_C(g(\hat{\theta})'(\hat{\Omega}^{-1})g(\hat{\theta}))$ and $t_C(u) = 1$ for $u \le C$ following its definition. By (A.12) and (A.13), we have for any $C \ge C_{1,\delta}$,

$$\mathbb{P}_{n_i}\left(R_C(\hat{\xi}) > c_\alpha(\hat{d}) + \varepsilon\right) > \alpha + 3\delta/4. \tag{A.14}$$

Since $L(\upsilon, \hat{d}) \ge 0$ for any $\upsilon \in \mathbb{R}^k$ by Lemma A3 and $t_C(u) \le 1$ for any $u \in \mathbb{R}$, we have $L_C(\upsilon, \hat{d}) \le L(\upsilon, \hat{d})$ for any $\upsilon \in \mathbb{R}^k$, which further implies that $c_{\alpha,C}(\hat{d}) \le c_\alpha(\hat{d})$. Therefore, by (A.14) we deduce that for any $C \ge C_{1,\delta}$,

$$\mathbb{P}_{n_i}\left(R_C(\hat{\xi}) - c_{\alpha,C}(\hat{d}) \ge \varepsilon\right) > \alpha + 3\delta/4. \tag{A.15}$$

Let $U_{C,n}$ be a random variable which has the same distribution as $R_C(\hat{\xi}) - c_{\alpha,C}(\hat{d})$ under the law $\mathbb{P}_n$. Let $U_{\infty,C,n}$ be a random variable which has the same distribution as $R_C(\xi^*) - c_{\alpha,C}(d^*)$. By Lemma A4 and Lemma A5, $R_C(\xi) - c_{\alpha,C}(d)$ is bounded and Lipschitz in ξ. Therefore, by Lemma A2,

$$\lim_{n\to\infty} \sup_{f \in BL_1} \|\mathbb{E}\left[f(U_{C,n})\right] - E\left[f(U_{\infty,C,n})\right]\| = 0. \tag{A.16}$$

Since $U_{\infty,C,n}$ is bounded for any n, by Prokhorov's theorem, there exists a subsequence n_j (of n_i) and a random variable U_C such that $U_{\infty,C,n_j} \to_d U_C$, which together with (A.16) implies that $U_{C,n_j} \to_d U_C$. Since (A.15) can be written as $\mathbb{P}_{n_i}(U_{C,n_i} \ge \varepsilon) > \alpha + 3\delta/4$, by Portmanteau theorem,

$$\begin{aligned}\liminf_{n_j\to\infty} P\left(U_{\infty,C,n_j} > \varepsilon/2\right) &\ge P\left(U_C > \varepsilon/2\right) \ge P\left(U_C \ge \varepsilon\right) \\ &\ge \limsup_{n_j\to\infty} \mathbb{P}_{n_j}\left(U_{C,n_j} \ge \varepsilon\right) \ge \alpha + 3\delta/4, \quad \text{for any } C \ge C_{1,\delta}.\end{aligned} \tag{A.17}$$

We next show that for all large C, $P\left(U_{\infty,C,n_j} > \varepsilon/2\right) \le \alpha + \delta/4$ for any n_j, which contradicts (A.17), and hence the claim of the theorem holds. To this end, for $C \ge C_{2,\delta}$ in Lemma A6,

$$P\left(U_{\infty,C,n_j} > \varepsilon/2\right) = P\left(R_C(\xi^*) > c_{\alpha,C}(d^*) + \varepsilon/2\right) \le \alpha + \delta/4, \tag{A.18}$$

where the equality holds because U_{∞,C,n_j} and $R_C(\xi^*) - c_{\alpha,C}(d^*)$ have the same distribution and the inequality follows from Lemma A6. *Q.E.D.*

Let $\hat{\theta}^* \equiv \arg\min_{\theta\in\Theta} \|\hat{m}(\theta) + \hat{V}(\theta)\hat{\Omega}^{1/2}\hat{M}\upsilon^*\|^2_{\hat{\Omega}_0(\theta)}$ and $\tilde{M}_0 \equiv (\Omega_0^{-1/2}S_0\Omega^{1/2})'M_0(\Omega_0^{-1/2}S_0\Omega^{1/2})$.

Lemma A7. *Under Assumptions 1, 2, 3 and 4, we have uniformly over $\mathbb{P} \in \mathcal{P}_0 \cap \mathcal{P}_{00}$:*
(a) $n^{1/2}(\hat{\theta}^* - \theta_0) = -(Q'\Omega^{-1}Q)^{-1}Q'\Omega^{-1}g(\theta_0) - \left(Q_0'\Omega_0^{-1}Q_0\right)^{-1} Q_0'\Omega_0^{-1}S_0\Omega^{1/2}M\upsilon^* + o_p(1)$;
(b) $L(\upsilon^*, \hat{d}) = \upsilon^{*\prime}(M - \tilde{M}_0)\upsilon^* + o_p(1)$;
(c) $\upsilon^{*\prime}(M - \tilde{M}_0)\upsilon^* \sim \chi^2_{k_1}$.

Proof of Theorem 2. (i) Under Assumptions 1 – 3, Lemma A1 gives

$$g(\hat{\theta}) = \Omega^{1/2}M\Omega^{-1/2}g(\theta_0) + o_p(1) \quad \text{and} \quad \hat{\Omega} \equiv \hat{\Omega}(\hat{\theta}) = \Omega + o_p(1), \tag{A.19}$$

uniformly over $\mathbb{P} \in \mathcal{P}_0$. Let $\hat{\theta}_0 \equiv \arg\min_{\theta\in\Theta} g_0(\theta)'(\hat{\Omega}_0(\theta))^{-1}g_0(\theta)$. Adding Assumption 4, we have

$$g_0(\hat{\theta}_0) = \Omega_0^{1/2}M_0\Omega_0^{-1/2}g_0(\theta_0) + o_p(1) \quad \text{and} \quad \hat{\Omega}_0 \equiv \hat{\Omega}_0(\hat{\theta}_0) = \Omega_0 + o_p(1) \tag{A.20}$$

uniformly over $\mathbb{P} \in \mathcal{P}_0 \cap \mathcal{P}_{00}$, where $M_0 \equiv I_{k_0} - \Omega_0^{-1/2}Q_0(Q_0'\Omega_0^{-1}Q_0)^{-1}Q_0'\Omega_0^{-1/2\prime}$. Therefore, $\mathcal{T} \to_d \chi^2_{k_1}$ uniformly over $\mathbb{P} \in \mathcal{P}_0$ by the standard arguments in the literature (e.g., Eichenbaum, Hansen, and Singleton, 1988; Hall, 2005, Section 5).

We next prove part (ii). The critical value is simulated from

$$L(\upsilon^*, \hat{d}) = \upsilon^{*\prime}\hat{M}\upsilon^* - \left\|\hat{m}(\hat{\theta}^*) + \hat{V}(\hat{\theta}^*)\hat{\Omega}^{1/2}\hat{M}\upsilon^*\right\|^2_{\hat{\Omega}_0(\hat{\theta}^*)}. \tag{A.21}$$

By Lemma A7(b, c), we have uniformly over $\mathbb{P} \in \mathcal{P}_0 \cap \mathcal{P}_{00}$,

$$L(\upsilon^*, \hat{d}) = L^* + o_p(1), \text{ where } L^* \equiv \upsilon^{*\prime}(M - \tilde{M}_0)\upsilon^* \sim \chi^2_{k_1}. \tag{A.22}$$

By (A.22), there exists a positive sequence $\delta_n = o(1)$ such that for any $\varepsilon > 0$,

$$\mathbb{P}^*\left(|L(\upsilon^*, \hat{d}) - L^*| \geq \varepsilon/2\right) = o(\delta_n), \text{ uniformly over } \mathbb{P} \in \mathcal{P}_0 \cap \mathcal{P}_{00}, \tag{A.23}$$

where $\mathbb{P}^* \equiv P^* \otimes \mathbb{P}$ denotes the product measure of υ^* and the data. Due to the independence between P^* and $\mathbb{P}$, for any $\varepsilon > 0$ and for all large n,

$$\mathbb{P}^*\left(|L(\upsilon^*, \hat{d}) - L^*| \geq \varepsilon/2 \middle| \hat{d}\right) \leq \delta_n \text{ wpa1}. \tag{A.24}$$

Note that $c_\alpha(\hat{d})$ is the $1-\alpha$ conditional quantile of $L(\upsilon^*, \hat{d})$ given $\hat{d}$ and $L^* \sim \chi^2_{k_1}$ is independent of $\hat{d}$. Therefore, (A.24) implies

$$q_{1-\alpha-\delta_n}(\chi^2_{k_1}) - \varepsilon/2 \leq c_\alpha(\hat{d}) \leq q_{1-\alpha+\delta_n}(\chi^2_{k_1}) + \varepsilon/2 \text{ wpa1} \tag{A.25}$$

because by (A.24) and the union bound of (conditional) probability, we have

$$\mathbb{P}^*\left(L(\upsilon^*, \hat{d}) > q_{1-\alpha+\delta_n}(\chi^2_{k_1}) + \varepsilon/2 \middle| \hat{d}\right) \leq \mathbb{P}^*\left(L^* > q_{1-\alpha+\delta_n}(\chi^2_{k_1}) \middle| \hat{d}\right) + \delta_n = \alpha, \tag{A.26}$$
$$\mathbb{P}^*\left(L^* > c_\alpha(\hat{d}) + \varepsilon/2 \middle| \hat{d}\right) \leq \mathbb{P}^*\left(L(\upsilon^*, \hat{d}) > c_\alpha(\hat{d}) \middle| \hat{d}\right) + \delta_n \leq \alpha + \delta_n.$$

Since $\delta_n = o(1)$ and $\chi^2_{k_1}$ is continuous with a strictly increasing quantile function, for all large n,

$$q_{1-\alpha+\delta_n}(\chi^2_{k_1}) - \varepsilon/2 \leq q_{1-\alpha}(\chi^2_{k_1}) \leq q_{1-\alpha-\delta_n}(\chi^2_{k_1}) + \varepsilon/2, \tag{A.27}$$

which together with (A.25) implies that, for any $\varepsilon > 0$, $|c_\alpha(\hat{d}) - q_{1-\alpha}(\chi^2_{k_1})| \leq \varepsilon$ wpa1. *Q.E.D.*

References

Abel, A. B. (1999): "Risk Premia and Term Premia in General Equilibrium," *Journal of Monetary Economics*, 43(1), 3 – 33.

Anatolyev, S., and A. Mikusheva (2020): "Factor Models with Many Assets: Strong Factors, Weak Factors, and the Two-pass Procedure," *Journal of Econometrics*, Forthcoming.

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

Andrews, D. W. K., X. Cheng, and P. Guggenberger (2020): "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," *The Journal of Econometrics*, Forthcoming.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74(3), 715–752.

Andrews, D. W. K., and J. H. Stock (2007): "Inference with Weak Instruments," in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, vol. 3 of *Econometric Society Monographs*, pp. 122–173. Cambridge University Press.

Andrews, I., and A. Mikusheva (2016a): "Conditional Inference with a Functional Nuisance Parameter," *Econometrica*, 84(4), 1571–1612.

——— (2016b): "A Geometric Approach to Nonlinear Econometric Models," *Econometrica*, 84(3), 1249–1264.

——— (2020): "Optimal Decision Rules for Weak GMM," Working papers, Harvard and MIT.

Andrews, I., J. H. Stock, and L. Sun (2019): "Weak Instruments in IV Regression: Theory and Practice," *Annual Review of Economics*, 11, 727–753.

Ascari, G., L. M. Magnusson, and S. Mavroeidis (2019): "Empirical Evidence on the Euler Equation For Consumption in the Us," *Journal of Monetary Economics*.

Backus, D., M. Chernov, and I. Martin (2011): "Disasters Implied by Equity Index Options," *Journal of Finance*, 66(6), 1969–2012.

Bansal, R., D. Kiku, and A. Yaron (2012): "An Empirical Evaluation of the Long-Run Risks Model for Asset Prices," *Critical Finance Review*, 1(1), 183–221.

Bansal, R., and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59(4), 1481–1509.

Barro, R. J. (2006): "Rare Disasters and Asset Markets in the Twentieth Century," *Quarterly Journal of Economics*, 121(3), 823–866.

——— (2009): "Rare Disasters, Asset Prices, and Welfare Costs," *American Economic Review*, 99(1), 243–64.

Beeler, J., and J. Y. Campbell (2012): "The Long-Run Risks Model and Aggregate Asset Prices: An Empirical Assessment," *Critical Finance Review*, 1(1), 141–182.

Brunnermeier, M., and Y. Sannikov (2016): "Chapter 18 - Macro, Money, and Finance: A Continuous-Time Approach," vol. 2 of *Handbook of Macroeconomics*, pp. 1497 – 1545. Elsevier.

Campbell, J. Y. (2003): "Consumption-based Asset Pricing," in *Handbook of the Economics of Finance*, ed. by G. Constantinides, M. Harris, and R. M. Stulz, vol. 1, Part 2, chap. 13, pp. 803–887. Elsevier.

——— (2018): *Financial Decisions and Markets: A Course in Asset Pricing*. Princeton University Press.

Campbell, J. Y., and J. H. Cochrane (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107(2), 205–251.

Campbell, J. Y., and R. J. Shiller (1988): "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1(3), 195–228.

Campbell, J. Y., and M. Yogo (2006): "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics*, 81(1), 27–60.

Chen, H., W. W. Dou, and L. Kogan (2020): "Measuring the 'Dark Matter' in Asset Pricing Models," *Journal of Finance*, Forthcoming.

Cheng, X., W. W. Dou, and Z. Liao (2021): "A Note on Additional Materials for 'Macro-Finance Decoupling: Robust Evaluations of Macro Asset Pricing Models'," *working paper, http://ssrn.com/abstract=3787125, also available on the authors' personal websites.*

Cochrane, J. H. (2017): "Macro-Finance," *Review of Finance*, 21(3), 945–985.

Dou, W., and A. Verdelhan (2017): "The Volatility of International Capital Flows and Foreign Assets," Unpublished working paper, MIT.

Dou, W. W., X. Fang, A. W. Lo, and H. Uhlig (2020): "Macro-Finance Models with Nonlinear Dynamics," Working papers, Becker Friedman Institute for Economics at University of Chicago.

Eichenbaum, M. S., L. P. Hansen, and K. J. Singleton (1988): "A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty," *Quarterly Journal of Economics*, 103(1), 51–78.

Elliott, G., U. K. Müller, and M. W. Watson (2015): "Nearly Optimal Tests when a Nuisance Parameter Is Present under the Null Hypothesis," *Econometrica*, 83(2), 771–811.

Elliott, G., and J. H. Stock (1994): "Inference in Time Series Regression When the Order of Integration of a Regressor Is Unknown," *Econometric Theory*, 10(3/4), 672–700.

Gabaix, X. (2012): "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance," *Quarterly Journal of Economics*, 127(2), 645–700.

Garlappi, L., R. Uppal, and T. Wang (2007): "Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach," *Review of Financial Studies*, 20(1), 41–81.

Gospodinov, N., R. Kan, and C. Robotti (2017): "Spurious Inference in Reduced-Rank Asset-Pricing Models," *Econometrica*, 85(5), 1613–1628.

Gourio, F. (2012): "Disaster Risk and Business Cycles," *American Economic Review*, 102(6), 2734–66.

Guggenberger, P., F. Kleibergen, and S. Mavroeidis (2019): "A More Powerful Subvector Anderson Rubin Test In Linear Instrumental Variables Regression," *Quantitative Economics*, 10(2), 487–526.

Guggenberger, P., F. Kleibergen, S. Mavroeidis, and L. Chen (2012): "On the Asymptotic Sizes of Subset Anderson - Rubin and Lagrange Multiplier Tests in Linear Instrumental Variables Regression," *Econometrica*, 80(6), 2649–2666.

Hahn, J., J. C. Ham, and H. R. Moon (2011): "The Hausman Test and Weak Instruments," *Journal of Econometrics*, 160(2), 289 – 299.

Hall, A. R. (2005): *Generalized Method of Moments*. Oxford university press.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50(4), 1029–54.

Hansen, L. P., J. Heaton, and A. Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14(3), 262–80.

Hansen, L. P., J. C. Heaton, and N. Li (2008): "Consumption Strikes Back? Measuring Long-run Risk," *Journal of Political Economy*, 116(2), 260–302.

Hansen, L. P., and T. J. Sargent (2001): "Robust Control and Model Uncertainty," *American Economic Review*, 91(2), 60–66.

He, Z., and A. Krishnamurthy (2018): "Intermediary Asset Pricing and the Financial Crisis," *Annual Review of Financial Economics*, 10(1), 173–197.

Julliard, C., and A. Ghosh (2012): "Can Rare Events Explain the Equity Premium Puzzle?," *Review of Financial Studies*, 25(10), 3037–3076.

Kan, R., and C. Zhang (1999): "GMM Tests of Stochastic Discount Factor Models with Useless Factors," *Journal of Financial Economics*, 54(1), 103 – 127.

Kandel, S., and R. F. Stambaugh (1991): "Asset Returns and Intertemporal Preferences," *Journal of Monetary Economics*, 27(1), 39–71.

Keim, D. B., and R. F. Stambaugh (1986): "Predicting Returns in The Stock and Bond Markets," *Journal of Financial Economics*, 17(2), 357 – 390.

Kilic, M., and J. A. Wachter (2018): "Risk, Unemployment, and the Stock Market: A Rare-Event-Based Explanation of Labor Market Volatility," *The Review of Financial Studies*, 31(12), 4762–4814.

Kleibergen, F. (2005): "Testing Parameters in GMM without Assuming that They Are Identified," *Econometrica*, 73(4), 1103–1123.

——— (2009): "Tests Of Risk Premia In Linear Factor Models," *Journal of Econometrics*, 149(2), 149 – 173.

——— (2021): "Efficient Size Correct Subset Inference In Homoskedastic Linear Instrumental Variables Regression," *Journal of Econometrics*, 221(1), 78–96.

Kleibergen, F., and Z. Zhan (2020): "Robust Inference for Consumption-Based Asset Pricing," *Journal of Finance*, 75(1), 507–550.

Lewis, K. K., and E. X. Liu (2017): "Disaster Risk and Asset Returns: an International Perspective," *Journal of International Economics*, 108, S42 – S58.

Longstaff, F. A., and M. Piazzesi (2004): "Corporate Earnings And The Equity Premium," *Journal of Financial Economics*, 74(3), 401–421.

Martin, I. (2013): "The Lucas Orchard," *Econometrica*, 81(1), 55–111.

Mikusheva, A. (2007): "Uniform Inference in Autoregressive Models," *Econometrica*, 75(5), 1411–1452.

——— (2013): "Survey On Statistical Inferences in Weakly-identified Instrumental Variable Models," *Applied Econometrics*, 29(1), 117–131.

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

Müller, U. K., and M. W. Watson (2008): "Testing Models of Low-Frequency Variability," *Econometrica*, 76(5), 979–1016.

——— (2018): "Long-Run Covariability," *Econometrica*, 86(3), 775–804.

Newey, W. K. (1985): "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29(3), 229–256.

Newey, W. K., and R. J. Smith (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72(1), 219–255.

Phillips, P. C. B. (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74(3), 535–547.

Rietz, T. A. (1988): "The Equity Premium: A Solution," *Journal of Monetary Economics*, 22(1), 117–131.

Schorfheide, F., D. Song, and A. Yaron (2018): "Identifying Long-Run Risks: A Bayesian Mixed-Frequency Approach," *Econometrica*, 86(2), 617–654.

Stambaugh, R. F. (1999): "Predictive Regressions," *Journal of Financial Economics*, 54(3), 375 – 421.

Stock, J., J. Wright, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20, 518 – 529.

Stock, J. H., and J. H. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68(5), 1055–1096.

Wachter, J. A. (2013): "Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?," *Journal of Finance*, 68(3), 987–1035.

Yogo, M. (2004): "Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak," *Review of Economics and Statistics*, 86(3), 797–810.

Supplemental Appendix to Macro-Finance Decoupling: Robust Evaluations of Macro Asset Pricing Models

Xu Cheng*, Winston Wei Dou†, Zhipeng Liao‡

August 23, 2021

Abstract

This supplemental appendix provides the following supporting materials. Sections SA – SC provide the proofs of Lemmas A1 – A7 in the appendix to the main text Cheng, Dou, and Liao (2021). Section SA provides the proofs of several lemmas on the asymptotic convergence of the random components in the test statistic $\mathcal{T}$ and the conditional critical value $c_\alpha(\hat{d})$. Section SB verifies the bounded Lipschitz properties of the test statistic and the conditional critical value, which are used to show their weak convergence in large samples. Section SC includes some auxiliary lemmas. Section SD provides additional theoretical results on the power of the proposed conditional test. Section SE provides comparison with some power envelopes through simulations. Section SF collects details and additional results of the empirical application.

*Department of Economics, University of Pennsylvania; Email: xucheng@econ.upenn.edu.
†Finance Department, The Wharton School, University of Pennsylvania; Email: wdou@wharton.upenn.edu.
‡Department of Economics, University of California, Los Angeles; Email: zhipeng.liao@econ.ucla.edu.

SA Proofs for Asymptotic Convergence

This section provides the proofs of Lemmas A1, A2, A6 and A7 in the appendix to Cheng, Dou, and Liao (2021) under their Assumptions 1, 2, 3 and 4. The proofs of Lemmas A3, A4 and A5 in the appendix to Cheng, Dou, and Liao (2021) are in Sections SB and SC of this supplemental appendix. Throughout this supplemental appendix, we use $\lambda_{min}(A)$ to denote the minimal eigenvalue of a real symmetric matrix A, and $\|\cdot\|$ denotes the matrix Frobenius norm.

Proof of Lemma A1. (a) Define $C_{1,n} \equiv \sup_{\theta\in\Theta} |\bar{g}(\theta)'(\hat{\Omega}(\theta))^{-1}\bar{g}(\theta) - G(\theta)'(\Omega(\theta))^{-1}G(\theta)|$ where $G(\theta) \equiv \mathbb{E}[\bar{g}(\theta)]$. Then by Assumptions 1(i, iii), 2(i) and 3(iii),

$$C_{1,n} = o_p(1), \tag{SA.1}$$

uniformly over $\mathbb{P} \in \mathcal{P}_0$. Consider any $\varepsilon > 0$. By the definition of $\hat{\theta}$ and θ_0,

$$\begin{aligned}
&\sup_{\mathbb{P}\in\mathcal{P}_0} \mathbb{P}\left(||\hat{\theta} - \theta_0|| \geq \varepsilon\right) \\
\leq &\sup_{\mathbb{P}\in\mathcal{P}_0} \mathbb{P}\left(\min_{\theta\in B_\varepsilon^c(\theta_0)} \bar{g}(\theta)'(\hat{\Omega}(\theta))^{-1}\bar{g}(\theta) \leq \bar{g}(\theta_0)'(\hat{\Omega}(\theta_0))^{-1}\bar{g}(\theta_0)\right) \\
\leq &\sup_{\mathbb{P}\in\mathcal{P}_0} \mathbb{P}\left(\min_{\theta\in B_\varepsilon^c(\theta_0)} G(\theta)'(\Omega(\theta))^{-1}G(\theta) \leq 2C_{1,n}\right) \leq \sup_{\mathbb{P}\in\mathcal{P}_0} \mathbb{P}\left(c_\lambda^{-1}\delta_\varepsilon^2 \leq 2C_{1,n}\right),
\end{aligned} \tag{SA.2}$$

where the second inequality is by the definition of $C_{1,n}$ and the third inequality is by Assumption 3. Combining the results in (SA.1) and (SA.2), we deduce that

$$\lim_{n\to\infty} \sup_{\mathbb{P}\in\mathcal{P}_0} \mathbb{P}\left(||\hat{\theta} - \theta_0|| \geq \varepsilon\right) = 0 \text{ for any } \varepsilon > 0. \tag{SA.3}$$

Let $C_{2,n} \equiv \sup_{\theta\in\Theta} \|q(\theta) - Q(\theta)\|$. Then by Assumption 1(ii),

$$C_{2,n} = o_p(1) \text{ uniformly over } \mathbb{P} \in \mathcal{P}. \tag{SA.4}$$

Applying the first order expansion, we get

$$g(\hat{\theta}) = g(\theta_0) + q(\tilde{\theta})n^{1/2}(\hat{\theta} - \theta_0), \tag{SA.5}$$

where $\tilde{\theta}$ is the mean value between θ_0 and $\hat{\theta}$ and it may vary across rows. By Assumption 1(ii), the consistency of $\hat{\theta}$ and (SA.4),

$$q(\tilde{\theta}) = Q(\tilde{\theta}) + o_p(1) = Q + o_p(1) = O_p(1), \text{ uniformly over } \mathbb{P} \in \mathcal{P}_0. \tag{SA.6}$$

Similarly, we can show that

$$q(\hat{\theta}) = Q(\hat{\theta}) + o_p(1) = Q + o_p(1) = O_p(1), \text{ uniformly over } \mathbb{P} \in \mathcal{P}_0. \tag{SA.7}$$

By Assumption 2(i, ii) and the consistency of $\hat{\theta}$,

$$\hat{\Omega} \equiv \hat{\Omega}(\hat{\theta}) = \Omega + o_p(1) \tag{SA.8}$$

uniformly over $\mathbb{P} \in \mathcal{P}_0$. Applying the chain rule, we get the first order condition of $\hat{\theta}$:

$$0_{d_\theta \times 1} = 2q(\hat{\theta})'\hat{\Omega}^{-1}g(\hat{\theta}) - \left(n^{-1/2} g(\hat{\theta})'\hat{\Omega}^{-1} \frac{\partial \hat{\Omega}(\hat{\theta})}{\partial \theta_j} \hat{\Omega}^{-1} g(\hat{\theta}) \right)_{j=1,\ldots,d_\theta}, \tag{SA.9}$$

where $(a_j)_{j=1,\ldots,d_\theta} \equiv (a_1, \ldots, a_{d_\theta})'$ for any real numbers $a_1, \ldots, a_{d_\theta}$. By Assumptions 1(i, iii), 2(iv) and 3(iii), the consistency of $\hat{\theta}$, (SA.5), (SA.6), (SA.7) and (SA.8),

$$\begin{aligned} q(\hat{\theta})'\hat{\Omega}^{-1}g(\hat{\theta}) &= q(\hat{\theta})'\hat{\Omega}^{-1}g(\theta_0) + q(\hat{\theta})'\hat{\Omega}^{-1}q(\tilde{\theta})n^{1/2}(\hat{\theta} - \theta_0) \\ &= Q'\Omega^{-1}g(\theta_0) + \left(Q'\Omega^{-1}Q + o_p(1)\right) n^{1/2}(\hat{\theta} - \theta_0) + o_p(1) \end{aligned} \tag{SA.10}$$

uniformly over $\mathbb{P} \in \mathcal{P}_0$. Similarly, we can show that for $j = 1, \ldots, d_\theta$,

$$n^{-1/2} g(\hat{\theta})'\hat{\Omega}^{-1} \frac{\partial \hat{\Omega}(\hat{\theta})}{\partial \theta_j} \hat{\Omega}^{-1} g(\hat{\theta}) = n^{1/2}(\hat{\theta} - \theta_0) o_p(1) + o_p(1) \text{ uniformly over } \mathbb{P} \in \mathcal{P}_0. \tag{SA.11}$$

Combining the results in (SA.9), (SA.10), (SA.11), and applying Assumptions 1(iii) and 3(iii), we deduce that

$$n^{1/2}(\hat{\theta} - \theta_0) = -(Q'\Omega^{-1}Q)^{-1}Q'\Omega^{-1}g(\theta_0) + o_p(1) = O_p(1) \tag{SA.12}$$

uniformly over $\mathbb{P} \in \mathcal{P}_0$, which proves the first claim of the lemma.

(b) This claim follows by Assumptions 1 and 3(iii), (SA.5), (SA.6) and (SA.12).

(c) This claim has been proved in (SA.8).

(d) By Assumptions 2(iv) and 3(iii), and (SA.8),

$$\lim_{n\to\infty} \inf_{\mathbb{P}\in\mathcal{P}_0} \mathbb{P}\left(K^{-1} \leq \lambda_{\min}(\hat{\Omega}) \leq \lambda_{\max}(\hat{\Omega}) \leq K \right) = 1. \tag{SA.13}$$

By Assumptions 1(iii) and 3(iii), (SA.7) and (SA.8), we have

$$q(\hat{\theta})'\hat{\Omega}^{-1}q(\hat{\theta}) = Q'\Omega^{-1}Q + o_p(1) \text{ uniformly over } \mathbb{P} \in \mathcal{P}_0, \tag{SA.14}$$

which together with Assumption 1(iii), 2(iv), and 3(iii) implies that

$$\lim_{n\to\infty} \inf_{\mathbb{P}\in\mathcal{P}_0} \mathbb{P}\left(K^{-1} \leq \lambda_{\min}(q(\hat{\theta})'(\hat{\Omega})^{-1}q(\hat{\theta})) \leq \lambda_{\max}(q(\hat{\theta})'(\hat{\Omega})^{-1}q(\hat{\theta})) \leq K\right) = 1. \tag{SA.15}$$

Let $\|\cdot\|_S$ denote the matrix operator norm. By Exercise 7.2.18 in Horn and Johnson (1990),

$$\left\|\hat{\Omega}^{1/2} - \Omega^{1/2}\right\|_S \leq ||\hat{\Omega} - \Omega||_S \left\|\Omega^{-1}\right\|_S, \tag{SA.16}$$

which together with Assumption 3(iii) and (SA.8) implies that

$$\left\|\hat{\Omega}^{1/2} - \Omega^{1/2}\right\|_S = o_p(1) \text{ uniformly over } \mathbb{P} \in \mathcal{P}_0. \tag{SA.17}$$

By (SA.17) and the relation between the operator norm and the Frobenius norm,

$$\left\|\hat{\Omega}^{1/2} - \Omega^{1/2}\right\| = o_p(1) \text{ uniformly over } \mathbb{P} \in \mathcal{P}_0. \tag{SA.18}$$

By Assumptions 1(iii) and 3(iii), (SA.7) and (SA.18),

$$\hat{\Omega}^{-1/2}q(\hat{\theta}) = \Omega^{-1/2}Q + o_p(1) = O_p(1) \tag{SA.19}$$

uniformly over $\mathbb{P} \in \mathcal{P}_0$. The claim in the lemma follows by (SA.14), (SA.15) and (SA.19).

(e) By Assumption 2(i, ii) and the consistency of $\hat{\theta}$,

$$\sup_{\theta\in\Theta} \left\|\hat{\Omega}(\theta, \hat{\theta}) - \Omega(\theta, \theta_0)\right\| = o_p(1) \text{ uniformly over } \mathbb{P} \in \mathcal{P}_0. \tag{SA.20}$$

By (SA.13) and (SA.20),

$$\sup_{\theta\in\Theta} \left\|(\hat{\Omega}(\theta, \hat{\theta}) - \Omega(\theta, \theta_0))\hat{\Omega}^{-1}\right\| \leq (\lambda_{\min}(\hat{\Omega}))^{-1} \sup_{\theta\in\Theta} \left\|\hat{\Omega}(\theta, \hat{\theta}) - \Omega(\theta, \theta_0)\right\| = o_p(1) \tag{SA.21}$$

uniformly over $\mathbb{P} \in \mathcal{P}_0$, where the inequality is by the Cauchy-Schwarz inequality. Similarly,

$$\begin{aligned} &\sup_{\theta\in\Theta} \left\|\Omega(\theta, \theta_0)(\hat{\Omega}^{-1} - \Omega^{-1})\right\| = \sup_{\theta\in\Theta} \left\|\Omega(\theta, \theta_0)\hat{\Omega}^{-1}(\hat{\Omega} - \Omega)\Omega^{-1}\right\| \\ &\leq (\lambda_{\min}(\hat{\Omega})\lambda_{\min}(\Omega))^{-1} \sup_{\theta,\tilde{\theta}\in\Theta} \left\|\Omega(\theta, \tilde{\theta})\right\| ||\hat{\Omega} - \Omega|| = o_p(1) \end{aligned} \tag{SA.22}$$

uniformly over $\mathbb{P} \in \mathcal{P}_0$, where the inequality is by the Cauchy-Schwarz inequality, and the last equality is by Assumptions 2(iv) and 3(iii), (SA.8) and (SA.13). Collecting the results in (SA.21)

and (SA.22), we deduce that

$$\begin{aligned}&\sup_{\theta\in\Theta}\left\|\hat{V}(\theta)-V(\theta)\right\| \\ \leq &\sup_{\theta\in\Theta}\left\|(\hat{\Omega}(\theta,\hat{\theta})-\Omega(\theta,\theta_0))\hat{\Omega}^{-1}\right\|+\sup_{\theta\in\Theta}\left\|\Omega(\theta,\theta_0)(\hat{\Omega}^{-1}-\Omega^{-1})\right\|=o_p(1)\end{aligned} \tag{SA.23}$$

uniformly over $\mathbb{P}\in\mathcal{P}_0$. By Assumptions 2(iv) and 3(iii),

$$\sup_{\theta\in\Theta}\|V(\theta)\|\leq\sup_{\theta\in\Theta}\left\|S_0\Omega(\theta,\theta_0)\Omega^{-1}\right\|\leq(\lambda_{\min}(\Omega))^{-1}\sup_{\theta\in\Theta}\|\Omega(\theta,\theta_0)\|\leq c_\lambda^{-1}C_\Omega \tag{SA.24}$$

which finishes the proof. *Q.E.D.*

Proof of Lemma A2. To link $\hat{\xi}$ and ξ^*, we first define

$$\tilde{\xi}\equiv\left(\tilde{v}',\tilde{m}\left(\cdot\right)',\mathrm{vec}(V\left(\cdot\right))',\mathrm{vech}(\Omega)',\mathrm{vech}(\Omega_0\left(\cdot\right))',\mathrm{vech}(M)'\right)', \tag{SA.25}$$

where $\tilde{v}\equiv\Omega^{1/2}M\Omega^{-1/2}g(\theta_0)$ and $\tilde{m}\left(\cdot\right)\equiv g_0\left(\cdot\right)-V\left(\cdot\right)\tilde{v}$. The difference between $\tilde{\xi}$ and ξ^* lies in the first two elements, where $\tilde{v}$ and $\tilde{m}(\cdot)$ in $\tilde{\xi}$ involve the empirical process $g\left(\cdot\right)$, but $\Omega^{1/2}M\upsilon^*$ and $m^*(\cdot)$ in ξ^* involve the limiting Gaussian process $\psi(\cdot)$. Under Assumption 1(i), $g\left(\cdot\right)-\mathbb{E}[g\left(\cdot\right)]$ weakly converges to $\psi(\cdot)$, which is equivalent to

$$\lim_{n\to\infty}\sup_{\mathbb{P}\in\mathcal{P}_0}\sup_{f\in BL_1}\|\mathbb{E}\left[f(g-\mathbb{E}[g])\right]-E\left[f(\psi)\right]\|=0. \tag{SA.26}$$

Furthermore, Assumptions 1(iii), 2(iv) and 3(iii) imply that $\tilde{v}$ and $\tilde{m}(\cdot)$ are Lipschitz in $g(\cdot)$, and hence any bounded Lipschitz function of $\tilde{\xi}$ can be written as a bounded Lipschitz function of $\tilde{\xi}_1$, where $\tilde{\xi}_1$ replaces $g_0(.)$ in $\tilde{\xi}$ with $g_0(.)-E[g_0(.)]$, which together with (SA.26) implies that

$$\lim_{n\to\infty}\sup_{\mathbb{P}\in\mathcal{P}_0}\sup_{f\in BL_1}\left\|\mathbb{E}[f(\tilde{\xi})]-E\left[f(\xi^*)\right]\right\|=0. \tag{SA.27}$$

Next, note that the difference between $\hat{\xi}$ and $\tilde{\xi}$ is that $\hat{\Omega}$, $\hat{M}$, and $\hat{V}(\theta)$ in $\hat{\xi}$ are replaced by their probability limits in Lemma A1(c, d, e) of Cheng, Dou, and Liao (2021). Thus, we have $\hat{\xi}=\tilde{\xi}+o_p(1)$ uniformly over $\mathbb{P}\in\mathcal{P}_0$ following Lemma A1(c, d, e) in Cheng, Dou, and Liao (2021), which implies

$$\lim_{n\to\infty}\sup_{\mathbb{P}\in\mathcal{P}_0}\sup_{f\in BL_1}\left\|\mathbb{E}[f(\hat{\xi})]-\mathbb{E}[f(\tilde{\xi})]\right\|=0. \tag{SA.28}$$

The desirable result follows from (SA.27), (SA.28), and the triangle inequality. *Q.E.D.*

Proof of Lemma A6. Since $0\leq L(\upsilon;d^*)\leq\upsilon'M\upsilon$ for any $\upsilon\in\mathbb{R}^k$ and $ut_C(u)\leq C$ for any $u\geq 0$,

we have

$$\left|L_C(\upsilon;d^*)-\overline{L}_C(\upsilon;d^*)\right| = L(\upsilon;d^*)t_C(\upsilon' M\upsilon)I\{\|\upsilon\|^2>C\} \le CI\{\|\upsilon\|^2>C\} \tag{SA.29}$$

for any $\upsilon \in \mathbb{R}^k$, which implies that

$$P\left(\left|L_C(\upsilon^*,d^*)-\overline{L}_C(\upsilon^*,d^*)\right|>\varepsilon\right) \le P\left(I\{\|\upsilon^*\|^2>C\}>\varepsilon/C\right) \le P\left(\|\upsilon^*\|^2>C\right). \tag{SA.30}$$

Since $\|\upsilon^*\|^2$ follows the chi-square distribution with degree of freedom k, there exists a constant C_δ such that $P(\|\upsilon^*\|^2>C_\delta)\le\delta/4$ which together with (SA.30) implies that for any $C\ge C_\delta$

$$P\left(\left|L_C(\upsilon^*,d^*)-\overline{L}_C(\upsilon^*,d^*)\right|>\varepsilon\right)\le\delta/4. \tag{SA.31}$$

By the union bound of probability and (SA.31), we have for any $C\ge C_\delta$,

$$\begin{aligned} P\left(L_C(\upsilon^*,d^*)>c_{\alpha,C}(d^*)+\varepsilon\right) &\le P\left(\overline{L}_C(\upsilon^*,d^*)+\left|L_C(\upsilon^*,d^*)-\overline{L}_C(\upsilon^*,d^*)\right|>c_{\alpha,C}(d^*)+\varepsilon\right)\\ &\le P\left(\overline{L}_C(\upsilon^*,d^*)>c_{\alpha,C}(d^*)\right)+\delta/4\le\alpha+\delta/4, \end{aligned} \tag{SA.32}$$

where the last inequality is by the definition of $c_{\alpha,C}(d^*)$. Since $R_C(\xi^*)=L_C(\upsilon^*,d^*)$ by definition, the claim of the lemma follows from (SA.32). *Q.E.D.*

Proof of Lemma A7. (a). Since $\upsilon^*=O_p(1)$, by Assumptions 1(iii), 2(iv) and 3(iii), Lemma A1(b, d) in Cheng, Dou, and Liao (2021), and (SA.18), we have uniformly over $\mathbb{P}\in\mathcal{P}_0$,

$$\hat{\Omega}^{1/2}\hat{M}\upsilon^*-g(\hat{\theta})=\Omega^{1/2}M\Omega^{-1/2}(\Omega^{1/2}\upsilon^*-g(\theta_0))+o_p(1)=O_p(1), \tag{SA.33}$$

which together with Lemma A1(e) in Cheng, Dou, and Liao (2021) implies that

$$\sup_{\theta\in\Theta}\left\|\hat{V}(\theta)(\hat{\Omega}^{1/2}\hat{M}\upsilon^*-g(\hat{\theta}))\right\|=O_p(1)\text{ uniformly over }\mathbb{P}\in\mathcal{P}_0. \tag{SA.34}$$

Therefore by the triangle inequality, Assumption 1 and (SA.34),

$$\begin{aligned} &\sup_{\theta\in\Theta}\left\|n^{-1/2}(\hat{m}(\theta)+\hat{V}(\theta)\hat{\Omega}^{1/2}\hat{M}\upsilon^*)-\mathbb{E}[g_0(\theta)]\right\|\\ \le&\sup_{\theta\in\Theta}\left\|n^{-1/2}g_0(\theta)-\mathbb{E}[g_0(\theta)]\right\|+\sup_{\theta\in\Theta}\left\|n^{-1/2}\hat{V}(\theta)(\hat{\Omega}^{1/2}\hat{M}\upsilon^*-g(\hat{\theta}))\right\|=o_p(1) \end{aligned} \tag{SA.35}$$

uniformly over $\mathbb{P}\in\mathcal{P}_0$. By Assumptions 1(iii), 2(i), 3(iii) and 4(i), and (SA.35), we can apply similar arguments in the proof of (SA.3) to deduce that

$$\hat{\theta}^*=\theta_0+o_p(1) \tag{SA.36}$$

uniformly over $\mathbb{P} \in \mathcal{P}_0 \cap \mathcal{P}_{00}$. Applying the chain rule, we get the first order condition of $\hat{\theta}^*$:

$$0_{d_\theta \times 1} = 2\left(\frac{\partial}{\partial \theta'}(\hat{m}(\hat{\theta}^*) + \hat{V}_S(\hat{\theta}^*)\upsilon^*)\right)' (\hat{\Omega}_0^*)^{-1}(\hat{m}(\hat{\theta}^*) + \hat{V}_S(\hat{\theta}^*)\upsilon^*)$$
$$- \left((\hat{m}(\hat{\theta}^*) + \hat{V}_S(\hat{\theta}^*)\upsilon^*)'(\hat{\Omega}_0^*)^{-1}\frac{\partial \hat{\Omega}_0(\hat{\theta}^*)}{\partial \theta_j}(\hat{\Omega}_0^*)^{-1}(\hat{m}(\hat{\theta}^*) + \hat{V}_S(\hat{\theta}^*)\upsilon^*)\right)_{j=1,\ldots,d_\theta}, \tag{SA.37}$$

where $\hat{\Omega}_0^* \equiv \hat{\Omega}_0(\hat{\theta}^*)$ and $\hat{V}_S(\hat{\theta}^*) \equiv \hat{V}(\hat{\theta}^*)\hat{\Omega}^{1/2}\hat{M}$. Using (SA.36) and similar arguments for showing (SA.5) and (SA.6), we obtain

$$g_0(\hat{\theta}^*) = g_0(\theta_0) + n^{1/2}(\hat{\theta}^* - \theta_0)(Q_0 + o_p(1)) = O_p(1) + n^{1/2}(\hat{\theta}^* - \theta_0)O_p(1) \tag{SA.38}$$

uniformly over $\mathbb{P} \in \mathcal{P}_0$. By (SA.34) and (SA.38),

$$\hat{m}(\hat{\theta}^*) + \hat{V}_S(\hat{\theta}^*)\upsilon^* = g_0(\hat{\theta}^*) + \hat{V}(\hat{\theta}^*)(\hat{\Omega}^{1/2}\hat{M}\upsilon^* - g(\hat{\theta})) = O_p(1) + n^{1/2}(\hat{\theta}^* - \theta_0)O_p(1), \tag{SA.39}$$

which together with Assumptions 2 and 3(iii), (SA.36) and the Cauchy-Schwarz inequality implies that for any $j = 1, ..., d_\theta$,

$$n^{-1/2}(\hat{m}(\hat{\theta}^*) + \hat{V}_S(\hat{\theta}^*)\upsilon^*)'(\hat{\Omega}_0^*)^{-1}\frac{\partial \hat{\Omega}_0(\hat{\theta}^*)}{\partial \theta_j}(\hat{\Omega}_0^*)^{-1}(\hat{m}(\hat{\theta}^*) + \hat{V}_S(\hat{\theta}^*)\upsilon^*)$$
$$= n^{1/2}(\hat{\theta}^* - \theta_0)o_p(1) + o_p(1), \text{ uniformly over } \mathbb{P} \in \mathcal{P}_0 \cap \mathcal{P}_{00}. \tag{SA.40}$$

By Assumptions 2(iii, iv) and 3(iii),

$$\max_{1 \le j \le d_\theta} \sup_{\theta \in \Theta} \left\| \hat{V}(\theta)/\partial \theta_j \right\| = O_p(1) \text{ uniformly over } \mathbb{P} \in \mathcal{P}_0, \tag{SA.41}$$

which combined with (SA.33) implies that

$$\frac{\partial}{\partial \theta'}\hat{V}(\hat{\theta}^*)(\hat{\Omega}^{1/2}\hat{M}\upsilon^* - g(\hat{\theta})) = O_p(1) \tag{SA.42}$$

uniformly over $\mathbb{P} \in \mathcal{P}_0 \cap \mathcal{P}_{00}$. By Assumption 1(ii) and (SA.36),

$$n^{-1/2}\frac{\partial g_0(\hat{\theta}^*)}{\partial \theta'} = Q_0 + o_p(1) \text{ uniformly over } \mathbb{P} \in \mathcal{P}_0 \cap \mathcal{P}_{00}. \tag{SA.43}$$

Collecting the results in (SA.42) and (SA.43), we have

$$\begin{aligned}
&n^{-1/2}\frac{\partial}{\partial\theta'}(\hat{m}(\hat{\theta}^*)+\hat{V}_S(\hat{\theta}^*)\upsilon^*)\\
&=n^{-1/2}\frac{\partial}{\partial\theta'}g_0(\hat{\theta}^*)+n^{-1/2}\frac{\partial}{\partial\theta'}\hat{V}(\hat{\theta}^*)(\hat{\Omega}^{1/2}\hat{M}\upsilon^*-g(\hat{\theta}))=Q_0+o_p(1)
\end{aligned}\tag{SA.44}$$

uniformly over $\mathbb{P}\in\mathcal{P}_0\cap\mathcal{P}_{00}$. Applying the first order expansion to get

$$\begin{aligned}
&\hat{m}(\hat{\theta}^*)+\hat{V}_S(\hat{\theta}^*)\upsilon^*\\
&=g_0(\hat{\theta}^*)+\hat{V}(\hat{\theta}^*)(\hat{\Omega}^{1/2}\hat{M}\upsilon^*-g(\hat{\theta}))\\
&=g_0(\theta_0)+\hat{V}(\theta_0)(\hat{\Omega}^{1/2}\hat{M}\upsilon^*-g(\hat{\theta}))\\
&\quad+\frac{\partial g_0(\tilde{\theta}^*)}{\partial\theta'}(\hat{\theta}^*-\theta_0)+\frac{\partial\hat{V}(\tilde{\theta}^*)}{\partial\theta'}(\hat{\theta}^*-\theta_0)(\hat{\Omega}^{1/2}\hat{M}\upsilon^*-g(\hat{\theta}))\\
&=g_0(\theta_0)+\hat{V}(\theta_0)(\hat{\Omega}^{1/2}\hat{M}\upsilon^*-g(\hat{\theta}))+(Q_0+o_p(1))n^{1/2}(\hat{\theta}^*-\theta_0)+o_p(1)
\end{aligned}\tag{SA.45}$$

uniformly over $\mathbb{P}\in\mathcal{P}_0\cap\mathcal{P}_{00}$, where the third equality is by Assumption 1(ii), (SA.33), (SA.36) and (SA.41). By $V(\theta_0)=S_0$, Lemma A1(b, e) in Cheng, Dou, and Liao (2021) and (SA.33),

$$\hat{V}(\theta_0)(\hat{\Omega}^{1/2}\hat{M}\upsilon^*-g(\hat{\theta}))=S_0\Omega^{1/2}M\Omega^{-1/2}(\Omega^{1/2}\upsilon^*-g(\theta_0))+o_p(1)\tag{SA.46}$$

uniformly over $\mathbb{P}\in\mathcal{P}_0$. By (SA.45) and (SA.46), we have uniformly over $\mathbb{P}\in\mathcal{P}_0\cap\mathcal{P}_{00}$,

$$\begin{aligned}
\hat{m}(\hat{\theta}^*)+\hat{V}_S(\hat{\theta}^*)\upsilon^*&=(Q_0+o_p(1))n^{1/2}(\hat{\theta}^*-\theta_0)\\
&\quad+Q_0(Q'\Omega^{-1}Q)^{-1}Q'\Omega^{-1}g(\theta_0)+S_0\Omega^{1/2}M\upsilon^*+o_p(1).
\end{aligned}\tag{SA.47}$$

By Assumptions 3(iii) and 4(ii), $Q_0'\Omega_0^{-1}Q_0$ is positive definite. Therefore collecting the results in (SA.37), (SA.40), (SA.44), (SA.47), and applying Assumption 2(i, ii) and (SA.36) to $\hat{\Omega}_0^*$, we obtain

$$n^{1/2}(\hat{\theta}^*-\theta_0)=-(Q'\Omega^{-1}Q)^{-1}Q'\Omega^{-1}g(\theta_0)-\left(Q_0'\Omega_0^{-1}Q_0\right)^{-1}Q_0'\Omega_0^{-1}S_0\Omega^{1/2}M\upsilon^*+o_p(1)\tag{SA.48}$$

uniformly over $\mathbb{P}\in\mathcal{P}_0\cap\mathcal{P}_{00}$, which proves part (a) of the lemma.

(b) By (SA.48), Assumptions 1(iii), 2(iv), 3(iii) and 4(ii),

$$n^{1/2}(\hat{\theta}^*-\theta_0)=O_p(1)\text{ uniformly over }\mathbb{P}\in\mathcal{P}_0\cap\mathcal{P}_{00}.\tag{SA.49}$$

By Assumptions 2(iv), 3(iii) and 4(ii), (SA.47) and (SA.48), we have uniformly over $\mathbb{P} \in \mathcal{P}_0 \cap \mathcal{P}_{00}$,

$$\begin{aligned}\hat{m}(\hat{\theta}^*) + \hat{V}_S(\hat{\theta}^*)\upsilon^* &= -Q_0\left(Q_0'\Omega_0^{-1}Q_0\right)^{-1}Q_0'\Omega_0^{-1}S_0\Omega^{1/2}M\upsilon^* + S_0\Omega^{1/2}M\upsilon^* + o_p(1)\\ &= \Omega_0^{1/2}M_0\Omega_0^{-1/2}S_0\Omega^{1/2}M\upsilon^* + o_p(1)\\ &= \Omega_0^{1/2}M_0\Omega_0^{-1/2}S_0\Omega^{1/2}\upsilon^* + o_p(1) = O_p(1),\end{aligned} \tag{SA.50}$$

where $M_0 \equiv I_{k_0} - \Omega_0^{-1/2}Q_0\left(Q_0'\Omega_0^{-1}Q_0\right)^{-1}Q_0'\Omega_0^{-1/2}$ and the third equality is by $M_0\Omega_0^{-1/2}Q_0 = 0_{k_0\times 1}$. By Assumptions 2(i, ii) and 3(iii), (SA.36) and (SA.50), we deduce that uniformly over $\mathbb{P} \in \mathcal{P}_0 \cap \mathcal{P}_{00}$,

$$(\hat{m}(\hat{\theta}^*) + \hat{V}_S(\hat{\theta}^*)\upsilon^*)'(\hat{\Omega}_0^*)^{-1}(\hat{m}(\hat{\theta}^*) + \hat{V}_S(\hat{\theta}^*)\upsilon^*) = \upsilon^{*\prime}\tilde{M}_0\upsilon^* + o_p(1), \tag{SA.51}$$

where $\tilde{M}_0 \equiv (\Omega_0^{-1/2}S_0\Omega^{1/2})'M_0(\Omega_0^{-1/2}S_0\Omega^{1/2})$. Since $\upsilon^* = O_p(1)$, Lemma A1(d) implies that

$$\upsilon^{*\prime}\hat{M}\upsilon^* = \upsilon^{*\prime}M\upsilon^* + o_p(1) \tag{SA.52}$$

uniformly over $\mathbb{P} \in \mathcal{P}_0$, which together with (SA.51) finishes the proof.

(c). Since $M \equiv I_k - \Omega^{-1/2}Q(Q'\Omega^{-1}Q)^{-1}Q'\Omega^{-1/2}$, $M^2 = M$. Moreover,

$$\begin{aligned}\tilde{M}_0^2 &= (\Omega_0^{-1/2}S_0\Omega^{1/2})'M_0(\Omega_0^{-1/2}S_0\Omega^{1/2})(\Omega_0^{-1/2}S_0\Omega^{1/2})'M_0(\Omega_0^{-1/2}S_0\Omega^{1/2})\\ &= (\Omega_0^{-1/2}S_0\Omega^{1/2})'M_0(\Omega_0^{-1/2}S_0\Omega^{1/2}) = \tilde{M}_0\end{aligned} \tag{SA.53}$$

and

$$\begin{aligned}\tilde{M}_0M &= \tilde{M}_0 - (\Omega_0^{-1/2}S_0\Omega^{1/2})'M_0(\Omega_0^{-1/2}S_0\Omega^{1/2})\Omega^{-1/2}Q(Q'\Omega^{-1}Q)^{-1}Q'\Omega^{-1/2}\\ &= \tilde{M}_0,\end{aligned} \tag{SA.54}$$

where the second equality is by $M_0\Omega_0^{-1/2}Q_0 = 0_{k_0\times 1}$. Similarly, $M\tilde{M}_0 = \tilde{M}_0$. Therefore, $(M - \tilde{M}_0)^2 = M^2 - M\tilde{M}_0 - \tilde{M}_0M + \tilde{M}_0^2 = M - \tilde{M}_0$ which implies that $M - \tilde{M}_0$ is an idempotent matrix. The rank of $M - \tilde{M}_0$ equals the trace of $M - \tilde{M}_0$ since $M - \tilde{M}_0$ is idempotent. By the definition of M and $\tilde{M}_0$,

$$tr(M) = k - tr(\Omega^{-1/2}Q(Q'\Omega^{-1}Q)^{-1}Q'\Omega^{-1/2}) = k - d_\theta \tag{SA.55}$$

and

$$\begin{aligned}tr(\tilde{M}_0) &= tr((\Omega_0^{-1/2}S_0\Omega^{1/2})'M_0(\Omega_0^{-1/2}S_0\Omega^{1/2}))\\ &= tr(M_0(\Omega_0^{-1/2}S_0\Omega S_0'\Omega_0^{-1/2})) = tr(M_0) = k_0 - d_\theta,\end{aligned} \tag{SA.56}$$

which implies that $tr(M - \tilde{M}_0) = k - k_0 = k_1$. Therefore, $M - \tilde{M}_0$ is an idempotent matrix with rank k_1 which together with $\upsilon^* \sim N(0, I_k)$ proves the claim (c). *Q.E.D.*

SB Proofs for Bounded Lipschitz Conditions

This section contains the proofs of Lemmas A4 and A5 in Cheng, Dou, and Liao (2021), and the auxiliary results used to show them.

Lemma SB1. *Consider $\xi \equiv (x', d')'$ where d satisfies Assumptions 2 and 3. Then we have for any given $C > 0$:*
(i) *$R(\xi)$ is bounded and Lipschitz in ξ on the set $\{\xi : R(\xi) \geq 0$ and $x'\Omega_d^{-1}x \leq C\}$;*
(ii) *$L(\upsilon; d)$ is bounded and Lipschitz in d on the set $\{(\upsilon, d) : L(\upsilon; d) \geq 0$ and $\|\upsilon\| \leq C\}$.*

Proof of Lemma SB1. (i) Let $S_\xi \equiv \{\xi : R(\xi) \geq 0$ and $x'\Omega_d^{-1}x \leq C\}$. By the definition of $R(\xi)$, $R(\xi) \leq x'\Omega_d^{-1}x \leq C$ for any $\xi \in S_\xi$, which shows that $R(\xi)$ is bounded on S_ξ. Next, we want to show that for any $\xi_1, \xi_2 \in S_\xi$,

$$|R(\xi_1) - R(\xi_2)| \leq C_R \|\xi_1 - \xi_2\|_s \tag{SB.57}$$

for some constant C_R. By the triangle inequality,

$$|R(\xi_1) - R(\xi_2)| \leq R(\xi_1) + R(\xi_2) \leq x_2'\Omega_{d,2}^{-1}x_2 + x_1'\Omega_{d,1}^{-1}x_1 = 2C, \tag{SB.58}$$

which implies that the claimed result holds with a Lipschitz constant $C_R = 2$ if $\|\xi_1 - \xi_2\|_s > C$. Thus, it is only necessary to consider the case that $\|\xi_1 - \xi_2\|_s \leq C$.

Define $A_j(\theta) \equiv m_{d,j}(\theta) + V_{d,j}(\theta)x_j$ for $j = 1, 2$. Consider any $\xi_1, \xi_2 \in S_\xi$, by the triangle inequality,

$$\begin{aligned} |R(\xi_1) - R(\xi_2)| \leq & \left| x_1'\Omega_{d,1}^{-1}x_1 - x_2'\Omega_{d,2}^{-1}x_2 \right| \\ & + \left| \min_{\theta\in\Theta} A_1(\theta)'(\Omega_{0,d,1}(\theta))^{-1}A_1(\theta) - \min_{\theta\in\Theta} A_2(\theta)'(\Omega_{0,d,2}(\theta))^{-1}A_2(\theta) \right|. \end{aligned} \tag{SB.59}$$

By the triangle inequality, the Cauchy-Schwarz inequality, $x_1'\Omega_{d,1}^{-1}x_1 \leq C$ and $x_2'\Omega_{d,2}^{-1}x_2 \leq C$,

$$\begin{aligned} \left| x_1'\Omega_{d,1}^{-1}x_1 - x_2'\Omega_{d,2}^{-1}x_2 \right| &\leq \left| (x_1 - x_2)'\Omega_{d,1}^{-1}x_1 \right| + \left| x_2'\Omega_{d,2}^{-1}(\Omega_{d,1} - \Omega_{d,2})\Omega_{d,1}^{-1}x_1 \right| + \left| x_2'\Omega_{d,2}^{-1}(x_1 - x_2) \right| \\ &\leq \left[\frac{(x_1'\Omega_{d,1}^{-1}x_1)^{1/2}}{(\lambda_{\min}(\Omega_{d,1}))^{1/2}} + \frac{(x_2'\Omega_{d,2}^{-1}x_2)^{1/2}}{(\lambda_{\min}(\Omega_{d,2}))^{1/2}} \right] \|x_1 - x_2\| \\ &\quad + \frac{(x_1'\Omega_{d,1}^{-1}x_1)^{1/2}(x_2'\Omega_{d,2}^{-1}x_2)^{1/2}}{(\lambda_{\min}(\Omega_{d,1})\lambda_{\min}(\Omega_{d,2}))^{1/2}} \|\Omega_{d,1} - \Omega_{d,2}\| \\ &\leq 2c_\lambda^{-1/2}C^{1/2} \|x_1 - x_2\| + c_\lambda^{-1}C \|\Omega_{d,1} - \Omega_{d,2}\| \leq C_1 \|\xi_1 - \xi_2\|_s \end{aligned} \tag{SB.60}$$

for some constant C_1.

Let θ_j denote the minimizer of $A_j(\theta)'(\Omega_{0,d,j}(\theta))^{-1}A_j(\theta)$ for $j = 1, 2$. By the triangle inequality, we have

$$\begin{aligned}
&\left|\min_{\theta\in\Theta} A_1(\theta)'(\Omega_{0,d,1}(\theta))^{-1}A_1(\theta) - \min_{\theta\in\Theta} A_2(\theta)'(\Omega_{0,d,2}(\theta))^{-1}A_2(\theta)\right| \\
&\leq \max_{\{\theta_1,\theta_2\}} \left|A_1(\theta)'(\Omega_{0,d,1}(\theta))^{-1}A_1(\theta) - A_2(\theta)'(\Omega_{0,d,2}(\theta))^{-1}A_2(\theta)\right| \\
&\leq \max_{\{\theta_1,\theta_2\}} \left|(A_1(\theta) - A_2(\theta))' (\Omega_{0,d,1}(\theta))^{-1}A_1(\theta)\right| \\
&+ \max_{\{\theta_1,\theta_2\}} \left|A_2(\theta)'(\Omega_{0,d,1}(\theta))^{-1} (A_1(\theta) - A_2(\theta))\right| \\
&+ \max_{\{\theta_1,\theta_2\}} \left|A_2(\theta)'((\Omega_{0,d,1}(\theta))^{-1} - (\Omega_{0,d,2}(\theta))^{-1})A_2(\theta)\right|. \qquad \text{(SB.61)}
\end{aligned}$$

We next investigate the three terms after the second inequality of (SB.61) one by one. By the triangle inequality and the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\max_{\{\theta_1,\theta_2\}} \|A_1(\theta) - A_2(\theta)\| \\
&\leq \sup_{\theta\in\Theta} \|m_{d,1}(\theta) - m_{d,2}(\theta)\| + \|x_1\| \sup_{\theta\in\Theta} \|V_{d,1}(\theta) - V_{d,2}(\theta)\| + \sup_{\theta\in\Theta} \|V_{d,2}(\theta)\| \|x_1 - x_2\| \\
&\leq \|\xi_1 - \xi_2\|_s + \lambda_{\max}(\Omega_{d,1})(x_1'\Omega_{d,1}^{-1}x_1)^{1/2} \|\xi_1 - \xi_2\|_s + C_V \|\xi_1 - \xi_2\|_s, \qquad \text{(SB.62)}
\end{aligned}$$

where $C_V \equiv c_\lambda^{-1}C_\Omega$ and the second inequality is by the definition of $\|\xi_1 - \xi_2\|_s$ and $\sup_{\theta\in\Theta} \|V_{2,d}(\theta)\| \leq c_\lambda^{-1}C_\Omega$ (which is proved in Lemma A1(e) of Cheng, Dou, and Liao (2021)). Therefore, by Assumption 2(iv) and (SB.62),

$$\max_{\{\theta_1,\theta_2\}} \|A_1(\theta) - A_2(\theta)\| \leq (1 + (CC_\Omega)^{1/2} + C_V) \|\xi_1 - \xi_2\|_s, \qquad \text{(SB.63)}$$

which together with the triangle inequality, Assumption 2(iv), the restrictions on ξ_1 and ξ_2, and $\|\xi_1 - \xi_2\|_s \leq C$ implies that

$$\begin{aligned}
&\max_{\{\theta_1,\theta_2\}} \|A_1(\theta)\| \leq \|A_1(\theta_1)\| + \|A_1(\theta_2)\| \\
&\leq \|A_1(\theta_1)\| + \|A_2(\theta_2)\| + \|A_1(\theta_2) - A_2(\theta_2)\| \\
&\leq C_\Omega^{1/2} \left((A_1(\theta_1)'(\Omega_{d,1,0}(\theta_1))^{-1}A_1(\theta_1))^{1/2} + (A_2(\theta_2)'(\Omega_{d,2,0}(\theta_2))^{-1}A_2(\theta_2))^{1/2}\right) \\
&+ (1 + (CC_\Omega)^{1/2} + C_V) \|\xi_1 - \xi_2\|_s \\
&\leq 2(C_\Omega C)^{1/2} + (1 + (CC_\Omega)^{1/2} + C_V)C. \qquad \text{(SB.64)}
\end{aligned}$$

By the same arguments, the inequality in (SB.64) applies to $\max_{\{\theta_1,\theta_2\}} ||A_2(\theta)||$. By the Cauchy-

Schwarz inequality, Assumption 3(iii), (SB.63) and (SB.64),

$$\begin{aligned}&\max_{\{\theta_1,\theta_2\}}\left|(A_1(\theta)-A_2(\theta))'\left(\Omega_{d,1,0}(\theta)\right)^{-1}A_1(\theta)\right|\\ \leq&\max_{\{\theta_1,\theta_2\}}\left(\lambda_{\min}(\Omega_{0,d,1}(\theta))\right)^{-1}\|A_1(\theta)-A_2(\theta)\|\,\|A_1(\theta)\|\leq C_2\,\|\xi_1-\xi_2\|_s\end{aligned}\tag{SB.65}$$

for some constant C_2. Similarly, we can show that

$$\max_{\{\theta_1,\theta_2\}}\left|A_2(\theta)'(\Omega_{0,d,1}(\theta))^{-1}\left(A_1(\theta)-A_2(\theta)\right)\right|\leq C_2\,\|\xi_1-\xi_2\|_s\,.\tag{SB.66}$$

By the Cauchy-Schwarz inequality, Assumption 3(iii) and (SB.64),

$$\begin{aligned}&\max_{\{\theta_1,\theta_2\}}\left|A_2(\theta)'((\Omega_{0,d,1}(\theta))^{-1}-(\Omega_{0,d,2}(\theta))^{-1})A_2(\theta)\right|\\ \leq&\frac{\max_{\{\theta_1,\theta_2\}}\|A_2(\theta)\|^2\,\|\Omega_{0,d,1}(\theta)-\Omega_{0,d,2}(\theta)\|}{(\inf_{\theta\in\Theta}\lambda_{\min}(\Omega_{0,d,1}(\theta))\inf_{\theta\in\Theta}\lambda_{\min}(\Omega_{0,d,2}(\theta)))^{-1}}\leq C_3\sup_{\theta\in\Theta}\|\Omega_{0,d,1}(\theta)-\Omega_{0,d,2}(\theta)\|\end{aligned}\tag{SB.67}$$

for some constant C_3. Collecting the results in (SB.61), (SB.65), (SB.66) and (SB.67), we get

$$\left|\min_{\theta\in\Theta}A_1(\theta)'(\Omega_{0,d,1}(\theta))^{-1}A_1(\theta)-\min_{\theta\in\Theta}A_2(\theta)'(\Omega_{0,d,2}(\theta))^{-1}A_2(\theta)\right|\leq(2C_2+C_3)\,\|\xi_1-\xi_2\|_s\,,\tag{SB.68}$$

which together with (SB.59) and (SB.60) implies that the Lipschitz constant is $C_R=C_1+2C_2+C_3$.

(ii) Note that under the condition $L(\upsilon;d)\geq 0$ and $\|\upsilon\|\leq C$, we have $0\leq L(\upsilon;d)\leq\upsilon' M_d\upsilon\leq C^2$. To show $L(\upsilon;d)$ is Lipschitz in d, we write

$$L(\upsilon;d)\equiv\upsilon' M_d\upsilon-\min_{\theta\in\Theta}\left(m_d(\theta)+V_{d,S}(\theta)\upsilon\right)'\left(\Omega_{d,0}(\theta)\right)^{-1}\left(m_d(\theta)+V_{d,S}(\theta)\upsilon\right),\tag{SB.69}$$

where $V_{d,S}(\cdot)\equiv V_d(\cdot)\,\Omega_d^{1/2}M_d$. This functional form is analogous to $R(\xi)$, with Ω_d and $V_d(\cdot)$ in $R(\xi)$ replaced by M_d and $V_{d,S}(\cdot)$, respectively. Given that $V_{d,S}(\cdot)$ is Lipschitz in d (established in Lemma SC4 below) and $\sup_{\theta\in\Theta}\|V_{d,S}(\theta)\|\leq C_\Omega^{1/2}C_V$ (by Assumptions 2(iv) and 3(iii) and Lemma A1(e) of Cheng, Dou, and Liao (2021)), showing $L(\upsilon;d)$ is Lipschitz in d is analogous to showing $R(\xi)$ is Lipschitz in ξ. The only difference is that M_d is not a full rank matrix, unlike Ω_d, which is the reason that we have to bound $\|\upsilon\|$ directly instead of bounding $\upsilon' M_d\upsilon$. Because (SB.60) in the proof of part (i) uses the full rank condition of Ω_d, we replace (SB.60) with the following argument to show $\upsilon' M_d\upsilon$ is Lipschitz in M_d. Given $\|\upsilon\|\leq C$, we have

$$\left|\upsilon' M_{d,1}\upsilon-\upsilon' M_{d,2}\upsilon\right|=\left|\upsilon'(M_{d,1}-M_{d,2})\upsilon\right|\leq\|\upsilon\|^2\,\|M_{d,1}-M_{d,2}\|\leq C^2\,\|M_{d,1}-M_{d,2}\|\tag{SB.70}$$

by the Cauchy-Schwarz inequality. The rest of the proof is analogous to those in the proof of

Lemma SB1(i) and hence is omitted. *Q.E.D.*

Proof of Lemma A4. The truncation function $t_C(u)$ satisfies the following properties: (i) for any $u \in \mathbb{R}^+$, $0 \le t_C(u) \le 1$ and $ut_C(u) \le u$; (ii) for any u_1, $u_2 \in \mathbb{R}$, $|t_C(u_1) - t_C(u_2)| \le C^{-1} |u_1 - u_2|$, which implies that $t_C(u)$ is Lipschitz in u. Therefore,

$$0 \le R_C(\xi) \le \left(x'\Omega_d^{-1}x\right) t_C(x'\Omega_d^{-1}x) \le C \tag{SB.71}$$

which means that $R_C(\xi)$ is bounded.

Next, we show that $R_C(\xi)$ is Lipschitz in ξ. That is for any ξ_j with $R(\xi_j) \ge 0$ $(j = 1, 2)$,

$$|R_C(\xi_1) - R_C(\xi_2)| \le C_R \|\xi_1 - \xi_2\|_s , \tag{SB.72}$$

where C_R is a constant. Without loss of generality, we assume that $x_2'\Omega_{d,2}^{-1}x_2 \le x_1'\Omega_{d,1}^{-1}x_1$. By the triangle inequality,

$$|R_C(\xi_1) - R_C(\xi_2)| \le \left|(R(\xi_1) - R(\xi_2))t_C(x_1'\Omega_{d,1}^{-1}x_1)\right| + \left|(t_C(x_1'\Omega_{d,1}^{-1}x_1) - t_C(x_2'\Omega_{d,2}^{-1}x_2))R(\xi_2)\right| . \tag{SB.73}$$

We have (SB.72) holds with $C_R = C_{R_1} + C_{R_2}$ if we can show that

$$\left|(R(\xi_1) - R(\xi_2))t_C(x_1'\Omega_{d,1}^{-1}x_1)\right| \le C_{R_1} \|\xi_1 - \xi_2\|_s \tag{SB.74}$$

and

$$\left|(t_C(x_1'\Omega_{d,1}^{-1}x_1) - t_C(x_2'\Omega_{d,2}^{-1}x_2))R(\xi_2)\right| \le C_{R_2} \|\xi_1 - \xi_2\|_s \tag{SB.75}$$

for some constants C_{R_1} and C_{R_2}.

We first consider (SB.74). First, note that it holds trivially if $x_1'\Omega_{d,1}^{-1}x_1 > 2C$ because, in this case, $t_C(x_1'\Omega_{d,1}^{-1}x_1) = 0$. Second, given $x_2'\Omega_{d,2}^{-1}x_2 \le x_1'\Omega_{d,1}^{-1}x_1 \le 2C$, we deduce that

$$\left|(R(\xi_1) - R(\xi_2))t_C(x_1'\Omega_{d,1}^{-1}x_1)\right| \le |R(\xi_1) - R(\xi_2)| \le C_{R_1} \|\xi_1 - \xi_2\|_s , \tag{SB.76}$$

where the first inequality is by property (i) of $t_C(u)$, and the second inequality is by Lemma SB1(i).

Next, we show (SB.75). First, note that it holds trivially if $2C < x_2'\Omega_{d,2}^{-1}x_2$. In this case,

$$t_C(x_1'\Omega_{d,1}^{-1}x_1) = t_C(x_2'\Omega_{d,2}^{-1}x_2) = 0 \tag{SB.77}$$

following the definition of $t_C(u)$. Second, given $x_2'\Omega_{d,2}^{-1}x_2 \le 2C$, we have

$$\left|(t_C(x_1'\Omega_{d,1}^{-1}x_1) - t_C(x_2'\Omega_{d,2}^{-1}x_2))R(\xi_2)\right| \le |R(\xi_2)| \le x_2'\Omega_{d,2}^{-1}x_2 \le 2C. \tag{SB.78}$$

Thus, (SB.75) holds with $C_{R_2} = 1$ if $\|\xi_1 - \xi_2\|_s > 2C$. Third, it remains to consider the case where $x_2'\Omega_{d,2}^{-1}x_2 \leq 2C$ and $\|\xi_1 - \xi_2\|_s \leq 2C$. In this case, Lemma SC3 in Section SC implies that $x_1'\Omega_{d,1}^{-1}x_1 \leq C^*$ for some constant C^*. In this case,

$$\begin{aligned} &\left|(t_C(x_1'\Omega_{d,1}^{-1}x_1) - t_C(x_2'\Omega_{d,2}^{-1}x_2))R(\xi_2)\right| \\ \leq &\left|t_C(x_1'\Omega_{d,1}^{-1}x_1) - t_C(x_2'\Omega_{d,2}^{-1}x_2)\right| x_2'\Omega_{d,2}^{-1}x_2 \leq 2C\left|x_1'\Omega_{d,1}^{-1}x_1 - x_2'\Omega_{d,2}^{-1}x_2\right| \end{aligned} \tag{SB.79}$$

using $0 \leq R(\xi_2) \leq x_2'\Omega_{d,2}^{-1}x_2 \leq 2C$ and $|t_C(u_1) - t_C(u_2)| \leq C^{-1}|u_1 - u_2|$ which follows from property (ii) of $t_C(u)$. Then we can show $|x_1'\Omega_{d,1}^{-1}x_1 - x_2'\Omega_{d,2}^{-1}x_2| \leq C_{R_2}\|\xi_1 - \xi_2\|_S$ for some constant C_{R_2} by the same arguments that show (SB.60) but with $x_1'\Omega_{d,1}^{-1}x_1 \leq 2C$ replaced by $x_1'\Omega_{d,1}^{-1}x_1 \leq C^*$. *Q.E.D.*

Lemma SB2. *Given $L(\upsilon; d) \geq 0$, $\overline{L}_C(\upsilon; d)$ is bounded and Lipschitz in d.*

Proof of Lemma SB2. The proof is analogous to that of Lemma A4 with the truncation function $t_C(x'\Omega_d^{-1}x)$ replaced by $t_C(\upsilon' M_d \upsilon)I\{\|\upsilon\|^2 \leq C\}$ and Lemma SB1(i) replaced by Lemma SB1(ii), and hence is omitted. *Q.E.D.*

Proof of Lemma A5. Since $0 \leq L(\upsilon; d) \leq \upsilon' M_d \upsilon$ for any $\upsilon \in \mathbb{R}^k$ and $ut_C(u) \leq C$ for any $u \geq 0$, we have

$$\overline{L}_C(\upsilon; d) = L(\upsilon; d)t_C(\upsilon' M_d \upsilon)I\{\|\upsilon\|^2 \leq C\} \leq L(\upsilon; d)t_C(\upsilon' M_d \upsilon) \leq C, \tag{SB.80}$$

which implies that $c_{\alpha,C}(d)$ is bounded. For any υ, any d_1 and d_2, by Lemma SB2 there exists a constant C_L such that

$$\left|\overline{L}_C(\upsilon; d_1) - \overline{L}_C(\upsilon; d_2)\right| \leq C_L\|d_1 - d_2\|_s. \tag{SB.81}$$

Since $\overline{L}_C(\upsilon; d_1) \geq \overline{L}_C(\upsilon; d_2) - C_L\|d_1 - d_2\|_s$ for any υ and $P^*\left(\overline{L}_C(\upsilon^*; d_1) > c_{\alpha,C}(d_1)\right) \leq \alpha$, we have $P\left(\overline{L}_C(\upsilon^*; d_2) > c_{\alpha,C}(d_1) + C_L\|d_1 - d_2\|_s\right) \leq \alpha$, which implies that

$$c_{\alpha,C}(d_2) \leq c_{\alpha,C}(d_1) + C_L\|d_1 - d_2\|_s. \tag{SB.82}$$

Similarly, we also have

$$c_{\alpha,C}(d_1) \leq c_{\alpha,C}(d_2) + C_L\|d_1 - d_2\|_s. \tag{SB.83}$$

Combining (SB.82) and (SB.83), we get

$$|c_{\alpha,C}(d_1) - c_{\alpha,C}(d_2)| \leq C_L\|d_1 - d_2\|_s, \tag{SB.84}$$

which shows the claim of the lemma. *Q.E.D.*

SC Additional Auxiliary Lemmas

This section contains the proof of Lemma A3 in Cheng, Dou, and Liao (2021) and some other auxiliary results.

Proof of Lemma A3. For any square matrices A_1 and A_2, let $\text{diag}\,(A_1, A_2)$ denote the block diagonal matrix created by aligning the input matrices A_1 and A_2 along the diagonal. Since $\hat{\theta} \in \Theta$ is the minimizer of $g(\theta)'(\hat{\Omega}(\theta))^{-1}g(\theta)$,

$$R(\hat{\xi}) \geq g(\hat{\theta})'\hat{\Omega}^{-1}g(\hat{\theta}) - g_0(\hat{\theta})'\hat{\Omega}_0^{-1}g_0(\hat{\theta}) = g(\hat{\theta})'\left(\hat{\Omega}^{-1} - \text{diag}\left(\hat{\Omega}_0^{-1}, 0_{k_1\times k_1}\right)\right)g(\hat{\theta}), \tag{SC.85}$$

where $\hat{\Omega} \equiv \hat{\Omega}(\hat{\theta}, \hat{\theta})$ and $\hat{\Omega}_0$ is the leading $k_0 \times k_0$ submatrix of $\hat{\Omega}$. Let $\hat{\Omega}_{0,1}$ denote the upper-right $k_0 \times k_1$ submatrix of $\hat{\Omega}$ and $\hat{\Omega}_{1,0} \equiv \hat{\Omega}_{0,1}'$. Since $\hat{\Omega}$ is positive definite,

$$\hat{\Omega}(\hat{\Omega}^{-1} - \text{diag}(\hat{\Omega}_0^{-1}, 0_{k_1\times k_1}))\hat{\Omega} = \text{diag}(0_{k_0\times k_0}, \hat{\Omega}_1 - \hat{\Omega}_{1,0}\hat{\Omega}_0^{-1}\hat{\Omega}_{0,1}) \tag{SC.86}$$

is a positive semi-definite matrix, which together with (SC.85) proves the first claim of the lemma.

To prove the second claim, we first notice that $\hat{m}(\hat{\theta}) = g_0(\hat{\theta}) - \hat{V}(\hat{\theta})g(\hat{\theta}) = 0_{k_0\times 1}$ which implies that for any $\upsilon \in \mathbb{R}^k$

$$L(\upsilon; \hat{d}) \geq \upsilon'\hat{M}\upsilon - \upsilon'\hat{V}_S(\hat{\theta})'\hat{\Omega}_0^{-1}\hat{V}_S(\hat{\theta})\upsilon, \tag{SC.87}$$

where $\hat{V}_S(\theta) \equiv S_0\hat{\Omega}^{1/2}\hat{M}$. Therefore

$$L(x; \hat{d}) \geq \upsilon'\hat{M}\hat{\Omega}^{1/2}\left(\hat{\Omega}^{-1} - \text{diag}(\hat{\Omega}_0^{-1}, 0_{k_1\times k_1})\right)\hat{\Omega}^{1/2}\hat{M}\upsilon \geq 0, \tag{SC.88}$$

where the second inequality holds since the matrix in (SC.86) is positive semi-definite. *Q.E.D.*

Lemma SC3. *For any ξ_1 and ξ_2 with $x_2'\Omega_{d,2}^{-1}x_2 \leq C$ and $\|\xi_1 - \xi_2\|_s \leq C$, where C is a constant, we have $x_1'\Omega_{d,1}^{-1}x_1 \leq C^*$ for some constant C^*, which depends the constant C of the lemma, C_Ω and c_λ in Assumptions 2(iv) and 3(iii) respectively.*

Proof of Lemma SC3. Since $\Omega_{d,1}^{-1}$ is symmetric and positive definite under Assumption 3(iii),

$$\begin{aligned} x_1'\Omega_{d,1}^{-1}x_1 &\leq 2x_2'\Omega_{d,1}^{-1}x_2 + 2(x_1 - x_2)'\Omega_{d,1}^{-1}(x_1 - x_2) \\ &= 2x_2'\Omega_{d,2}^{-1}x_2 + 2x_2'(\Omega_{d,1}^{-1} - \Omega_{d,2}^{-1})x_2 + 2(x_1 - x_2)'\Omega_{d,1}^{-1}(x_1 - x_2) \\ &\leq 2C + 2x_2'(\Omega_{d,1}^{-1} - \Omega_{d,2}^{-1})x_2 + 2(x_1 - x_2)'\Omega_{d,1}^{-1}(x_1 - x_2), \end{aligned} \tag{SC.89}$$

where the second inequality is by $x_2'\Omega_{d,2}^{-1}x_2 \leq C$ as assumed in the lemma. By Assumption 3(iii)

and $\|\xi_1 - \xi_2\|_s \leq C$,

$$(x_1 - x_2)'\Omega_{d,1}^{-1}(x_1 - x_2) \leq (\lambda_{\min}(\Omega_{d,1}))^{-1} \|x_1 - x_2\|^2 \leq C^2 c_\lambda^{-1}. \tag{SC.90}$$

Similarly, by Assumption 3(iii) and $\|\xi_1 - \xi_2\|_s \leq C$,

$$\begin{aligned}\left|x_2'(\Omega_{d,1}^{-1} - \Omega_{d,2}^{-1})x_2\right|^2 &= \left|x_2'\Omega_{d,1}^{-1}(\Omega_{d,1} - \Omega_{d,2})\Omega_{d,2}^{-1}x_2\right|^2 \leq (x_2'\Omega_{d,1}^{-2}x_2)(x_2'\Omega_{d,2}^{-2}x_2)\|\Omega_{d,1} - \Omega_{d,2}\|^2 \\ &\leq \frac{\lambda_{\max}(\Omega_{d,2})(x_2'\Omega_{d,2}^{-1}x_2)^2}{\lambda_{\min}(\Omega_{d,2})(\lambda_{\min}(\Omega_{d,1}))^2}\|\Omega_{d,1} - \Omega_{d,2}\|^2 \leq C^4 c_\lambda^{-3} C_\Omega, \end{aligned} \tag{SC.91}$$

where the last inequality is by $x_2'\Omega_{d,2}^{-1}x_2 \leq C$, Assumptions 2(iv) and 3(iii) and $\|\xi_1 - \xi_2\|_s \leq C$. The claim of the lemma follows from (SC.89)-(SC.91). *Q.E.D.*

Lemma SC4. *For any ξ_1 and ξ_2, define $V_{d,S,j}(\theta) \equiv V_{d,j}(\theta)\,\Omega_{d,j}^{1/2} M_{d,j}$ for $j = 1, 2$. Then we have*

$$\sup_{\theta\in\Theta} \|V_{d,S,1}(\theta) - V_{d,S,2}(\theta)\| \leq C_\Omega^{1/2}(1 + c_\lambda^{-1}C_\Omega^{1/2} + c_\lambda^{-1}C_\Omega)\|\xi_1 - \xi_2\|_s, \tag{SC.92}$$

where C_Ω and c_λ are in Assumptions 2(iv) and 3(iii) respectively.

Proof of Lemma SC4. By definition,

$$\begin{aligned}V_{d,S,1}(\theta) - V_{d,S,2}(\theta) &= [V_{d,1}(\theta) - V_{d,2}(\theta)]\,\Omega_{d,1}^{1/2}M_{d,1} \\ &\quad + V_{d,2}(\theta)(\Omega_{d,1}^{1/2} - \Omega_{d,2}^{1/2})M_{d,1} + V_{d,2}(\theta)\Omega_{d,2}^{1/2}(M_{d,1} - M_{d,2}). \end{aligned} \tag{SC.93}$$

By the properties of $\Omega_{d,j}$ and $M_{d,j}$,

$$\sup_{\theta\in\Theta}\left\|[V_{d,1}(\theta) - V_{d,2}(\theta)]\,\Omega_{d,1}^{1/2}M_{d,1}\right\| \leq C_\Omega^{1/2}\sup_{\theta\in\Theta}\|V_{d,1}(\theta) - V_{d,2}(\theta)\|. \tag{SC.94}$$

By the properties of $V_{d,j}(\theta)$ and $M_{d,j}$, and Exercise 7.2.18 in Horn and Johnson (1990),

$$\begin{aligned}\sup_{\theta\in\Theta}\left\|V_{d,2}(\theta)(\Omega_{d,1}^{1/2} - \Omega_{d,2}^{1/2})M_{d,1}\right\| &\leq \sup_{\theta\in\Theta}\|V_{d,2}(\theta)\|\left\|\Omega_{d,1}^{1/2} - \Omega_{d,2}^{1/2}\right\|_S \\ &\leq c_\lambda^{-1}C_\Omega\left\|\Omega_{d,2}^{-1}\right\|_S\|\Omega_{d,1} - \Omega_{d,2}\| \leq c_\lambda^{-2}C_\Omega\|\Omega_{d,1} - \Omega_{d,2}\|. \end{aligned} \tag{SC.95}$$

Similarly,

$$\sup_{\theta\in\Theta}\left\|V_{d,2}(\theta)\Omega_{d,2}^{1/2}(M_{d,1} - M_{d,2})\right\| \leq c_\lambda^{-1}C_\Omega^{3/2}\|M_{d,1} - M_{d,2}\|. \tag{SC.96}$$

The desirable result follows by (SC.93)-(SC.96) and the triangle inequality. *Q.E.D.*

SD Theoretical Power Properties of the New Test

In this section, we investigate the power properties of the conditional specification test in two cases. First, when the asset pricing moments are globally misspecified, we show that the conditional specification test rejects these moments wpa1, and thus is consistent regardless of the identification strength in the baseline moments. Second, when baseline moments provide strong identification and the asset pricing moments are locally misspecified, we show that the conditional test has the same asymptotic local power as the C test. Thus, it shares the power optimality of the C test in standard scenarios.

Assumption SD1. *The following conditions hold for any $\mathbb{P} \in \mathcal{P}_{1,\infty} \subset \mathcal{P}$:* (i) $\inf_{\theta\in\Theta} \|\mathbb{E}[\bar{g}_1(\theta)]\| > c_{g_1}$ *for some* $c_{g_1} > 0$*;* (ii) $\lambda_{\min}(\Omega_0(\theta_0)) \geq c_\lambda$*,* $\lambda_{\min}(\hat{\Omega}) \geq c_\lambda$ *and* $\lambda_{\min}(\hat{Q}'\hat{Q}) \geq c_\lambda$ *wpa1.*

Assumption SD1(i) implies that there are globally misspecified moments in $\mathbb{E}[\bar{g}_1(\theta_0)] = 0_{k_1\times 1}$. Assumption SD1(ii) requires that the eigenvalues of $\hat{\Omega}$ and $\hat{Q}'\hat{Q}$ are bounded away from zero wpa1. In view of Assumptions 1(ii) and 2(i), this condition holds if the eigenvalues of $\Omega(\theta_1)$ and $Q(\theta_1)Q(\theta_1)'$ are bounded away from zero, where θ_1 denotes the pseudo true value under misspecification. Therefore Assumption SD1(ii) is the counterpart of Assumption 3(iii) under the alternative.

Theorem SD1. *Suppose Assumptions 1, 2 and SD1 hold. For any $\mathbb{P} \in \mathcal{P}_{1,\infty}$, $\mathbb{P}(\mathcal{T} > c_\alpha(\hat{d})) \to 1$ as $n \to \infty$.*

Proof of Theorem SD1. We first show that the test statistic $\mathcal{T}$, written as $R(\hat{\xi})$, diverges at rate n under global misspecification. By Assumptions 2(i, iv) and SD1(ii),

$$\begin{aligned} R(\hat{\xi}) &= \min_{\theta\in\Theta} g(\theta)'(\hat{\Omega}(\theta))^{-1}g(\theta) - \min_{\theta\in\Theta} g_0(\theta)'(\hat{\Omega}_0(\theta))^{-1}g_0(\theta) \\ &\geq (C_\Omega + 1)^{-1} \min_{\theta\in\Theta} \|g(\theta)\|^2 - c_\lambda^{-1} \|g_0(\theta_0)\|^2 \quad \text{wpa1}, \end{aligned} \tag{SD.1}$$

where

$$\|g(\theta)\|^2 \geq \frac{1}{2} \|\mathbb{E}[g(\theta)]\|^2 - \|g(\theta) - \mathbb{E}[g(\theta)]\|^2 . \tag{SD.2}$$

By Assumption SD1(i), there exists a constant $c_{g_1} > 0$ such that $\min_{\theta\in\Theta} \|\mathbb{E}[\bar{g}(\theta)]\|^2 \geq c_{g_1}$, which combined with (SD.1), (SD.2), and Assumptions 1(i) and 2(iv) implies that

$$R(\hat{\xi}) \geq n \left(K^{-1} \min_{\theta\in\Theta} \|\mathbb{E}[\bar{g}(\theta)]\|^2 - o_p(1) \right) \geq n c_{g_1} K^{-1} \text{ wpa1}. \tag{SD.3}$$

The critical value satisfies $c_\alpha(\hat{d}) \leq q_{1-\alpha}(\chi^2_k)$ wpa1, because $L(\upsilon^*; \hat{d}) \leq \upsilon^{*\prime}\hat{M}\upsilon^* \leq \|\upsilon^*\|^2$ wpa1 given that $\hat{M}$ is an idempotent matrix wpa1 under Assumption SD1(ii) and $q_{1-\alpha}(\chi^2_k)$ is the $1-\alpha$

quantile of $\|\upsilon^*\|^2$. Therefore, by (SD.3) and $c_\alpha(\hat{d}) \leq q_{1-\alpha}(\chi^2_k)$ wpa1, we have

$$\mathbb{P}\left(R(\hat{\xi}) > c_\alpha(\hat{d})\right) \geq \mathbb{P}\left(nc_{g_1}K^{-1} > q_{1-\alpha}(\chi^2_k)\right) - o(1), \tag{SD.4}$$

where the right hand side of the above inequality goes to 1 as $n \to \infty$. *Q.E.D.*

The consistency of the conditional specification test holds no matter the parameter θ_0 (or its subvector) is strongly, weakly or not identified by the baseline moments. We next study the local power of the conditional specification test when the baseline moments provide strong identification.

Assumption SD2. *The following conditions hold for any $\mathbb{P} \in \mathcal{P}_{1,A} \subset \mathcal{P}$:*
(i) $\mathbb{E}\left[\bar{g}_1(\theta_0)\right] = an^{-1/2}$ *for some* $a \in \mathbb{R}^{k_1}$ *with* $\|a\| < \infty$*;*
(ii) *Assumptions 3(ii) and 3(iii) hold for any* $\mathbb{P} \in \mathcal{P}_{1,A}$*.*

Theorem SD2. *Suppose Assumptions 1, 2, 4 and SD2 hold. For any $\mathbb{P} \in \mathcal{P}_{00} \cap \mathcal{P}_{1,A}$, we have*

$$\mathbb{P}\left(\mathcal{T} > c_\alpha(\hat{d})\right) \to P\left(\chi^2_{k_1}(a'_\Omega M a_\Omega) > q_{1-\alpha}(\chi^2_{k_1})\right), \text{ as } n \to \infty,$$

where $a_\Omega \equiv \Omega^{-1/2}a$ and $\chi^2_{k_1}(a'_\Omega M a_\Omega)$ denotes a non-central chi-square random variable with degree of freedom k_1 and non-central parameter $a'_\Omega M a_\Omega$.

Proof of Theorem SD2. Under Assumptions 1 and 2, the strong identification in baseline moments in Assumption 4, and the local misspecification in Assumption SD2, $\hat{\theta}$ and $\hat{\theta}_0$ are consistent by the standard arguments and results in (A.19) and (A.20) of Cheng, Dou, and Liao (2021) remain valid. Therefore,

$$R(\hat{\xi}) \to_d (\Omega^{-1/2}\upsilon + a_\Omega)' M (\Omega^{-1/2}\upsilon + a_\Omega) - \upsilon'_0 \Omega_0^{-1/2} M_0 \Omega_0^{-1/2} \upsilon_0, \tag{SD.5}$$

where $a_\Omega \equiv \Omega^{-1/2}a$, υ is a multivariate normal random variable with mean zero and variance Ω, and υ_0 denotes the leading k_0 subvector of υ. By the standard arguments in the GMM literature (e.g., Hall, 2005, Section 5), we have

$$(\Omega^{-1/2}\upsilon + a_\Omega)' M (\Omega^{-1/2}\upsilon + a_\Omega) - \upsilon'_0 \Omega_0^{-1/2} M_0 \Omega_0^{-1/2} \upsilon_0 \sim \chi^2_{k_1}(a'_\Omega M a_\Omega). \tag{SD.6}$$

We next study $c_\alpha(\hat{d})$ under the local misspecification. Since $\hat{\theta}$ is $n^{1/2}$ consistent under the local misspecification, Lemma A7 of Cheng, Dou, and Liao (2021) remains valid for any $\mathbb{P} \in \mathcal{P}_{00} \cap \mathcal{P}_{1,A}$. Therefore, for any $\mathbb{P} \in \mathcal{P}_{00} \cap \mathcal{P}_{1,A}$,

$$L(\upsilon^*, \hat{d}) = \upsilon^{*\prime}(M - \tilde{M}_0)\upsilon^* + o_p(1) \sim \chi^2_{k_1}. \tag{SD.7}$$

By (SD.7) and arguments analogous to those used to show Theorem 2(ii), we have $c_\alpha(\hat{d}) \to_p q_{1-\alpha}(\chi^2_{k_1})$, which together with (SD.5) and (SD.6) proves the claim of the theorem. *Q.E.D.*

As long as $a'_\Omega M a_\Omega > 0$, we have $P\left(\chi^2_{k_1}(a'_\Omega M a_\Omega) > q_{1-\alpha}(\chi^2_{k_1})\right) > \alpha$. Moreover, this probability is strictly increasing in the non-central parameter $a'_\Omega M a_\Omega$. If the baseline moments $\mathbb{E}\left[\bar{g}_0(\theta_0)\right] = 0_{k_0\times 1}$ only depend on a subvector $\theta_{c,0}$ of θ_0 with dimension d_c and strongly identify $\theta_{c,0}$, arguments analogous to those used to show Theorem SD2 give

$$\mathbb{P}\left(\mathcal{T} > c_\alpha(\hat{d})\right) \to P\left(\chi^2_{k_1+d_c-d_\theta}(a'_\Omega M a_\Omega) > q_{1-\alpha}(\chi^2_{k_1+d_c-d_\theta})\right) \text{ as } n \to \infty. \tag{SD.8}$$

When the baseline moments provide strong identification, the conditional specification test is asymptotically equivalent to the C test following Theorems 2, SD1 and SD2.[1] In particular, it shares the same (asymptotic) local power function with the C test and thus achieves optimality under local misspecification (Newey, 1985). Nevertheless, the conditional specification test compares favorably to the C test for its correct asymptotic size even with weak identification in the baseline moments, an important property for its applications to many macro-finance asset pricing models.

SE Comparison to Some Power Envelopes

In this section, we derive some power envelopes in a Gaussian experiment as in Section 4.1 of Cheng, Dou, and Liao (2021). These power envelopes are akin to those in Section 3.4 of Andrews and Mikusheva (2016). We compare the power of the proposed conditional specification test to these power envelopes through simulation studies.

Setup. We observe (i) a Gaussian process $g_{0,\infty}(\cdot)$ with covariance matrix $\Omega_0(\cdot,\cdot)$, and (ii) a Gaussian random vector $g_\infty(\hat{\theta})$ which satisfies

$$g_\infty(\hat{\theta}) \equiv (I_k - Q(Q'\Omega^{-1}Q)^{-1}Q'\Omega^{-1})g_\infty(\theta_0) = \Omega^{1/2}M\Omega^{-1/2}g_\infty(\theta_0), \tag{SE.1}$$

where $g_\infty(\theta_0) \equiv (g_{0,\infty}(\theta_0)', g_{1,\infty}(\theta_0)')'$ is normal with covariance matrix Ω, $g_{0,\infty}(\theta_0)$ and $g_{1,\infty}(\theta_0)$ are $k_0 \times 1$ and $k_1 \times 1$ respectively, $Q \equiv (Q'_0, Q'_1)'$, Q_0 and Q_1 are $k_0 \times d_\theta$ and $k_1 \times d_\theta$ ($k_1 \geq d_\theta$) matrices respectively. We assume that Q_1 has full rank, and $\Omega_0(\cdot,\cdot)$, Ω, Q and the covariance between $g_{0,\infty}(\cdot)$ and $g_\infty(\theta_0)$ are known.

We are interested in testing

$$H_0 : \eta = 0_{k_1\times 1} \text{ where } \eta \equiv \mathbb{E}[g_{1,\infty}(\theta_0)], \tag{SE.2}$$

[1] See e.g., Hall (2005) for detailed derivations for the C test.

while maintaining $\mathbb{E}[g_{0,\infty}(\theta_0)] = 0_{k_0\times 1}$ under both the null and the alternative hypotheses. The alternative hypothesis is written as

$$H_1 : \eta \neq 0. \tag{SE.3}$$

The true value of θ_0 is unknown under both the null and the alternative.

Power Envelopes. Let $Q^{\perp}$ denote the orthogonal complement of Q. It is clear that $Q'\Omega^{-1}$ and $Q^{\perp\prime}$ are the left eigenvectors of $I_k - Q(Q'\Omega^{-1}Q)^{-1}Q'\Omega^{-1}$ with respect to the (left) eigenvalues 0 and 1 respectively. Let $D = (Q^{\perp}, \Omega^{-1}Q)'$, then D is non-singular. Moreover,

$$Dg_{\infty}(\hat{\theta}) = D(I_k - Q(Q'\Omega^{-1}Q)^{-1}Q'\Omega^{-1})g_{\infty}(\theta_0) = \begin{pmatrix} Q^{\perp\prime}g_{\infty}(\theta_0) \\ 0_{d_\theta\times k} \end{pmatrix}. \tag{SE.4}$$

Based on (SE.4), observing $g_{\infty}(\hat{\theta})$ is equivalent to observing

$$Y \equiv Q^{\perp\prime}g_{\infty}(\theta_0) \sim N\left(A(\eta), Q^{\perp\prime}\Omega Q^{\perp}\right), \text{ where } A(\eta) \equiv Q^{\perp\prime}\begin{pmatrix} 0_{k_0\times 1} \\ \eta \end{pmatrix}. \tag{SE.5}$$

We next consider inference of η based only on Y.

Since the uniformly most powerful (UMP) test does not exists for (SE.3), we follow Andrews and Mikusheva (2016) and derive several power envelopes by reducing the alternative hypothesis (SE.3) and/or imposing restrictions on the class of tests. If the alternative hypothesis (SE.3) is reduced to a single value η^* with $A(\eta^*) \neq 0$, then the Neyman-Pearson lemma implies that the UMP test rejects H_0 if

$$\frac{A(\eta^*)'(Q^{\perp\prime}\Omega Q^{\perp})^{-1}Y}{(A(\eta^*)'(Q^{\perp\prime}\Omega Q^{\perp})^{-1}A(\eta^*))^{1/2}} > z_{1-\alpha}, \tag{SE.6}$$

where $z_{1-\alpha}$ denotes the $1-\alpha$ quantile of the standard normal distribution. It is clear that the optimality of the test in (SE.6) depends on η^* by construction. Its power may be low if the true value η under the alternative is different from η^*. In the simulation study below, we let the test in (SE.6) depend on the true value under the alternative η (i.e., we replace η^* by η) and call its power (as a function of η) as PE-1. Next, we consider a subset of alternative hypothesis (SE.3) which are proportional to a known vector η^* with $A(\eta^*) \neq 0$, i.e., $H_1 : \eta = a\eta^*$. Since η^* is known, the subset of alternative hypothesis becomes $H_1 : a \neq 0$. As noticed in Andrews and Mikusheva (2016), the UMP unbiased test for this reduced problem rejects H_0 if

$$\left|\frac{A(\eta^*)'(Q^{\perp\prime}\Omega Q^{\perp})^{-1}Y}{(A(\eta^*)'(Q^{\perp\prime}\Omega Q^{\perp})^{-1}A(\eta^*))^{1/2}}\right| > z_{1-\alpha/2}. \tag{SE.7}$$

In the simulation study below, we let the test in (SE.7) also depend on the true value η under the alternative and call its power (as a function of η) as PE-2. Both PE-1 and PE-2 are infeasible

because they require the knowledge of the true value η under the alternative. Finally, the following feasible test

$$Y'(Q^{\perp\prime}\Omega Q^{\perp})^{-1}Y > q_{1-\alpha}(\chi^2_{k-d_\theta}) \tag{SE.8}$$

is the UMP invariant test, whose power function is called PE-3. This is equivalent to the J-test.

Conditional Specification Test. In this setup, the test statistic $\mathcal{T}$ in the paper is the QLR statistic written as

$$\mathcal{T} \equiv g_\infty(\hat{\theta})'\Omega^{-1}g_\infty(\hat{\theta}) - \min_{\theta\in\Theta} g_{0,\infty}(\theta)'(\Omega_0(\theta))^{-1}g_{0,\infty}(\theta), \tag{SE.9}$$

where $\Omega_0(\theta) \equiv \Omega_0(\theta,\theta)$. We apply the conditional inference based on this test statistic. Define

$$m_{0,\infty}(\theta) \equiv g_{0,\infty}(\theta) - V(\theta)g_\infty(\hat{\theta}), \tag{SE.10}$$

where $V(\theta) \equiv \mathrm{Cov}(g_{0,\infty}(\theta), g_\infty(\theta_0))\Omega^{-1}$ is a known function of θ. Then under the null hypothesis, $\mathrm{Cov}(m_{0,\infty}(\theta), g_\infty(\hat{\theta})) = 0$ which implies that $m_{0,\infty}(\theta)$ and $g_\infty(\hat{\theta})$ are independent by their joint normal distribution. The conditional inference is conducted using the critical value of

$$\mathcal{T} = g_\infty(\hat{\theta})'\Omega^{-1}g_\infty(\hat{\theta}) - \min_{\theta\in\Theta}(m_{0,\infty}(\theta) + V(\theta)g_\infty(\hat{\theta}))'(\Omega_0(\theta))^{-1}(m_{0,\infty}(\theta) + V(\theta)g_\infty(\hat{\theta})) \tag{SE.11}$$

conditioning on $m_{0,\infty}(\theta)$.

Simulation. Next, we compare the power of the proposed test with the three power envelopes through simulation studies. To this end, we consider a specific example where $d_\theta = 1$, $k_0 = qk_1$, $k_1 = 2$, and

$$g_{0,\infty}(\theta) \equiv g_{0,\infty}(\theta_0) + (\theta - \theta_0)Q_{0,\infty} \tag{SE.12}$$

where $Q_{0,\infty}$ is a $k_0 \times 1$ random vector. The distribution of the random vector $(g_\infty(\theta_0)', Q'_{0,\infty})'$ is specified as follows:

$$\begin{pmatrix} g_{0,\infty}(\theta_0) \\ g_{1,\infty}(\theta_0) \\ Q_{0,\infty} \end{pmatrix} \sim N\left(\begin{pmatrix} 0_{k_0\times 1} \\ \eta \\ \mu_g \end{pmatrix}, \Sigma\right) \text{ where } \Sigma \equiv \begin{pmatrix} \Omega_{00} & \Omega_{01} & \Omega_{0g} \\ \Omega_{10} & \Omega_{11} & \Omega_{1g} \\ \Omega_{g0} & \Omega_{g1} & \Omega_{gg} \end{pmatrix} \tag{SE.13}$$

where μ_g is a $k_0 \times 1$ real vector. We shall consider two cases for $Q_{0,\infty}$. In the first case, $Q_{0,\infty}$ is a nonrandom vector as in the simple disaster risk model in Section 2 of Cheng, Dou, and Liao (2021). In this case, Ω_{gg}, Ω_{g1}, Ω_{1g}, Ω_{0g} and Ω_{g0} are zero matrices, and $Q_{0,\infty} = \mu_g$. In the second case, $Q_{0,\infty}$ is a non-degenerate normal random vector.

By the definition of $g_{0,\infty}(\theta)$ and the joint distribution of $(g_\infty(\theta_0)', Q'_{0,\infty})'$, $\Omega_0(\theta)$ and $V(\theta)$, both

of which show up in the conditional specification test, take the following form:

$$\Omega_0(\theta) = \Omega_{00} + (\theta - \theta_0)(\Omega_{0g} + \Omega_{g0}) + (\theta - \theta_0)^2 \Omega_{gg},$$

$$V(\theta) = \begin{pmatrix} \Omega_{00} & \Omega_{01} \end{pmatrix} \Omega^{-1} + (\theta - \theta_0) \begin{pmatrix} \Omega_{g0} & \Omega_{g1} \end{pmatrix} \Omega^{-1}, \text{ where } \Omega = \begin{pmatrix} \Omega_{00} & \Omega_{01} \\ \Omega_{10} & \Omega_{11} \end{pmatrix}. \quad \text{(SE.14)}$$

We generate the covariance matrix Ω as follows

$$\Omega = \begin{pmatrix} (1+\rho^2)^{-1}(I_{k_0} + \rho^2 1_{q\times q} \otimes \Omega_u) & \rho 1_{q\times 1} \otimes \Omega_u \\ \rho 1_{1\times q} \otimes \Omega_u & \Omega_u \end{pmatrix} \text{ where } \Omega_u = \begin{pmatrix} 1 & \lambda \\ \lambda & 1 \end{pmatrix}, \quad \text{(SE.15)}$$

where $A \otimes B$ denotes the Kronecker product of two real matrices A and B. The parameter λ determines the correlation between the two moments in $g_{1,\infty}(\theta_0)$, while ρ mainly controls the correlations between moments in $g_{0,\infty}(\theta_0)$ and $g_{1,\infty}(\theta_0)$. In the case that $Q_{0,\infty}$ is a non-degenerate normal random vector, we let

$$\Omega_{gg} = I_{k_0}, \;\; \Omega_{0g} = \Omega'_{g0} = \lambda I_{k_0} \;\text{ and }\; \Omega_{1g} = \Omega'_{g1} = \lambda 1_{1\times q} \otimes I_{k_1} \quad \text{(SE.16)}$$

where we also use λ to control the correlation between $Q_{0,\infty}$ and $g_\infty(\theta_0)$.

Throughout this simulation, we let $\theta_0 = 0$, $\Theta = [-1, 1]$, $Q_0 = c_g 1_{k_0\times 1}$, $Q_1 = (j^{-1})_{j=1,\ldots,k_1}$, $\mu_g = c_\mu 1_{k_0\times 1}$, $\eta = a 1_{k_1\times 1}$, and $\lambda = 0.1$. We consider a benchmark case and three deviations from the benchmark, which are defined as follows.

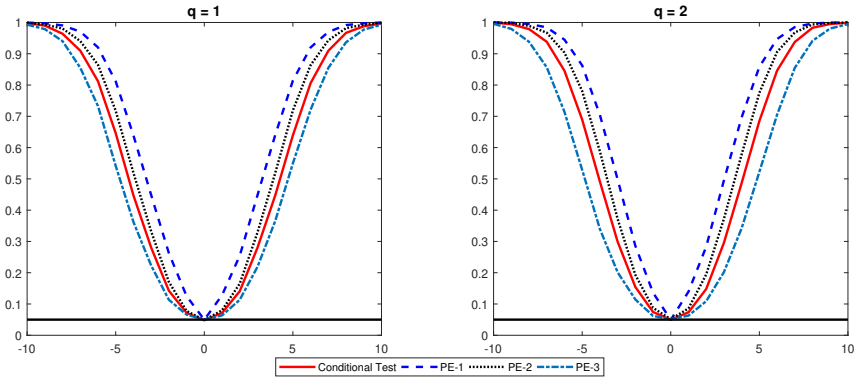
Benchmark case:	$\rho = 0.4$, $c_g = 0$, $c_\mu = 1$, nonrandom $Q_{0,\infty}$, $q = k_0/k_1 = 1$ or 2;
Deviation case 1:	$\rho = 0.2$ or 0.8, and $q = 2$;
Deviation case 2:	$c_g = 0.1$;
Deviation case 3:	random $Q_{0,\infty}$.

In the benchmark case, we set $c_g = 0$ to model weak baseline moments whose derivatives Q_0 are 0 in the limiting experiment. The deviation cases enable us to investigate how the power properties of the conditional test change when: (1) the baseline moments and the asset pricing moments have correlation; (2) the baseline moments provide non-trivial identification when combined with the asset pricing moments; (3) the matrix $Q_{0,\infty}$ is random. The simulation results in the benchmark case, and in the 3 deviation cases are presented in Figure 1 and Figures 2-4, respectively. We plot the finite-sample rejection probability against a, where $\eta = (a, a)'$ under the alternative. All the results are calculated with 10,000 simulation replications.

Discussion. In all cases, the power of the proposed conditional specification test is between PE-2 and PE-3 (J-test). PE-2 is the power of the UMP unbiased test with respect to a smaller subset of

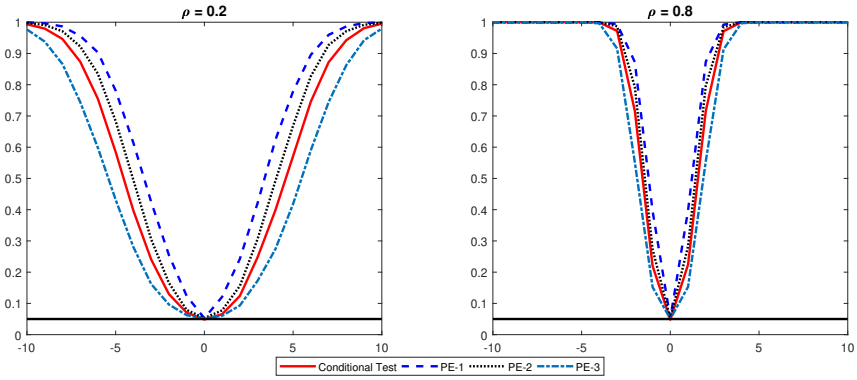
the general alternative hypothesis in (SE.3) and it is constructed using the true alternative value η, whereas the conditional specification test does not require such information. Simulation results show that the power function of the conditional specification test is rather close to PE-2 in many cases with a substantial improvement from PE-3. The benchmark case in Figure 1 shows that increasing the number of baseline moments significantly enlarges the power gain compared to PE-3 while roughly maintains the same amount of power loss compared to PE-2. Figure 2 and Figure 3 show that increasing the correlation between the baseline moments and the asset-pricing moments, or increasing the identification strength of the baseline moments to the structural parameter make all powers higher and reduce the power difference between the conditional specification test and PE-2. Figure 4 shows that reducing the signal-to-noise ratio in the baseline moments results in a larger gap between the power of the conditional specification test and PE-2. Nevertheless, we still see noticeable improvement over PE-3 in the two scenarios of this case.

Figure 1: Power Comparison in the Benchmark Case



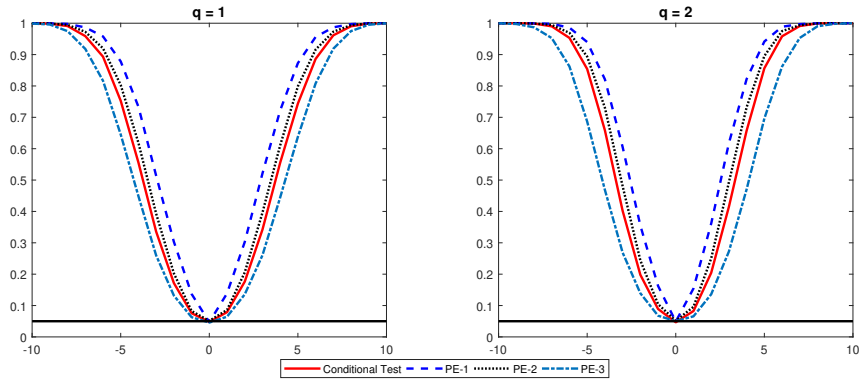
Note: In the Benchmark case, we have $\rho = 0.4$, $c_g = 0$, $c_\mu = 1$, nonrandom $Q_{0,\infty}$, $q = k_0/k_1 = 1$ or 2.

Figure 2: Power Comparison in the Deviation Case 1



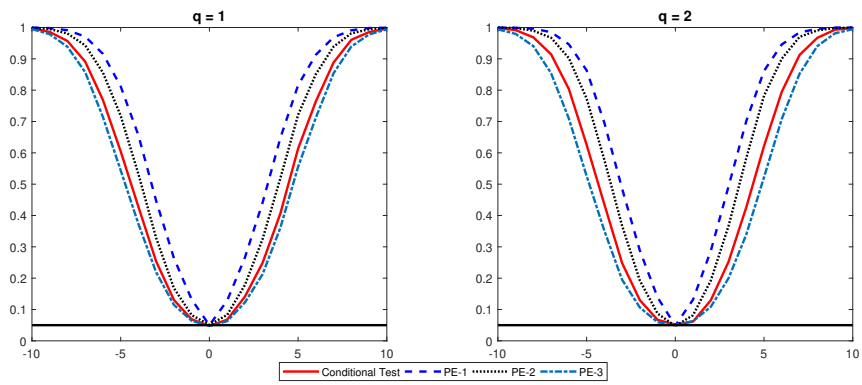
Note: In the deviation case 1, we have $\rho = 0.2$ or 0.8, $c_g = 0$, $c_\mu = 1$, non-random $Q_{0,\infty}, q = 2$.

Figure 3: Power Comparison in the Deviation Case 2



Note: In the deviation case 2, we have $c_g = 0.1$, $\rho = 0.4$, $c_\mu = 1$, non-random $Q_{0,\infty}$, $q = 1$ or 2.

Figure 4: Power Comparison in the Deviation Case 3



Note: In the deviation case 4, we have random $Q_{0,\infty}$, $\rho = 0.4$, $c_g = 0$, $c_\mu = 1$, and $q = 1$ or 2.

SF Additional Details of the Empirical Application

We have 8 baseline moment conditions $\mathbb{E}[\bar{g}_0(\theta)] = 0_{8\times 1}$ when $\theta = \theta_0$, where $\theta \equiv (\theta_1, \ldots, \theta_4)$ is the reparametrized parameter defined as

$$\theta_1 \equiv \frac{p}{\alpha - \gamma},\ \theta_2 \equiv \frac{\sigma_p^2}{1-\rho^2},\ \theta_3 \equiv \rho, \text{ and } \theta_4 \equiv \gamma. \tag{SF.1}$$

In the model, $\bar{g}_0(\theta)$ only depends on a subvector of θ. We have 6 asset pricing moment conditions $\mathbb{E}[\bar{g}_1(\theta_0)] = 0_{6\times 1}$ where $\bar{g}_1(\theta)$ depends on all the components in θ.

We consider the following calibrated values for the nuisance parameters

$$(\delta,\ g_c,\ g_d,\ \sigma_c,\ \phi,\ \underline{\upsilon},\ q) = (0.97,\ 0.02,\ 0.02,\ 0.02,\ 3.5,\ 0.07,\ 0.4). \tag{SF.2}$$

We consider $p \in \{0.3\%, 0.5\%, 0.7\%, 0.9\%, 1.1\%\}$ where $p = 0.7\%$ is our benchmark case and the

other four values of p are used for the robustness check. The parameter space Θ for the unknown parameter is set to $\Theta \equiv \Theta_1 \times \Theta_2 \times \Theta_3 \times \Theta_4$ where

$$\Theta_1 \equiv [0.001, 0.02],\ \Theta_2 \equiv [5, 12],\ \ \Theta_3 \equiv [0.95, 0.999], \text{ and } \ \Theta_4 \equiv [3, 6]. \tag{SF.3}$$

To compute the CUE estimator, the J statistic and the statistic of the conditional specification test, we search through equally spaced grid points with step size (i.e., the distance between two adjacent points) 0.001 in Θ_1 and Θ_3, and step size 0.01 in Θ_2 and Θ_4.[2] The critical values of the conditional specification test are simulated using $B = 2500$ Gaussian random vectors.

To calculate the model uncertainty set for $p \in \{0.5\%,\ 0.7\%,\ 0.9\%\}$ we consider a smaller parameter space $\Theta_2 \equiv [5, 8]$ and a larger step size 0.1 of the grid points in Θ_2 and Θ_4 to reduce the computational cost. The parameter spaces $\Theta_j (j = 1, 3, 4)$ and the grid points in Θ_1 and Θ_3 are unchanged. The reduced space Θ_2 still covers the CUE estimators of θ_2 for the three values of p considered. The model uncertainty sets of (η_1, η_3) and (η_3, η_4) are calculated through grid search with equally spaced grid points for η_j $(j = 1, 3, 4)$ with step size 0.001. The parameter spaces for η_j $(j = 1, 3, 4)$ are set large enough such that the model uncertainty sets from the J test are contained in the interior of these parameter spaces.

References

Andrews, I., and A. Mikusheva (2016): "Conditional Inference with a Functional Nuisance Parameter," *Econometrica*, 84(4), 1571–1612.

Cheng, X., W. W. Dou, and Z. Liao (2021): "Macro-Finance Decoupling: Robust Evaluations of Macro Asset Pricing Models," *working paper, UCLA and University of Pennsylvania.*

Hall, A. R. (2005): *Generalized Method of Moments.* Oxford university press.

Horn, R. A., and C. R. Johnson (1990): *Matrix Analysis.* Cambridge university press.

Newey, W. K. (1985): "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29(3), 229–256.

[2] We have also considered a much larger parameter space with $\Theta_1 \equiv [0.001, 1]$, $\Theta_2 \equiv [1, 15]$, $\Theta_3 \equiv [0.9, 0.999]$, $\Theta_4 \equiv [1, 20]$, and step size 0.002, 0.2, 0.001, 0.2, respectively. The results on the CUE estimators and the J tests are similar to those reported in Table 1 of the paper.