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

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

May 24, 2020

Abstract

This paper shows that robust inference under weak identification is important to the evaluation of many influential macro asset pricing models, including long-run risk models, disaster risk models, and multifactor linear asset pricing models. Building on recent developments in the conditional inference literature, we provide a new 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 our specification test when asset pricing theories are difficult to refute due to an imbalance in the information content about the key model parameters between macroeconomic moment restrictions and asset pricing cross-equation restrictions.

Keywords: Asset Pricing, Conditional Inference, Disaster Risk, Long-Run Risk, Factor Models, Specification Test, Weak Identification.
JEL Classification: C12, C32, C52, G12.

1 Introduction

Many influential macro asset pricing models, either structural or reduced form, have been developed and widely used by researchers, practitioners, and monetary authorities (for recent reviews,

---

[†]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.ecla.edu.

see Cochrane, 2017; Dou, Fang, Lo, and Uhlig, 2020). However, model evaluation is challenging when many models are seemingly able to match any asset pricing and macroeconomic moments of interest, either because of the "dark matter" that captures severe information imbalance (Chen, Dou, and Kogan, 2020) or because of the reduced rank with spurious factors (Kan and Zhang, 1999a; Gospodinov, Kan, and Robotti, 2017). Lewellen, Nagel, and Shanken (2010) argue that the existing standard specification tests overenthusiastically support a large number of linear factor models (see also Nagel and Singleton, 2011; Daniel and Titman, 2012; Nagel, 2013), and Gospodinov, Kan, and Robotti (2017) show that the power of the J test (Hansen, 1982; Hansen and Singleton, 1982) could be as small as the size of the test in their model with spurious factors. The econometric question of how to construct a robust and efficient evaluation of such models, including nonlinear structural models, is of great importance for the asset pricing literature.

To address this crucial question, this paper provides a specification test robust to information imbalance or a lack of full rank in a unified weak identification framework. We adopt recent developments in conditional inference and extend it to model evaluation. The proposed test is built on a sufficient statistic, which can be interpreted as a measure capturing macroeconomic information decoupled from underlying content of asset pricing theories. Such decoupling is crucial for preserving limited macroeconomic information and efficiently using it to evaluate the asset pricing moment restrictions that are difficult to refute. This proposed test is more robust than the C test, which is also referred to as the incremental J test and was proposed by Eichenbaum, Hansen, and Singleton (1988), in size, more efficient than the J test in power, and becomes equivalent to the optimal test in the classical scenario without weak identification (Newey, 1985).

We contribute to both the econometrics and the asset pricing literatures. First, we demonstrate that weak identification robust inference can be fruitfully applied to many influential macro asset pricing models, beyond the scope of existing investigation (for reviews on weak instruments and weak identification, see Stock, Wright, and Yogo, 2002; Andrews, Stock, and Sun, 2020). In the generalized method of moments (GMM) setup, we show that the long-run risk model (Bansal and Yaron, 2004), the disaster risk model (Barro, 2006), and the multifactor linear asset pricing model (Ross, 1976; Fama and French, 1996) 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 parameters, characterized by near flatness of the baseline moments (Stock and Wright, 2000). The remainder of the moments are asset-pricing moments implied by the specific theory, which by design provide tighter cross-equation restrictions and thus strong identification of the parameters. Identification-robust inference methods are desirable in these models because, even in large sample, estimators for weakly identified parameters are poorly approximated by the normal

distribution (e.g., Staiger and Stock, 1997; Stock and Wright, 2000; Andrews and Cheng, 2012).

Second, our approach is developed based on the recent advances in conditional inference, and we extend the conditional inference to specification tests of nonlinear structural models. Conditional inference has been successful in constructing confidence sets for weakly identified parameters, following the pioneering work of Moreira (2003) for linear instrumental variable (IV) models. Kleibergen (2005) broadens its application to nonlinear GMM models. Furthermore, Andrews and Mikusheva (2016a) provide a new perspective in viewing the near-flat population moment function as a functional nuisance parameter. Doing so enables a more powerful test based on the quasi-likelihood ratio (QLR) statistic. We build on the profound idea of Andrews and Mikusheva (2016a) and show how to use it to evaluate economic models, rather than draw statistical inference on specific parameters.

Third, we contribute new econometric tools to the empirical toolbox for financial economists. A growing literature is concerned with the efficacy of conventional methods for macro asset pricing models and the development of robust methods, including Gospodinov, Kan, and Robotti (2014), Burnside (2015), and Kleibergen and Zhan (2015, 2020) for linear asset pricing models and Stock and Wright (2000), Kleibergen (2005), Andrews (2016), Andrews and Mikusheva (2016a), and Andrews and Guggenberger (2019) for applications to nonlinear models. Under a general semiparametric framework, Chen, Dou, and Kogan (2020) formalize the idea of "dark matter" to model components that are difficult to measure directly in the data, despite having significant effects on the model's main implications. The authors show that high dark matter leads to low reputability and poor out-of-sample performance. Hodrick and Zhang (2001) also emphasize that the evaluation of asset pricing models is plagued by a lack of power to reject mis-specified models using the standard test procedures. Furthermore, in linear predictive models of stock returns, the standard asymptotic inference and test fail, when the predictor variable is highly persistent and its innovations are highly correlated with returns (Elliott and Stock, 1994; Mankiw and Shapiro, 1986; Stambaugh, 1999), and new methods have been proposed to provide valid and efficient procedures to correct this problem (e.g., Campbell and Yogo, 2006; Elliott, Müller, and Watson, 2015). Kelly, Pruitt, and Su (2019) and Feng, Giglio, and Xiu (2020) apply machine learning methods to evaluate linear asset pricing models. Nevertheless, the existing literature lacks reliable and powerful model evaluation methods in the presence of information imbalance, as modeled here by weak identification. This paper tackles this challenging yet important issue in the asset pricing literature.

The rest of the paper is organized as follows. Section 2 studies simple variants of three popular asset pricing models to illustrate the weak identification issue in baseline moments and the

mechanism through which asset pricing theories provide strong identification. Section 3 provides the general framework, the algorithm to conduct the test, and its validity in a linear Gaussian model, which captures the essence of the method. Sections 4 and 5 study the asymptotic size and power of the test in the general nonlinear GMM framework. Section 6 gives simulation results for all three motivating examples. Section 7 concludes. The appendix offers proofs. Cheng, Dou, and Liao (2020), available on SSRN and the authors' personal websites, contains a supplemental appendix with some additional supporting results.

2   Motivating Examples

2.1   A Long-Run Risk Model for the 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), U.S. real output growth contains 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, in the real output growth (e.g., Beeler and Campbell, 2012; Bansal, Kiku, and Yaron, 2012; Schorfheide, Song, and Yaron, 2018).

The long-run component of real output growth, x_t, is latent and follows

$$x_t = \rho x_{t-1} + \varepsilon_{x,t}. \tag{2.1}$$

The representative agent's consumption has the following log growth process:[1]

$$\Delta c_t \equiv \ln\left(\frac{C_t}{C_{t-1}}\right) = \phi x_{t-1} + \sigma_c \varepsilon_{c,t}, \tag{2.2}$$

where C_t is consumption per capita. The shocks $(\varepsilon_{x,t}, \varepsilon_{c,t})$ follow a standard multivariate normal distribution and are i.i.d. over t. By introducing parameter ϕ in (2.2), 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 ϕ is near zero, the consumption growth process is nearly i.i.d. In the baseline model of Bansal and Yaron (2004, Table I), the loading parameter ϕ is effectively 0.034% in monthly frequency. The specification of Δc_t in (2.2) provides

---

[1]We ignore the intercept in Δc_t to maintain the example's simplicity, because it plays little role in explaining the equity premium.

the baseline moment conditions

$$\mathbb{E}\left[m_{0,t}(\rho)\right] = 0, \quad \text{where } m_{0,t}(\rho) \equiv \begin{bmatrix} \Delta c_{t-1}\left(\Delta c_{t+1} - \rho \Delta c_t\right) \\ \\ \Delta c_t\left(\Delta c_{t+1} - \rho \Delta c_t\right) + \rho \sigma_c^2 \end{bmatrix}. \tag{2.3}$$

For illustrative purpose, we assume that econometricians know 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.[2,3]

The representative agent has recursive preferences as in Epstein and Zin (1989) and Weil (1989), and maximizes his lifetime utility

$$V_t = \left[(1-\delta)C_t^{1-1/\psi} + \delta\left(\mathbb{E}_t\left[V_{t+1}^{1-\gamma}\right]\right)^{\frac{1-1/\psi}{1-\gamma}}\right]^{\frac{1}{1-1/\psi}}, \tag{2.4}$$

where δ is the rate of time preference, γ is the coefficient of risk aversion for timeless gambles, and ψ is the elasticity of intertemporal substitution under certainty. The Euler equation for the utility maximization problem requires that the equilibrium excess log return r_t^e satisfies

$$\mathbb{E}\left[m_{1,t}(\rho)\right] = 0, \quad \text{where } m_{1,t}(\rho) \equiv r_t^e - \frac{1}{2}\left(2\gamma - \psi^{-1} - 1\right)\left(1 - \psi^{-1}\right)\frac{\phi^2}{(1-\rho)^2}. \tag{2.5}$$

We call this the asset pricing moment.

The key insight of the long-run risk model can be clearly seen from (2.5): 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 cash flows load on the long-run component (i.e., ϕ is positive) and the long-run component is persistent (i.e., ρ is close to unity). To ensure that the long-run risk is a meaningful economic mechanism for explaining the sizeable equity premium, $\phi/(1-\rho)$ must be a positive component that is bounded away from zero and from above in order to match the moment of r_t^e. To utilize this insight, we transform ρ to θ where

$$\theta \equiv \frac{1-\rho}{\phi} \quad \text{and} \quad \theta \in \Theta \equiv \{\theta \in [\underline{c}, \overline{c}] \text{ and } \theta\phi \in (0,1]\}, \tag{2.6}$$

for some constants $0 < \underline{c} < \overline{c}$. Our analysis allows ρ to be arbitrarily close to 1 and ϕ to be arbitrarily close to 0, while keeping the ratio θ to be bounded from above and below.

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

$$g_{0,t}(\theta) \equiv m_{0,t}(1-\theta\phi) \quad \text{and} \quad g_{1,t}(\theta) \equiv m_{1,t}(1-\theta\phi). \tag{2.7}$$

---

[2] The parameter ϕ, when unknown, can be strongly identified by the baseline moments.

[3] In practice, one could add more moments to the lists of baseline moments and asset pricing moments and estimate more parameters in all three motivating examples. However, additional moments do not change the nature of the problem, and the lists provided here sufficiently illustrate the key idea.

For the baseline moments $g_{0,t}(\theta)$, simple calculations give

$$\mathbb{E}\left[\frac{d}{d\theta}g_{0,t}(\theta)\right] = \frac{\phi^2}{(2-\theta\phi)\theta}\left(1-\theta\phi,1\right)'. \tag{2.8}$$

Since θ is bounded, we have

$$\lim_{\phi\to 0}\mathbb{E}\left[\frac{d}{d\theta}g_{0,t}(\theta)\right] = 0 \quad \text{and} \quad \lim_{\phi\to 0}\mathbb{E}\left[\frac{d}{d\theta}g_{1,t}(\theta)\right] = -\frac{\left(2\gamma-\psi^{-1}-1\right)\left(1-\psi^{-1}\right)}{\theta^3} \neq 0. \tag{2.9}$$

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

2.2 A Disaster Risk Model for the Equity Premium

We consider a simple variant of Barro (2006), Barro (2009), and Barro and Ursúa (2012) with macroeconomic disasters characterized by extremely large consumption declines. We assume that real consumption growth follows

$$\Delta c_t \equiv \ln\left(\frac{C_t}{C_{t-1}}\right) = \sigma\varepsilon_t - \zeta_t, \tag{2.10}$$

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

$$\zeta_t \equiv x_t(\underline{v} + J_t), \tag{2.11}$$

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 i.i.d. over t and contemporaneously mutually independent. Specification of Δc_t in (2.10) provides baseline moment conditions:

$$E[m_{0,t}(\alpha)] = 0, \quad \text{where } m_{0,t}(\alpha) \equiv \begin{bmatrix} \Delta c_t \\ \Delta c_t^2 \end{bmatrix} + \mu(\alpha) \text{ with } \mu(\alpha) \equiv \begin{bmatrix} p\mu_1(\alpha) \\ -\sigma^2 - p\mu_2(\alpha) \end{bmatrix}, \tag{2.12}$$

and $\mu_j(\alpha) \equiv \mathbb{E}[(\underline{v} + J_t)^j] > 0.$[4] 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.[5]

The representative agent maximizes his lifetime expected utility:

$$U_0 \equiv \mathbb{E}\left[\sum_{t=0}^{\infty} e^{-\delta t}\frac{C_t^{1-\gamma}}{1-\gamma}\right], \tag{2.13}$$

---

[4]Specifically, $\mu_1(\alpha) = \underline{v} + 1/\alpha$ and $\mu_2(\alpha) = \underline{v}^2 + 2\underline{v}/\alpha + 2/\alpha^2.$
[5]The parameter p, when unknown, can be strongly identified by the baseline moments.

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:

$$\mathbb{E}[m_{1,t}(\alpha)] = 0, \quad \text{with } m_{1,t}(\alpha) \equiv r_t^e + p(\underline{v} + \alpha^{-1}) - \frac{p}{\alpha - \gamma} h(\alpha) \text{ and}$$

$$h(\alpha) \equiv \alpha \left[e^{\gamma \underline{v}} - \frac{\alpha - \gamma}{\alpha - \gamma + 1} e^{(\gamma - 1)\underline{v}} \right]. \tag{2.14}$$

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

The asset pricing moment (2.14) 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_t^e. This 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

$$\theta \equiv \frac{\alpha - \gamma}{p} \quad \text{and} \quad \theta \in \Theta \equiv [\underline{c}, \overline{c}], \tag{2.15}$$

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

To parameterize all the moments in θ, write

$$g_{0,t}(\theta) \equiv m_{0,t}(\theta p + \gamma) \quad \text{and} \quad g_{1,t}(\theta) \equiv m_{1,t}(\theta p + \gamma). \tag{2.16}$$

Let $\mu_j^{(1)}(\alpha) \equiv (d/d\alpha)\mu_j(\alpha)$ for $j = 1, 2$. For the baseline moments $g_{0,t}(\theta)$, simple calculations give

$$\frac{d}{d\theta} g_{0,t}(\theta) = p^2 \left(\mu_1^{(1)}(\theta p + \gamma), \; -\mu_2^{(1)}(\theta p + \gamma) \right)', \tag{2.17}$$

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

$$\lim_{p \to 0} \frac{d}{d\theta} g_{0,t}(\theta) = 0 \quad \text{and} \quad \lim_{p \to 0} \frac{d}{d\theta} g_{1,t}(\theta) = \frac{\gamma}{\theta^2} e^{\gamma \underline{v}} \neq 0. \tag{2.18}$$

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

2.3 Multifactor Linear Asset Pricing Models

A large and growing literature explores inference issues in linear macro asset pricing models in which the stochastic discount factor (SDF) takes the affine form with constant SDF coefficients:

$$\pi_t(\theta) \equiv 1 - \gamma_f' f_t - \gamma_g' g_t \text{ with } \theta \equiv (\gamma_f', \gamma_g')', \tag{2.19}$$

where f_t is a ℓ_f-dimensional vector of observed priced risk factors, g_t is a ℓ_g-dimensional vector of potential risk factors that are orthogonal to f_t, and the factor loadings γ_f and γ_g are unknown constant vectors. The factors have zero mean and an identity covariance matrix. The functional form of the SDF $\pi_t(\theta)$ is behind the conditional versions of the classical CAPM and its multifactor extensions (e.g., Fama and French, 1996; Jagannathan and Wang, 1996; Hodrick and Zhang, 2001). It can also arise from the linearized structural asset pricing models in which $\pi_t(\theta)$ is a representative agent's marginal rate of substitution (e.g., Cochrane, 1996; Campbell and Cochrane, 1999; Lettau and Ludvigson, 2001; Gomes, Kogan, and Zhang, 2003; Bansal and Yaron, 2004).

Suppose there is a set of baseline portfolios with excess returns $r_t^e \equiv (r_{1,t}^e, \cdots, r_{k_0,t}^e)'$. The factors f_t are risk factors priced by the k_0 baseline portfolios. For example, the baseline priced risk factors f_t can be the Fama-French three factors, and the baseline portfolios can be the 3-month Treasury bill and the 25 common stock portfolios sorted based on the book-to-market ratio and firm size (Fama and French, 1996). The additional potential risk factors g_t can be the term structure factor and the default-premium factor (Fama and French, 1993).[6,7]

The Euler equation of the baseline test assets implies that

$$\mathbb{E}\left[g_{0,t}(\theta)\right] = 0_{k_0 \times 1} \quad \text{with } g_{0,t}(\theta) \equiv \pi_t(\theta) r_t^e. \tag{2.20}$$

Importantly, as emphasized in the empirical asset pricing literature (e.g., Jagannathan and Wang, 1996; Kan and Zhang, 1999a,b; Gospodinov, Kan, and Robotti, 2014, 2017; Kleibergen and Zhan, 2015; Burnside, 2015; Bryzgalova, 2016), the risk premium parameters are not well identified and the conventional estimation and inference approaches are unreliable, when some factors only weakly correlate with assets' returns or do not correlate at all. The concern of spurious factors is prevalent for potential macro risk factors and is nested in our framework. Let $\beta_g \equiv \mathbb{E}[r_t^e g_t']$, a $k_0 \times \ell_g$ matrix. Our framework allows the singular value(s) of β_g to be zero or arbitrarily close to 0 such that γ_g is not identified or is weakly identified.

In addition to the baseline portfolios, we can find some additional portfolios with excess returns $\tilde{r}_t^e \equiv (\tilde{r}_{1,t}^e, \cdots, \tilde{r}_{k_1,t}^e)'$ that are strongly correlated with g_t. Formally, the singular values of $\tilde{\beta}_g \equiv \mathbb{E}[\tilde{r}_t^e g_t']$ are all bounded away from 0, unlike that of β_g. Such additional portfolios usually arise from three sources: First, these portfolios can be obtained based on certain fundamental characteristics

---

[6]The term structure factor is the difference between the yield on a 30-year bond and the yield on a 1-month Treasury bill. The default-premium factor is the difference between the yields on baa and aaa corporate bonds.

[7]As another example, the baseline priced risk factors f_t can be the market portfolio, and the baseline portfolios can be the 3-month Treasury bill and the 10 common stock portfolios sorted based on the market beta. The additional potential risk factors g_t can be labor income growth, reflecting a return to human capital, and the default risk, measured by the difference between the yields on baa and aaa corporate bonds (as in Jagannathan and Wang, 1996; Hodrick and Zhang, 2001).

of firms or industries implied by theories and thus strongly identify the market price of risk of the potential factors g_t (e.g., Gomes, Kogan, and Yogo, 2009; Papanikolaou, 2011; Dou, Ji, Reibstein, and Wu, 2019; Dou, Ji, and Wu, 2020); second, these portfolios can be sorted based on the post-formation betas of stock returns on g_t (e.g., Yogo, 2006; Dou, Kogan, and Wu, 2020), and thus, by construction, their returns are highly correlated with g_t; and third, the baseline portfolios with excess returns r_t^e and the factors f_t can be international ones, whereas the additional portfolios and factors are domestic (say, U.S.) ones (e.g., Griffin, 2015; Dumas, Gabuniya, and Marston, 2020).

However, we are concerned with the possibility that these additional portfolios are not priced by the baseline factors f_t. Namely, we are concerned with whether the following additional cross-equation restrictions are correctly specified:

$$\mathbb{E}\left[g_{1,t}(\theta)\right] = 0_{k_1 \times 1} \quad \text{with } g_{1,t}(\theta) \equiv \pi_t(\theta)\tilde{r}_t^e, \tag{2.21}$$

given that we believe the baseline asset pricing moment conditions (2.20) hold.

3 Conditional Inference for Specification Test

Our objective is to test an asset pricing theory through specification test of the implied asset pricing moments

$$H_0 : \mathbb{E}[g_{1,t}(\theta_0)] = 0_{k_1 \times 1} \quad \text{versus} \quad H_1 : \mathbb{E}[g_{1,t}(\theta_0)] \neq 0_{k_1 \times 1}, \tag{3.1}$$

where $g_{1,t}(\theta) \equiv g_1(Y_t, \theta) \in \mathbb{R}^{k_1}$ 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,

$$\mathbb{E}[g_{0,t}(\theta_0)] = 0_{k_0 \times 1}, \tag{3.2}$$

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

The motivating examples discussed in the previous section share two key features. First, the baseline moments $\mathbb{E}[g_{0,t}(\theta_0)] = 0_{k_0 \times 1}$ only weakly identify θ_0, i.e., $\mathbb{E}[g_{0,t}(\theta)]$ is (nearly) flat around θ_0 and the singular values of the Jacobian matrix $Q_0 \equiv \mathbb{E}[\partial g_{0,t}(\theta_0)/\partial\theta']$ are (nearly) zero. Second, once the asset pricing theory is imposed, the full moments $\mathbb{E}[g_t(\theta_0)] = 0_{k \times 1}$ strongly identify θ_0,[8]

---

[8]If one is concerned that some subvector of θ, denoted by θ_c, is weakly identified by even the most informative asset pricing moments, we can apply the proposed test to the joint hypothesis $H_0 : \mathbb{E}[g_{1,t}(\theta_0)] = 0_{k_1 \times 1}$ and $\theta_c = \theta_{c,0}$ and reject the original null hypothesis if the joint hypothesis is rejected for all null values $\theta_{c,0}$ in its parameter space. This is a projection-based subvector inference method with weakly identified nuisance parameters. One can also apply the more efficient method in Andrews and Mikusheva (2016b) in this context.

i.e., the singular values of the associated Jacobian matrix $Q \equiv \mathbb{E}[\partial g_t(\theta_0)/\partial \theta']$ are bounded away from zero. For moments created with IVs, $\mathbb{E}[g_{0,t}(\theta_0)] = 0_{k_0 \times 1}$ and $\mathbb{E}[g_{1,t}(\theta_0)] = 0_{k_1 \times 1}$ are created with weak instruments and strong instruments, respectively.[9]

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

$$\mathcal{T} \equiv J - J_0, \quad \text{where } J \equiv \min_{\theta \in \Theta} g(\theta)'(\hat{\Omega}(\theta))^{-1} g(\theta) \text{ and } J_0 \equiv \min_{\theta \in \Theta} g_0(\theta)'(\hat{\Omega}_0(\theta))^{-1} g_0(\theta), \qquad (3.3)$$

with $g_0(\theta) \equiv n^{-1/2} \sum_{t=1}^n g_{0,t}(\theta) \in \mathbb{R}^{k_0}$, $g(\theta) \equiv n^{-1/2} \sum_{t=1}^n g_t(\theta) \in \mathbb{R}^k$, $\hat{\Omega}(\theta) \equiv \hat{\Omega}(\theta, \theta)$, where $\hat{\Omega}(\theta, \tilde{\theta})$ is an estimator of $\Omega(\theta, \tilde{\theta}) \equiv \lim_{n \to \infty} \text{Cov}(g(\theta), g(\tilde{\theta}))$ for any $\theta, \tilde{\theta} \in \Theta$, and $\hat{\Omega}_0(\theta)$ is the leading $k_0 \times k_0$ submatrix of $\hat{\Omega}(\theta)$. 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 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 to the finite-sample distribution of $\mathcal{T}$. We need an alternative critical value.[10]

Next, we provide the algorithm to compute the critical value based on the conditional inference approach. Following Andrews and Mikusheva (2016a), we view the baseline moment function $\mathbb{E}[g_{0,t}(\theta)]$ indexed by θ as a functional nuisance parameter and obtain a simulation-based critical value by conditioning on a sufficient statistic for $\mathbb{E}[g_{0,t}(\theta)]$. For the ease of presentation, we assume that the Jacobian matrix Q, the covariance function $\Omega(\theta, \tilde{\theta})$, and $\Omega \equiv \Omega(\theta_0, \theta_0)$ are all known for now. In practice, they can be replaced by consistent estimators, which are easily obtainable because the continuously updated estimator (CUE)

$$\hat{\theta} \equiv \arg\min_{\theta \in \Theta} g(\theta)'(\hat{\Omega}(\theta))^{-1} g(\theta) \qquad (3.4)$$

is consistent under the null, following standard arguments (e.g., Newey and Smith, 2004).

The sufficient statistic for the baseline moments $\mathbb{E}[g_{0,t}(\theta)]$ is constructed as follows. Under the null, $g(\hat{\theta})$, the full moment function evaluated at $\hat{\theta}$, is approximately a linear function of $g(\theta_0)$:

$$g(\hat{\theta}) = \Omega^{1/2} M \upsilon + \varepsilon_n, \quad \text{where } \upsilon \equiv \Omega^{-1/2} g(\theta_0) \to_d N(0, I_k), \qquad (3.5)$$

---

[9] In the context of linear IV model, Hahn, Ham, and Moon (2011) provide a generalized Hausman test robust to weak instruments. When applied to the same context, its power is lower than that of the conditional test here.

[10] When the baseline moments provide only weak identification, the J test still has correct size. However, it suffers from low power for two reasons. First, it ignores the information that the baseline moments are valid. Second, it effectively over-counts the degree of overidentification in the presence of a severe imbalance of the information content among moments. Chen, Dou, and Kogan (2020) demonstrate the consequence of such an information imbalance when estimation consistency is maintained and show that the J statistic is driven mainly by the asset pricing moments.

$M \equiv I_k - \Omega^{-1/2} Q (Q' \Omega^{-1} Q)^{-1} Q' \Omega^{-1/2\prime}$ 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 in large sample if $g_t(\theta)$ is nonlinear. Consider the following decomposition

$$g_0(\theta) = m(\theta) + V(\theta) g(\hat{\theta}), \quad \text{where } m(\theta) \equiv g_0(\theta) - V(\theta) g(\hat{\theta}) \tag{3.6}$$

is 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 $S_0 \equiv [I_{k_0}, 0_{k_0 \times k_1}]$. In large sample, the random component in $m(\theta)$, i.e., $m(\theta) - \mathbb{E}[g_{0,t}(\theta)]$, is a Gaussian process indexed by θ, and thus is independent of $g(\hat{\theta})$ by orthogonality. Conditioning on $m(\theta)$, the distribution of $g_0(\theta)$ and the test statistic do not depend on the functional nuisance parameter $\mathbb{E}[g_{0,t}(\theta)]$, which means that $m(\theta)$ is a sufficient statistic for $\mathbb{E}[g_{0,t}(\theta)]$. The sufficient statistic $m(\theta)$ contains identification information in the baseline moments and is independent of randomness in the full moments in large sample.

Conditioning on the sufficient statistic $m(\theta)$, we approximate the conditional distribution of the test statistic $\mathcal{T}$ by replacing $g(\hat{\theta})$ with $\Omega^{1/2} M \upsilon$ as in (3.5) and generate independent realizations of υ from the standard multivariate normal distribution. Specifically, we define

$$L(\upsilon; d_0) \equiv \upsilon' M \upsilon - \min_{\theta \in \Theta} \left[m(\theta) + V(\theta) \Omega^{1/2} M \upsilon \right]' (\Omega_0(\theta))^{-1} \left[m(\theta) + V(\theta) \Omega^{1/2} M \upsilon \right], \tag{3.7}$$

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 $\Omega(\theta) \equiv \Omega(\theta, \theta)$. This is the analog of the test statistic $\mathcal{T}$ with $g_0(\cdot)$ replaced by the decomposition in (3.6), $g(\hat{\theta})$ replaced by its linear approximation in (3.5), $\hat{\Omega}(\hat{\theta})$ replaced by Ω, and $\hat{\Omega}_0(\cdot)$ replaced by $\Omega_0(\cdot)$. To conduct the conditional specification test, we execute the following algorithm in practice.

Algorithm (Conditional Specification Test):

1. Approximate the unknown quantities $\Omega(\theta, \tilde{\theta})$, Q, $V(\theta)$ and M with their consistent estimators. Specifically let $\hat{\Omega}(\theta, \tilde{\theta})$ be the consistent estimator of $\Omega(\theta, \tilde{\theta})$ for any $\theta, \tilde{\theta} \in \Theta$, i.e., a sample analog for i.i.d. data and heteroskedasticity and autocorrelation consistent (HAC) estimator for time series data. Calculate the CUE $\hat{\theta}$ in (3.4) with $\hat{\Omega}(\theta) \equiv \hat{\Omega}(\theta, \theta)$. Then $\hat{Q} \equiv n^{-1/2} \partial g(\hat{\theta}) / \partial \theta'$, $\hat{\Omega} \equiv \hat{\Omega}(\hat{\theta}, \hat{\theta})$, $\hat{M} \equiv I_k - \hat{\Omega}^{-1/2} \hat{Q} (\hat{Q}' \hat{\Omega}^{-1} \hat{Q})^{-1} \hat{Q}' \hat{\Omega}^{-1/2\prime}$, and $\hat{V}(\theta) \equiv S_0 \hat{\Omega}(\theta, \hat{\theta}) \hat{\Omega}^{-1}$.

2. Compute the test statistic $\mathcal{T}$ following (3.3).

3. Compute the sufficient statistic $m(\cdot)$ following (3.6).

4. For $b = 1, \ldots, B$, take independent draws $\upsilon_b^* \sim N(0, I_k)$ and produce simulated process $\mathcal{T}_b^* = L(\upsilon_b^*; \hat{d})$ using the process $L(\upsilon; d_0)$ defined in (3.7), where d_0 being replaced by $\hat{d}$ means that $m(\cdot)$ in d_0 is estimated in step 3 and other components are estimated in step 1.

5. Let $b_0 = \lceil(1-\alpha)B\rceil$ be the smallest integer lager than or equal to $(1-\alpha)B$ and α be the nominal size of the test. The critical value $c_{B,\alpha}(\hat{d})$ is the b_0^{th} smallest value among $\{\mathcal{T}_b^*, b = 1, \ldots, B\}$.

6. Reject the null hypothesis in (3.1) if the statistic $\mathcal{T}$ in step 2 is larger than the critical value $c_{B,\alpha}(\hat{d})$ in step 5. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

For the test statistic and the critical value in the algorithm, there are three types of approximation errors between the finite-sample distribution and the large sample distribution: (i) the linear approximation error ε_n in (3.5); (ii) the Gaussian approximation error for the distribution of $m(\cdot)$ and v; (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 disappear in large sample, below we first consider a linear Gaussian statistical experiment where all types of errors are exactly zero. Let $v^* \equiv \Omega^{-1/2}\psi(\theta_0)$ 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 v and $m(\cdot)$, respectively, where $\psi(\cdot)$ is a Gaussian process defined in Assumption 1 below. In the Gaussian experiment, the test statistic $\mathcal{T}$ is exactly $L(v^*; d^*)$, where d^* is defined similar to d_0 with $m(\cdot)$ in d_0 replaced by $m^*(\cdot)$. Define its conditional $1-\alpha$ quantile as

$$c_\alpha^*(d^*) \equiv \inf\{c \in \mathbb{R} : P(L(v^*; d^*) > c \,|\, d^*) \leq \alpha\}, \tag{3.8}$$

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

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

The critical value $c_\alpha^*(d^*)$ can be simulated using the marginal distribution of v^* because (1) v^* enters $L(v^*; d^*)$ through Mv^*, (2) Mv^* and $m^*(\cdot)$ are independent, (3) $m^*(\cdot)$ is the only random component in d^*. In large sample, the simulated critical value $c_{B,\alpha}(\hat{d})$ approximates $c_\alpha^*(d^*)$ with high accuracy when B is a large number.

4 Asymptotic Uniform Validity for Nonlinear Models

In this section, we bring back many finite-sample features that are important in empirical work, including: (i) the moment function $\mathbb{E}[g_t(\theta)]$ is nonlinear in θ; (ii) $g(\theta) - \mathbb{E}[g_t(\theta)]$ is non-Gaussian, but weakly converges to a Gaussian process $\psi(\theta)$ in large sample; (iii) θ_0, Q, $\Omega(\cdot, \cdot)$ and Ω are

all unknown but can be consistently estimated under the null hypothesis. Under some regularity conditions, we show the proposed test has correct asymptotic size.

We first state some 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 moment conditions (3.2) holds. 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 $q(\theta) \equiv n^{-1} \sum_{t=1}^n \partial g_t(\theta)/\partial \theta'$ and $Q(\theta) \equiv \mathbb{E}[q(\theta)]$. 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 $\| \cdot \|$ 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) $\sup_{\theta \in \Theta} [\|\mathbb{E}[g_t(\theta)]\| + \|Q(\theta)\| + \sum_{j=1}^k \|\partial Q_j(\theta)/\partial \theta'\|] \leq C_m$ *for some finite constant C_m.*

Assumption 1(i) requires that the moment condition is well approximated by a Gaussian limit. In our long-run risk example, 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 our example, the loading on this latent process is smaller as the root gets closer to unity, and thus the non-stationary component is dominated.[11] Assumption 1(ii) follows from the uniform law of large numbers. Assumption 1(iii) includes standard regularity conditions on uniformly bounded moment functions and their derivatives.

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) $\sup_{\theta, \tilde{\theta} \in \Theta} [\|\Omega(\theta, \tilde{\theta})\| + \sum_{j=1}^k \|\partial \Omega(\theta, \tilde{\theta})/\partial \theta_j\|] \leq C_\Omega$ *for some finite constant C_Ω.*

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 smoothness conditions. Assumptions

---

[11] See Section SE in Cheng, Dou, and Liao (2020) for the details. In a similar spirit, the assumption holds for the disaster risk model. The baseline moments there have a non-singular covariance matrix even if the disaster occurs with a small probability because the normally distributed regular shock dominates the disaster shock. For the disaster risk example, an alternative approximation that combines the extreme value theory and the central limit theorem could work better for extremely rare and large disasters, see Müller (2019, 2020).

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 analysis of the proposed conditional test. Let $\lambda_{min}(A)$ denote the minimal eigenvalue of a square matrix A.

Assumption 3. *The following conditions hold uniformly over $\mathbb{P} \in \mathcal{P}_0$:*
(i) *There exists $\theta_0 \in \Theta$ such that $\mathbb{E}[g_t(\theta_0)] = 0_{k \times 1}$;*
(ii) *for any $\varepsilon > 0$ and any n, there exists a fixed constant $\delta_\varepsilon > 0$ such that $\inf_{\theta \in B_\varepsilon^c(\theta_0)} ||\mathbb{E}[g_t(\theta)]|| > \delta_\varepsilon$, where $B_\varepsilon^c(\theta) \equiv \{\tilde{\theta} \in \Theta : ||\tilde{\theta} - \theta|| \geq \varepsilon\}$;*
(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 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 when all moments are used.[12]

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 $v^* \sim N(0, I_k)$ and obtain the critical value

$$c_\alpha(\hat{d}) \equiv \inf\left\{ c \in \mathbb{R} : P^*(v^* : L(v^*; \hat{d}) > c) \leq \alpha \right\}, \tag{4.9}$$

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

Theorem 1. *Suppose Assumptions 1, 2, 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 test controls the asymptotic size no matter the unknown parameter θ_0 (or its subvector) is strongly identified, weakly identified, or not identified by the baseline moments $\mathbb{E}[g_{0,t}(\theta_0)] = 0_{k_0 \times 1}$.[13]

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

---

[12] Andrews and Guggenberger (2019) point out that near singular covariance matrix may show up together with weak identification. If this is the case with our baseline moments, the methods there can be used to adjust our test.

[13] 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 Andrews, Cheng, and Guggenberger (2020) for discussions on asymptotic similarity.

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* $\inf_{\theta \in B^c_\varepsilon(\theta_0)} \|\mathbb{E}[g_{0,t}(\theta)]\| > \delta_\varepsilon$*; (ii)* $\lambda_{\min}(Q'_0 Q_0) \geq c_\lambda$*.*

Assumption 4 is similar to Assumption 3 which 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, 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 when the baseline moments $\mathbb{E}[g_{0,t}(\theta_0)] = 0_{k_0 \times 1}$ provide strong identification of θ_0, the conditional test is equivalent to the C test under the null.

If the baseline moments $\mathbb{E}[g_{0,t}(\theta_0)] = 0_{k_0 \times 1}$ 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

$$\mathcal{T} \to_d \chi^2_{k_1 + d_c - d_\theta} \quad \text{and} \quad c_\alpha(\hat{d}) \to_p q_{1-\alpha}(\chi^2_{k_1 + d_c - d_\theta}) \tag{4.10}$$

uniformly over $\mathcal{P}_0 \cap \mathcal{P}_{00}$. In this case, $k_1 \geq d_\theta - d_c$ in (4.10) because the asset pricing moments of dimension k_1 must strongly identify all the parameters not in the baseline moments.

5 Power of the Conditional Specification Test

In this section, we investigate the power properties of the conditional test in two cases. First, when the asset pricing moments are globally misspecified, we show that the conditional test rejects these moments with probability approaching 1 (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 5. *The following conditions hold for any* $\mathbb{P} \in \mathcal{P}_{1,\infty} \subset \mathcal{P}$*: (i)* $\inf_{\theta \in \Theta} \|\mathbb{E}[g_{1,t}(\theta)]\| > c_g$ *for some* $c_g > 0$*; (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 5(i) implies that there are globally misspecified moments in $\mathbb{E}[g_{1,t}(\theta_0)] = 0_{k_1 \times 1}$. Assumption 5(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 5(ii) is the counterpart of Assumption 3(iii) under the alternative.

Theorem 3. *Suppose Assumptions 1, 2, 5 hold. For any $\mathbb{P} \in \mathcal{P}_{1,\infty}$, $\mathbb{P}(\mathcal{T} > c_\alpha(\hat{d})) \to 1$ as $n \to \infty$.*

The consistency of the conditional 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 test when the baseline moments provide strong identification.

Assumption 6. *The following conditions hold for any $\mathbb{P} \in \mathcal{P}_{1,A} \subset \mathcal{P}$:*
(i) $\mathbb{E}[g_{1,t}(\theta_0)] = an^{-1/2}$ *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 4. *Suppose Assumptions 1, 2, 4, 6 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 $a'_\Omega M a_\Omega$.

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 $\mathbb{E}[g_{0,t}(\theta_0)] = 0_{k_0 \times 1}$ 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 show Theorem 4 give

$$\mathbb{P}\left(\mathcal{T} > c_\alpha(\hat{d})\right) \to P\left(\chi^2_{k_1+d_c-d_\theta}(a'_\Omega M a_\Omega) > q_{1-\alpha}(\chi^2_{k_1+d_c-d_\theta})\right) \text{ as } n \to \infty. \tag{5.11}$$

When the baseline moments provide strong identification, the conditional test is asymptotically equivalent to the C test following Theorems 2, 3, 4.[14] 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 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.

Optimal test under weakly identified baseline moments is beyond the scope of the paper. However, the literature has provided several encouraging power results[15] for various conditional tests for $H_0 : \theta = \theta_0$ (e.g., Andrews, Moreira, and Stock, 2006; Andrews and Mikusheva, 2016a) and we expect the conditional specification test to inherit this good property. Simulation results

---

[14]See e.g., Hall (2005) for detailed derivations for the C test and its equivalence.
[15]The conditional test is more powerful than alternative robust tests and is nearly optimal in some cases.

below indeed confirm that the proposed test is more powerful than the J test. Methods in Elliott, Müller, and Watson (2015) could help to evaluate how close the proposed test is to the power envelope. We leave this important question to future research.

6 Simulation Studies

6.1 Simulation Design for the Long-Run Risk Model

The model, moments, and parameters are as described in Section 2.1. We generate Δc_t following (2.1) and (2.2) and generate r_t^e as follows:

$$r_t^e = n^{-1/2}\eta + \frac{1}{2}\left(2\gamma - \psi^{-1} - 1\right)\left(1 - \psi^{-1}\right)\frac{\phi^2}{(1-\rho)^2} + \varepsilon_{r,t}^e, \tag{6.1}$$

where the error term $\varepsilon_{r,t}^e \equiv \sigma_c\varepsilon_{c,t} + (1 - \psi^{-1})(1 - \rho)^{-1}\phi\varepsilon_{x,t} + \sigma_r\varepsilon_{r,t}$ in the asset pricing moment follows the derivation of the solution of the long-run risk model.[16] Here, $\varepsilon_{r,t}$ is an i.i.d. standard normal variable capturing the measurement error, and it is independent of the other shocks. Under the null of the correct specification of the asset pricing moment (2.5), $\eta = 0$. Under the alternative, we consider numerous values of η to compare the local power of different specification tests.

The parameters for the simulation design are chosen, close to those used by Bansal, Kiku, and Yaron (2012), as follows: (i) we simulate quarterly time series with sample size $n = 500$, which is in ballpark of the U.S. sample for the past century used by Constantinides and Ghosh (2011); (ii) the volatility is set to $\sigma_c = 0.0072 \times \sqrt{3}$; (iii) the preference parameters are set to $\delta = 0.9989^3$, $\gamma = 10$, and $\psi = 1.5$; (iv) the coefficient in the latent autoregressive process is $\rho = 0.975^3$; (v) in accordance with the data, we set the model-implied equity premium to $2^{-1}\left(2\gamma - \psi^{-1} - 1\right)\left(1 - \psi^{-1}\right)\theta^{-2} = 2\%$ (i.e., to match the quarterly equity premium), which leads to $\theta = (1 - \rho)/\phi = 12.36$; (vi) we set $\sigma_r = 7.5\%$ to match the volatility of the quarterly stock return in the data; and (vii) in accordance with the data, we set $1.5\% \leq 2^{-1}\left(2\gamma - \psi^{-1} - 1\right)\left(1 - \psi^{-1}\right)\theta^{-2} \leq 2.5\%$, which leads to $\underline{c} = 11.06$ and $\bar{c} = 14.27$ for the parameter space of θ.

6.2 Simulation Design for the Disaster Risk Model

The model, moments, and parameters are as described in Section 2.2. We generate Δc_t following (2.10) and (2.11), and generate r_t^e as follows:

$$r_t^e = n^{-1/2}\eta - p_n(\underline{v} + 1/\alpha) + \frac{p}{\alpha - \gamma}h(\alpha) + \varepsilon_{r,t}^e, \tag{6.2}$$

---

[16]See Section SD in Cheng, Dou, and Liao (2020) for the derivation.

where the error term $\varepsilon_{r,t}^e \equiv \sigma\varepsilon_t - [x_t(\underline{v} + J_t) - p_n(\underline{v} + 1/\alpha_n)] + \sigma_r\varepsilon_{r,t}$ in the asset pricing moment follows the derivation of the solution of the disaster risk model.[17] Here, $\varepsilon_{r,t}$ is an i.i.d. standard normal variable capturing the measurement error, and it is independent of the other shocks. Under the null of the correct specification of (2.14), $\eta = 0$. Under the alternative, we consider numerous values of η to compare the local power of different specification tests.

The parameters for the simulation design are chosen as follows: (i) we simulate yearly time series with sample size $n = 500$. In the literature (e.g., Barro, 2006; Barro and Ursúa, 2008; Barro and Jin, 2011; Tsai and Wachter, 2015), international consumption data over 35 countries for the past 100 years are used to calibrate the disaster risk models;[18] (ii) the volatility of contemporaneous annual consumption growth is set to $\sigma = 3.5\%$; (iii) the minimum level of disaster size is set to $\underline{v} = 7\%$; (iv) the risk aversion coefficient is set to $\gamma = 4$, similar to that of Barro (2006); (v) we set $p = 3.6\%$, the estimated disaster probability from the empirical frequency of entry into disaster states per year for consumption (Barro and Ursúa, 2008); (vi) we set $\sigma_r = 15\%$ to match the volatility of the stock return in the data; (vii) in accordance with the data, we set the model-implied equity premium to be $\theta^{-1}h(\theta p + \gamma) - p(\underline{v} + 1/(\theta p + \gamma)) = 8\%$ (i.e., to match the annual equity premium 8%), which leads to $\theta = 36.95$; and (viii) in accordance with the data, we set $6\% \leq \theta^{-1}h(\theta p + \gamma) - p(\underline{v} + 1/(\theta p + \gamma)) \leq 10\%$, which leads to the bounds $\underline{c} = 31.33$ and $\bar{c} = 43.00$.

6.3 Simulation Design for the Multifactor Linear Asset Pricing Models

The model, moments, and parameters are as described in Section 2.3 with one-dimensional factors f_t and g_t. We consider k_0 baseline portfolios with excess returns $r_t^e \in \mathbb{R}^{k_0}$ satisfying the following factor structure:

$$r_t^e = \mu + \beta_f f_t + \beta_g g_t + \sigma\varepsilon_t, \quad \text{with } \mu = \beta_f\gamma_f + \beta_g\gamma_g. \tag{6.3}$$

We also consider k_1 additional portfolios with excess returns $\tilde{r}_t^e \in \mathbb{R}^{k_1}$ satisfying the following factor structure:

$$\tilde{r}_t^e = \tilde{\mu} + \tilde{\beta}_f f_t + \tilde{\beta}_g g_t + \sigma\tilde{\varepsilon}_t, \quad \text{with } \tilde{\mu} = n^{-1/2}\eta + \tilde{\beta}_f\gamma_f + \tilde{\beta}_g\gamma_g. \tag{6.4}$$

Here the factors $f_t \in \mathbb{R}$ and $g_t \in \mathbb{R}$, and the errors $\varepsilon_t \in \mathbb{R}^{k_0}$ and $\tilde{\varepsilon}_t \in \mathbb{R}^{k_1}$ are i.i.d. standard multivariate normal variables, and they are mutually independent. The value of μ is set in (6.3) to ensure that the baseline moments (2.20) are correctly specified, and $\tilde{\mu}$ is set in (6.4) so that

---

[17]See Section SD in Cheng, Dou, and Liao (2020) for the derivation.

[18]It is assumed that all countries face the same disaster probability and distribution. Particularly, the U.S. experience is far from an anomaly relative to other countries, as Nakamura, Steinsson, Barro, and Ursúa (2013) discuss.

additional moments (2.21) are misspecified if and only if $\eta \neq 0$. The SDF is specified in (2.19) with unknown coefficients γ_f and γ_g.

The parameters for the simulation design are as follows: (i) we simulate monthly time series with sample size $n = 500$; (ii) we set $k_0 = 10$ and $k_1 = 2$; (iii) β_f is a k_0-dimensional vector whose i-th element is $0.01/ + (i-1) \times 0.02/(k_0 - 1)$; namely, the elements of β_f are equally spaced between 0.01 and 0.03; (iv) we set $\beta_g = n^{-1/2} \pi_g \times \mathbf{1}_{k_0}$, a k_0-dimensional constant vector, with $\pi_g = 0.03$ for a weak factor setup and all elements of β_g equal the maximum element in β_f for a strong factor setup; (v) the additional assets' betas are $\tilde{\beta}_f = 0.01 \times \mathbf{1}_{k_1}$, and $\tilde{\beta}_g$ is a k_1-dimensional vector whose i-th element is $0.01 + (i-1) \times 0.01/(k_1 - 1)$; namely, the elements of $\tilde{\beta}_g$ are equally spaced between 0.01 and 0.02; (vi) we set $\sigma = 1.5\%$; and (vii) the market price of risk is $\gamma_f = \gamma_g = 0.5$ and their parameter space is unbounded in the simulation.

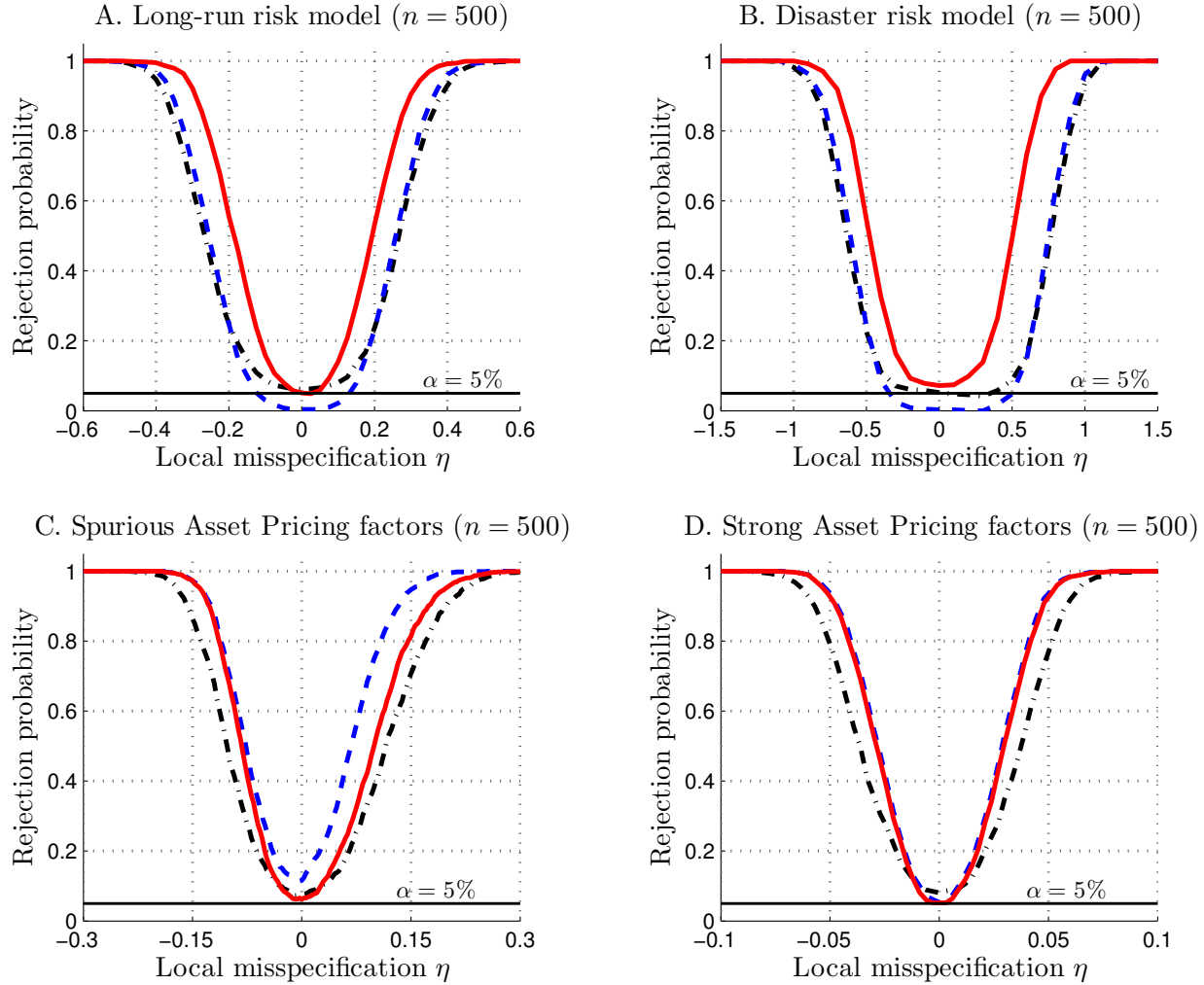## 6.4   Discussion of Simulation Results

Results for all three examples are based on 1000 simulation repetitions and $B = 1000$ random draws for the critical value of the conditional test in each repetition. Figure 1 reports the finite-sample rejection probabilities of the proposed conditional test, the J test, and the C test in the long-run risk model (panel A), the disaster risk model (panel B) and the linear asset pricing model (panels C and D).

In the long-run risk example and the disaster risk example, the simulation confirms that (i) the size of the conditional test is at the nominal level, (ii) the power of the conditional test is higher than that of the J test, (iii) the size of the C test deviates from the nominal level. In both cases, the C test severely under-rejects under the null. In both examples, the true parameters are chosen to follow those in the literature or calibrated to match the empirical counterparts. The patterns in panels A and B of Figure 1 confirm that weak identification is indeed prevalent in empirical studies with the long-run risk and disaster risk models. The proposed conditional test improves upon the existing standard test procedures substantially in terms of both size and power.

Panels C and D of Figure 1 provide results for the linear asset pricing model. Panel C displays the case with a spurious factor that is only weakly correlated with the portfolio returns in the baseline moments. The conditional test has correct size and higher power than the J test, as expected. In this case, the C test over-rejects under the null. Panel C together with panels A and B of Figure 1 shows that the size of the C test may largely deviate from the nominal level in both directions when weak identification occurs.[19] Panel D of Figure 1 is the standard scenario with

---

[19]Under weak identification in the baseline moments, rejection probability of the C test depends on the specific model and moments, parameter values, and the parameter space. In a homoskedastic linear IV model, Guggenberger,

Figure 1: A comparison of tests for the long-run risk, disaster risk, and linear factor models.

A. Long-run risk model ($n = 500$)

B. Disaster risk model ($n = 500$)

C. Spurious Asset Pricing factors ($n = 500$)

D. Strong Asset Pricing factors ($n = 500$)

Note: Panels A and B plot the rejection probabilities of three tests for the simulation study on the long-run risk and disaster risk model, respectively, and panels C and D plot those for the linear asset pricing model. In all four panels, the red solid curve represents the rejection probability of the conditional specification test, with the test statistic $\mathcal{T}$ in (3.3) and the conditional critical value in (4.9); the black dash-dotted curve represents the rejection probability of the J test (Hansen, 1982); the blue dashed curve represents the rejection probability of the C test (Eichenbaum, Hansen, and Singleton, 1988); and the black solid horizontal line represents the 5% nominal size for all three tests.

strong identification, in which case the C test is optimal. Here, the proposed conditional test is indistinguishable from the C test and achieves the same level of optimality.

Kleibergen, Mavroeidis, and Chen (2012) provide some related analytical results on a test that shares some features of the C test.

7 Conclusion

This paper provides a robust and powerful test to evaluate macro asset pricing models. Our test gains power by exploiting valid but noisy information in weakly identified baseline moments. To achieve robustness under weak identification, the conditional test decouples the useful "orthogonal" macroeconomic information embedded in the baseline moment restrictions from the additional asset pricing moment restrictions. Our novel approach is particularly useful when standard over-identification tests suffer from distorted size or poor power due to information imbalance or reduced rank. It will help researchers, practitioners, and monetary authorities to better understand the economic mechanisms behind macro asset pricing models.

<center>APPENDIX: PROOFS</center>

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, $\|x\|_A \equiv (x'A^{-1}x)^{1/2}$. Proof of all auxiliary Lemmas, i.e., Lemmas A2 – A8 below, are given in Cheng, Dou, and Liao (2020).

PROOF OF LEMMA 1. Since Mv^* and $m^*(\cdot)$ are mean-zero Gaussian, part (i) follows from $E[m^*(\cdot)(Mv^*)'] = 0$. For part (ii), by the law of iterated expectation and the definition of $c_\alpha^*(d^*)$,

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

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

The following results hold for the CUE $\hat{\theta}$ in (3.4) that uses the full moments and some estimators based on $\hat{\theta}$, regardless of the identification strength in the baseline moments.

Lemma A2. *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* $\sup_{\theta \in \Theta} \|V(\theta)\| \leq c_\lambda^{-1}C_\Omega$.

<center>21</center>

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

$$\xi \equiv (x', d')', \text{ where } d \equiv (m_d(\cdot)', \text{vec}(V_d(\cdot))', \text{vech}(\Omega_d)', \text{vech}(\Omega_{0,d}(\cdot))', \text{vech}(M_d)')'. \tag{A.2}$$

Define

$$R(\xi) \equiv \|x\|_{\Omega_d}^2 - \min_{\theta \in \Theta} \|m_d(\theta) + V_d(\theta)x\|_{\Omega_{0,d}(\theta)}^2 , \tag{A.3}$$

and

$$L(v; d) \equiv v'M_d v - \min_{\theta \in \Theta} \left\| m_d(\theta) + V_d(\theta)\Omega_d^{1/2}M_d v \right\|_{\Omega_{0,d}(\theta)}^2 . \tag{A.4}$$

The test statistic $\mathcal{T}$ in (3.3) can be written as

$$\mathcal{T} = R(\hat{\xi}), \text{ where } \hat{\xi} \equiv (g(\hat{\theta})', \hat{d}')' \text{ and}$$
$$\hat{d} \equiv (\hat{m}(\cdot)', \text{vec}(\hat{V}(\cdot))', \text{vech}(\hat{\Omega})', \text{vech}(\hat{\Omega}_0(\cdot))', \text{vech}(\hat{M})')'. \tag{A.5}$$

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

$$\|\xi\|_s = \|x\| + \sup_{\theta \in \Theta} \|m_d(\theta)\| + \sup_{\theta \in \Theta} \|V_d(\theta)\| + \|\Omega_d\| + \sup_{\theta \in \Theta} \|\Omega_{0,d}(\theta)\| + \|M_d\|. \tag{A.6}$$

Lemma A3. *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 $\xi^* \equiv ((\Omega^{1/2} M v^*)', d^{*\prime})'$ 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

$$R_C(\xi) \equiv R(\xi)t_C(x'\Omega^{-1}x) \tag{A.7}$$

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 $L(v; d)$ as

$$\overline{L}_C(v; d) \equiv L_C(v; d)I\{\|v\|^2 \le C\}, \text{ where } L_C(v; d) \equiv L(v; d)t_C(v'M_d v). \tag{A.8}$$

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

22

Lemma A4. *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 $\mathcal{T} \geq 0$. Moreover, if $\hat{Q}'\hat{Q}$ is nonsingular, $L(v; \hat{d}) \geq 0$ for any $x \in \mathbb{R}^k$.*

Lemma A5. *Given $R(\xi) \geq 0$, the functional $R_C(\xi)$ is bounded and Lipschitz in ξ.*

Lemma A6. *Let $c_{\alpha,C}(d) \equiv \inf \left\{ c : P^* \left(v^* : \overline{L}_C(v^*; d) > c \right) \leq \alpha \right\}$. Given $L(v; d) \geq 0$, $c_{\alpha,C}(d)$ is bounded and Lipschitz in d.*

The extra truncation $\|v\|^2 \leq C$ in $\overline{L}_C(v; d)$ causes a discrepancy between $c_{\alpha,C}(d^*)$ and the conditional $1 - \alpha$ quantile of $R_C(\xi^*)$ given d^*. Lemma A7 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 test.

Lemma A7. *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 $L(v^*; \hat{d})$, respectively. These two functions have to be truncated differently too, as in (A.7) and (A.8), respectively, to yield the bounded Lipschitz property. (ii) The additional truncation to $L(v^*; \hat{d})$ causes a discrepancy between $c_{\alpha,C}(d^*)$ and the conditional $1 - \alpha$ quantile of $R_C(\xi^*)$ given d^*. Lemma A7 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 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

$$\lim_{n \to \infty} \sup_{\mathbb{P} \in \mathcal{P}_0} \mathbb{P}\left(R(\hat{\xi}) > c_\alpha(\hat{d}) + \varepsilon \right) > \alpha, \tag{A.9}$$

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

$$\mathbb{P}_{n_i}\left(R(\hat{\xi}) > c_\alpha(\hat{d}) + \varepsilon \right) > \alpha + \delta \text{ for all } i. \tag{A.10}$$

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

$$\mathbb{P}_{n_i}\left(R(\hat{\xi}) > u \right) \leq \mathbb{P}_{n_i}\left(R(\hat{\xi}) > u, g(\hat{\theta})'(\hat{\Omega}^{-1})g(\hat{\theta}) \leq C \right) + \mathbb{P}_{n_i}\left(g(\hat{\theta})'(\hat{\Omega}^{-1})g(\hat{\theta}) > C \right). \tag{A.11}$$

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,

$$\mathbb{P}_{n_i}\left(g(\hat{\theta})'(\hat{\Omega}^{-1})g(\hat{\theta}) > C_{1,\delta} \right) \leq \delta/4, \tag{A.12}$$

23

which together with (A.10) and (A.11) implies that

$$\mathbb{P}_{n_i}\left(R(\hat{\xi}) > c_\alpha(\hat{d}) + \varepsilon,\ g(\hat{\theta})'(\hat{\Omega}^{-1})g(\hat{\theta}) \leq C\right) > \alpha + 3\delta/4, \tag{A.13}$$

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

$$I\left\{R_C(\hat{\xi}) > u\right\} \geq I\left\{R(\hat{\xi}) > u\right\} I\left\{g(\hat{\theta})'(\hat{\Omega}^{-1})g(\hat{\theta}) \leq C\right\} \text{ for any } u \in \mathbb{R}, \tag{A.14}$$

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 $u \leq C$ following its definition. By (A.13) and (A.14), we have for any $C \geq C_{1,\delta}$,

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

Since $L(v, \hat{d}) \geq 0$ for any $v \in \mathbb{R}^k$ by Lemma A4 and $t_C(u) \leq 1$ for any $u \in \mathbb{R}$, we have $L_C(v, \hat{d}) \leq L(v, \hat{d})$ for any $v \in \mathbb{R}^k$, which further implies that $c_{\alpha,C}(\hat{d}) \leq c_\alpha(\hat{d})$. Therefore, by (A.15) we deduce that for any $C \geq C_{1,\delta}$,

$$\mathbb{P}_{n_i}\left(R_C(\hat{\xi}) - c_{\alpha,C}(\hat{d}) \geq \varepsilon\right) > \alpha + 3\delta/4. \tag{A.16}$$

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 A5 and Lemma A6, $R_C(\xi) - c_{\alpha,C}(d)$ is bounded and Lipschitz in ξ. Therefore, by Lemma A3,

$$\lim_{n\to\infty} \sup_{f\in BL_1} \|\mathbb{E}\left[f(U_{C,n})\right] - E\left[f(U_{\infty,C,n})\right]\| = 0. \tag{A.17}$$

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.17) implies that $U_{C,n_j} \to_d U_C$. Since (A.16) can be written as $\mathbb{P}_{n_i}\left(U_{C,n_i} \geq \varepsilon\right) > \alpha + 3\delta/4$, by Portmanteau theorem,

$$\liminf_{n_j\to\infty} P\left(U_{\infty,C,n_j} > \varepsilon/2\right) \geq P\left(U_C > \varepsilon/2\right) \geq P\left(U_C \geq \varepsilon\right) \tag{A.18}$$

$$\geq \limsup_{n_j\to\infty} \mathbb{P}_{n_j}\left(U_{C,n_j} \geq \varepsilon\right) \geq \alpha + 3\delta/4, \quad \text{for any } C \geq C_{1,\delta}.$$

We next show that for all large C, $P\left(U_{\infty,C,n_j} > \varepsilon/2\right) \leq \alpha + \delta/4$ for any n_j, which contradicts (A.18), and hence the claim of the theorem holds. To this end, for $C \geq C_{2,\delta}$ in Lemma A7,

$$P\left(U_{\infty,C,n_j} > \varepsilon/2\right) = P\left(R_C(\xi^*) > c_{\alpha,C}(d^*) + \varepsilon/2\right) \leq \alpha + \delta/4, \tag{A.19}$$

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 A7. $\hspace{2cm}$ Q.E.D.

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

24

Lemma A8. *Under Assumptions 1, 2, 3 and 4, we have uniformly over $\mathbb{P} \in \mathcal{P}_0 \cap \mathcal{P}_{00}$:*

(a) $n^{1/2}(\hat{\theta}^* - \theta_0) = -(Q'\Omega^{-1}Q)^{-1}Q'\Omega^{-1}g(\theta_0) - \left(Q_0'\Omega_0^{-1}Q_0\right)^{-1}Q_0'\Omega_0^{-1}S_0\Omega^{1/2}Mv^* + o_p(1);$

(b) $L(v^*, \hat{d}) = v^{*\prime}(M - \tilde{M}_0)v^* + o_p(1);$

(c) $v^{*\prime}(M - \tilde{M}_0)v^* \sim \chi^2_{k_1}.$

PROOF OF THEOREM 2. (i) Under Assumptions 1 – 3, Lemma A2 gives

$$g(\hat{\theta}) = \Omega^{1/2}M\Omega^{-1/2}g(\theta_0) + o_p(1) \quad \text{and} \quad \hat{\Omega} \equiv \hat{\Omega}(\hat{\theta}) = \Omega + o_p(1), \tag{A.20}$$

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

$$g_0(\hat{\theta}_0) = \Omega_0^{1/2}M_0\Omega_0^{-1/2}g_0(\theta_0) + o_p(1) \quad \text{and} \quad \hat{\Omega}_0 \equiv \hat{\Omega}_0(\hat{\theta}_0) = \Omega_0 + o_p(1) \tag{A.21}$$

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

$$L(v^*, \hat{d}) = v^{*\prime}\hat{M}v^* - \left\|\hat{m}(\hat{\theta}^*) + \hat{V}(\hat{\theta}^*)\hat{\Omega}^{1/2}\hat{M}v^*\right\|^2_{\hat{\Omega}_0(\hat{\theta}^*)}. \tag{A.22}$$

By Lemma A8(b, c), we have uniformly over $\mathbb{P} \in \mathcal{P}_0 \cap \mathcal{P}_{00}$,

$$L(v^*, \hat{d}) = L^* + o_p(1), \quad \text{where } L^* \equiv v^{*\prime}(M - \tilde{M}_0)v^* \sim \chi^2_{k_1}. \tag{A.23}$$

By (A.23), there exists a positive sequence $\delta_n = o(1)$ such that for any $\varepsilon > 0$,

$$\mathbb{P}^* \left(|L(v^*, \hat{d}) - L^*| \geq \varepsilon/2\right) = o(\delta_n), \text{ uniformly over } \mathbb{P} \in \mathcal{P}_0 \cap \mathcal{P}_{00}, \tag{A.24}$$

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

$$\mathbb{P}^* \left(|L(v^*, \hat{d}) - L^*| \geq \varepsilon/2 \,\Big|\, \hat{d}\right) \leq \delta_n \text{ wpa1}. \tag{A.25}$$

Note that $c_\alpha(\hat{d})$ is the $1 - \alpha$ conditional quantile of $L(v^*, \hat{d})$ given $\hat{d}$ and $L^* \sim \chi^2_{k_1}$ is independent of $\hat{d}$. Therefore, (A.25) implies

$$q_{1-\alpha-\delta_n}(\chi^2_{k_1}) - \varepsilon/2 \leq c_\alpha(\hat{d}) \leq q_{1-\alpha+\delta_n}(\chi^2_{k_1}) + \varepsilon/2 \text{ wpa1} \tag{A.26}$$

because by (A.25) and the union bound of (conditional) probability, we have

$$\mathbb{P}^* \left(L(v^*, \hat{d}) > q_{1-\alpha+\delta_n}(\chi^2_{k_1}) + \varepsilon/2 \,\Big|\, \hat{d}\right) \leq \mathbb{P}^* \left(L^* > q_{1-\alpha+\delta_n}(\chi^2_{k_1}) \,\Big|\, \hat{d}\right) + \delta_n = \alpha, \tag{A.27}$$

$$\mathbb{P}^* \left(L^* > c_\alpha(\hat{d}) + \varepsilon/2 \,\Big|\, \hat{d}\right) \leq \mathbb{P}^* \left(L(v^*, \hat{d}) > c_\alpha(\hat{d}) \,\Big|\, \hat{d}\right) + \delta_n \leq \alpha + \delta_n.$$

25

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

$$q_{1-\alpha-\delta_n}(\chi^2_{k_1}) - \varepsilon/2 \leq q_{1-\alpha}(\chi^2_{k_1}) \leq q_{1-\alpha+\delta_n}(\chi^2_{k_1}) + \varepsilon/2, \tag{A.28}$$

which together with (A.26) implies that, for any $\varepsilon > 0$, $|c_\alpha(\hat{d}) - q_{1-\alpha}(\chi^2_{k_1})| \leq \varepsilon$ wpa1. $\quad$ *Q.E.D.*

PROOF OF THEOREM 3. We first show that the test statistic, written as $R(\hat{\xi})$, diverges at rate n under global misspecification. By Assumptions 2(i, iv) and 5(ii),

$$R(\hat{\xi}) = \min_{\theta \in \Theta} g(\theta)'(\hat{\Omega}(\theta))^{-1}g(\theta) - \min_{\theta \in \Theta} g_0(\theta)'(\hat{\Omega}_0(\theta))^{-1}g_0(\theta)$$

$$\geq (C_\Omega + 1)^{-1} \min_{\theta \in \Theta} \|g(\theta)\|^2 - c_\lambda^{-1} \|g_0(\theta_0)\|^2 \quad \text{wpa1}, \tag{A.29}$$

where

$$\|g(\theta)\|^2 \geq \frac{1}{2} \|\mathbb{E}[g(\theta)]\|^2 - \|g(\theta) - \mathbb{E}[g(\theta)]\|^2. \tag{A.30}$$

By Assumption 5(i), there exists a constant $c_g > 0$ such that $\min_{\theta \in \Theta} \left\| n^{-1/2} \mathbb{E}[g(\theta)] \right\|^2 \geq c_g$, which combined with (A.29), (A.30), and Assumptions 1(i) and 5(ii) implies that

$$R(\hat{\xi}) \geq n \left(K^{-1} \min_{\theta \in \Theta} \left\| n^{-1/2}\mathbb{E}[g(\theta)] \right\|^2 - o_p(1) \right) \geq nc_g K^{-1} \quad \text{wpa1}. \tag{A.31}$$

The critical value satisfies $c_\alpha(\hat{d}) \leq q_{1-\alpha}(\chi^2_k)$ wpa1, because $L(v^*; \hat{d}) \leq v^{*\prime}\hat{M}v^* \leq \|v^*\|^2$ wpa1 given that $\hat{M}$ is an idempotent matrix wpa1 under Assumption 5(ii) and $q_{1-\alpha}(\chi^2_k)$ is the $1 - \alpha$ quantile of $\|v^*\|^2$. Therefore, by (A.31) and $c_\alpha(\hat{d}) \leq q_{1-\alpha}(\chi^2_k)$ wpa1, we have

$$\mathbb{P}\left(R(\hat{\xi}) > c_\alpha(\hat{d})\right) \geq \mathbb{P}\left(nc_g K^{-1} > q_{1-\alpha}(\chi^2_k)\right) - o(1) \to 1, \text{ as } n \to \infty. \tag{A.32}$$

$$\text{Q.E.D.}$$

PROOF OF THEOREM 4. Under Assumptions 1 and 2, the strong identification in baseline moments in Assumption 4, and the local misspecification in Assumption 6, $\hat{\theta}$ and $\hat{\theta}_0$ are consistent by the standard arguments and results in (A.20) and (A.21) remain valid. Therefore,

$$R(\hat{\xi}) \to_d (v^* + a_\Omega)'M(v^* + a_\Omega) - v_0^{*\prime}M_0 v_0^*, \tag{A.33}$$

where $a_\Omega \equiv \Omega^{-1/2}a$ and v_0^* denotes the leading k_0 subvector of v^*. By the standard arguments in the GMM literature (e.g., Hall, 2005, Section 5), we have

$$(v^* + a_\Omega)'M(v^* + a_\Omega) - v_0^{*\prime}M_0 v_0^* \sim \chi^2_{k_1}(a'_\Omega M a_\Omega). \tag{A.34}$$

We next study $c_\alpha(\hat{d})$ under the local misspecification. Since $\hat{\theta}$ is $n^{1/2}$ consistent under the local misspecification, Lemma A8 remains valid for any $\mathbb{P} \in \mathcal{P}_{00} \cap \mathcal{P}_{1,A}$. Therefore, for any $\mathbb{P} \in \mathcal{P}_{00} \cap \mathcal{P}_{1,A}$,

$$L(v^*, \hat{d}) = v^{*\prime}(M - \tilde{M}_0)v^* + o_p(1) \sim \chi^2_{k_1}. \tag{A.35}$$

By (A.35) 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 (A.33) proves the claim of the theorem. $\quad$ *Q.E.D.*

References

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., AND P. GUGGENBERGER (2019): "Identification- and Singularity-Robust Inference for Moment Condition Models," *Quantitative Economics*, 10(4), 1703–1746.

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, I. (2016): "Conditional Linear Combination Tests for Weakly Identified Models," *Econometrica*, 84(6), 2155–2182.

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.

ANDREWS, I., J. H. STOCK, AND L. SUN (2020): "Weak Instruments in IV Regression: Theory and Practice," *Annual Review of Economics*, Forthcoming.

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," *The Quarterly Journal of Economics*, 121(3), 823–866.

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

BARRO, R. J., AND T. JIN (2011): "On the Size Distribution of Macroeconomic Disasters," *Econometrica*, 79(5).

BARRO, R. J., AND J. F. URSÚA (2008): "Macroeconomic Crises since 1870," *Brookings Papers on Economic Activity*, 39(1 (Spring), 255–350.

BARRO, R. J., AND J. F. URSÚA (2012): "Rare Macroeconomic Disasters," *Annual Review of Economics*, 4(1), 83–109.

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.

BRYZGALOVA, S. (2016): "Spurious Factors in Linear Asset Pricing Models," Working paper, Stanford University.

BURNSIDE, C. (2015): "Identification and Inference in Linear Stochastic Discount Factor Models with Excess Returns," *Journal of Financial Econometrics*, 14(2), 295–330.

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 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," Working papers.

CHENG, X., W. W. DOU, AND Z. LIAO (2020): "Supplemental Appendix to 'Macro-Finance Decoupling: Robust Evaluations of Macro Asset Pricing Models'," *working paper, http://ssrn.com/abstract=3609598, also available on the authors' personal websites*.

COCHRANE, J. H. (1996): "A Cross-Sectional Test of an Investment-Based Asset Pricing Model," *Journal of Political Economy*, 104(3), 572–621.

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

CONSTANTINIDES, G. M., AND A. GHOSH (2011): "Asset Pricing Tests with Long-run Risks in Consumption Growth," *The Review of Asset Pricing Studies*, 1(1), 96–136.

DANIEL, K., AND S. TITMAN (2012): "Testing Factor-Model Explanations of Market Anomalies," *Critical Finance Review*, 1(1), 103–139.

DOU, W., Y. JI, D. REIBSTEIN, AND W. WU (2019): "Inalienable Customer Capital, Corporate Liquidity, and Stock Returns," *Journal of Finance*, Forthcoming.

DOU, W., Y. JI, AND W. WU (2020): "Competition, Profitability, and Discount Rates," *Journal of Financial Economics*, Forthcoming.

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.

DOU, W. W., L. KOGAN, AND W. WU (2020): "Common Fund Flows: Flow Hedging and Factor Pricing," Working papers, The Wharton School at University of Pennsylvania.

DUMAS, B., T. GABUNIYA, AND R. C. MARSTON (2020): "Geographic Stock Returns Indexes," Working papers, INSEAD.

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," *The 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.

EPSTEIN, L., AND S. ZIN (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption Growth and Asset Returns I: A Theoretical Framework," *Econometrica*, 57, 937–969.

FAMA, E. F., AND K. R. FRENCH (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33(1), 3 – 56.

——— (1996): "Multifactor Explanations of Asset Pricing Anomalies," *The Journal of Finance*, 51(1), 55–84.

FENG, G., S. GIGLIO, AND D. XIU (2020): "Taming the Factor Zoo: A Test of New Factors," *The Journal of Finance*, 75(3), 1327–1370.

GOMES, J., L. KOGAN, AND L. ZHANG (2003): "Equilibrium Cross Section of Returns," *Journal of Political Economy*, 111(4), 693–732.

GOMES, J. F., L. KOGAN, AND M. YOGO (2009): "Durability of Output and Expected Stock Returns," *Journal of Political Economy*, 117(5), 941–986.

GOSPODINOV, N., R. KAN, AND C. ROBOTTI (2014): "Misspecification-Robust Inference in Linear Asset-Pricing Models with Irrelevant Risk Factors," *The Review of Financial Studies*, 27(7), 2139–2170.

——— (2017): "Spurious Inference in Reduced-Rank Asset-Pricing Models," *Econometrica*, 85(5), 1613–1628.

GRIFFIN, J. M. (2015): "Are the Fama and French Factors Global or Country Specific?," *The Review of Financial Studies*, 15(3), 783–803.

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., AND K. J. SINGLETON (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50(5), 1269–86.

HODRICK, R. J., AND X. ZHANG (2001): "Evaluating the Specification Errors of Asset Pricing Models," *Journal of Financial Economics*, 62(2), 327 – 376.

JAGANNATHAN, R., AND Z. WANG (1996): "The Conditional CAPM and the Cross-Section of Expected Returns," *The Journal of Finance*, 51(1), 3–53.

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

——— (1999b): "Two-Pass Tests of Asset Pricing Models with Useless Factors," *The Journal of Finance*, 54(1), 203–235.

KELLY, B. T., S. PRUITT, AND Y. SU (2019): "Characteristics Are Covariances: A Unified Model of Risk

and Return," *Journal of Financial Economics*, 134(3), 501 – 524.

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

KLEIBERGEN, F., AND Z. ZHAN (2015): "Unexplained Factors and Their Effects on Second Pass R-squared's," *Journal of Econometrics*, 189(1), 101 – 116.

KLEIBERGEN, F., AND Z. ZHAN (2020): "Robust Inference for Consumption-Based Asset Pricing," *The Journal of Finance*, 75(1), 507–550.

LETTAU, M., AND S. LUDVIGSON (2001): "Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying," *Journal of Political Economy*, 109(6), 1238–1287.

LEWELLEN, J., S. NAGEL, AND J. SHANKEN (2010): "A Skeptical Appraisal of Asset Pricing Tests," *Journal of Financial Economics*, 96(2), 175 – 194.

MANKIW, G., AND M. D. SHAPIRO (1986): "Do We Reject Too Often? Small Sample Properties of Tests of Rational Expectations Models," *Economics Letters*, 20(2), 139 – 145.

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

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

MÜLLER, U. K. (2019): "Refining the Central Limit Theorem Approximation via Extreme Value Theory," *Statistics & Probability Letters*, 155, 108564.

——— (2020): "A More Robust t-Test," Working papers, Princeton University.

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.

NAGEL, S. (2013): "Empirical Cross-Sectional Asset Pricing," *Annual Review of Financial Economics*, 5(1), 167–199.

NAGEL, S., AND K. J. SINGLETON (2011): "Estimation and Evaluation of Conditional Asset Pricing Models," *The Journal of Finance*, 66(3), 873–909.

NAKAMURA, E., J. STEINSSON, R. BARRO, AND J. URSÚA (2013): "Crises and Recoveries in an Empirical Model of Consumption Disasters," *American Economic Journal: Macroeconomics*, 5(3), 35–74.

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.

PAPANIKOLAOU, D. (2011): "Investment Shocks and Asset Prices," *Journal of Political Economy*, 119(4), 639–685.

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

ROSS, S. A. (1976): "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13(3), 341–360.

SCHORFHEIDE, F., D. SONG, AND A. YARON (2018): "Identifying Long-Run Risks: A Bayesian Mixed-Frequency Approach," *Econometrica*, 86(2), 617–654.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557–586.

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.

TSAI, J., AND J. A. WACHTER (2015): "Disaster Risk and Its Implications for Asset Pricing," *Annual Review of Financial Economics*, 7(1), 219–252.

WEIL, P. (1989): "The Equity Premium Puzzle and the Risk-Free Rate Puzzle," *Journal of Monetary Economics*, 24, 401–421.

YOGO, M. (2006): "A Consumption-Based Explanation of Expected Stock Returns," *Journal of Finance*, 61(2), 539–580.