Online Appendix for
"Measuring the 'Dark Matter' in Asset Pricing Models"

Hui Chen Winston Wei Dou Leonid Kogan*

November 23, 2021

Contents

*Chen is at MIT (Sloan) and NBER. Dou is at University of Pennsylvania (Wharton). Kogan is at MIT (Sloan) and NBER. Emails: huichen@mit.edu, wdou@wharton.upenn.edu, lkogan@mit.edu.

1 Technical and Regularity Assumptions

We first discuss the relevant regularity conditions, including smoothness, rank, and identification. These assumptions are also stated in the appendix of Chen, Dou, and Kogan (2021).

Assumption 1 (GMM Regularity Conditions)**.** *We assume that the moment function $m(\cdot,\theta)$, defined on a compact set Θ, satisfies the following regularity conditions:*

(i) there exists $\theta_0 \in int(\Theta)$ such that $\mathcal{Q}(\theta_0)$ is non-empty;

(ii) the moment restrictions are over-identified: $d_\theta < d_m$;

(iii) $\mathrm{E}^{\mathrm{Q}_0}[m_t^{(1)}(\theta^{(1)})] = 0$ and $\mathrm{E}^{\mathrm{Q}_0}[m_t(\theta)] = 0$ only when $\theta^{(1)} = \theta_0^{(1)}$ and $\theta = \theta_0$;

(iv) $m_t(\theta)$ is continuously differentiable in θ, and D has full column rank.

The compactness of Θ and the assumption $\theta_0 \in int(\Theta)$ are the standard regularity conditions to ensure the uniform law of large numbers (ULLN) and the first-order-condition characterization of GMM estimators, respectively. Condition (i) means that the moment restrictions are satisfied under θ_0 and Q_0, though Q_0 may not be the true DGP. Condition (ii) is the standard over-identification condition in GMM (see Hansen, 1982). Condition (iii) is also a standard identification assumption to ensure that the sequence of GMM estimators has a unique limit (see Hansen, 1982). Condition (iv) is the rank condition for moment restrictions, and is the sufficient condition for local identification enabling us to consistently estimate θ_0.

Assumption 2 (Markov Processes)**.** *$\{\mathbf{y}_t : t = 0, 1, \cdots\}$ is a time-homogeneous Harris ergodic and stationary Markov process satisfying the Doeblin condition.*

A Markov process is Harris ergodic if it is aperiodic, irreducible, and positive Harris recurrent (e.g. Jones, 2004; Meyn and Tweedie, 2009). Harris ergodicity guarantees the existence of a unique invariant probability measure (e.g., Meyn and Tweedie, 2009). Given Harris ergodicity, stationarity only requires that the initial distribution of $\mathbf{y}_0$ is the unique invariant probability measure. The Doeblin condition implies that the ϕ-mixing coefficients $\phi(n)$ decay to zero exponentially fast (e.g. Bradley, 2005, Section 3.2 and Theorem 3.4), which is useful for establishing the uniform law of large numbers (ULLN) (White and Domowitz, 1984) and the central limit theorem (CLT) (e.g., Jones, 2004, Theorem 9).

In Assumption 3, we impose additional assumptions about the heteroskedasticity of the locally unstable DGP under consideration, thereby extending the statistical setting of Andrews (1993), Sowell (1996) and Li and Müller (2009) to the semiparametric setting.

Assumption 3 (Tail Properties of Local Instability). *As $n \to \infty$, it holds that under* Q_0

(i) $n^{-1} \max_{1 \leq t \leq n} |g(\mathbf{y}_{t-1}, \mathbf{y}_t)|^2 = o_p(1)$*;*

(ii) $\mathrm{E}^{\mathrm{Q}_0}\left[|g(\mathbf{y}_{t-1}, \mathbf{y}_t)|^{2+\nu}\right] < \infty$*, for some* $\nu > 0$.

Condition (i) of Assumption 3 is needed for establishing the results on the law of large numbers (LLN) of Lemma 4 of Li and Müller (2009), which we use throughout our proofs. Condition (ii) of Assumption 3 implies $n^{-1} \sum_{t=1}^{n} \mathrm{E}_{t-1}^{\mathrm{Q}_0}\left[|g(\mathbf{y}_{t-1}, \mathbf{y}_t)|^{2+\nu}\right] = O_p(1)$ and $n^{-1} \sum_{t=1}^{n} |g(\mathbf{y}_{t-1}, \mathbf{y}_t)|^{2+\nu} = O_p(1)$. Condition (ii) is needed for establishing the local asymptotic normality (LAN) for time-inhomogeneous Markov processes (see Proposition 3 in Appendix 2.1) and thus ensuring that the locally unstable DGP is contiguous to the stable DGP (see Corollary 1 in Online Appendix 2.1). Condition (ii) is also a commonly adopted assumption (e.g., Li and Müller, 2009, Lemma 1). A direct implication of Assumption 3 is the LLN and CLT of partial summations of score functions.

Assumption 4 (Global Identification Condition). *There exists $\epsilon > 0$ such that $\vartheta(\mathrm{Q}_{s,f})$ is unique if it exists, for all $\mathrm{Q}_{s,f} \in \mathcal{N}(\mathrm{Q}_0)$ with the Hellinger distance $\mathbf{H}^2(\mathrm{Q}_{s,f}, \mathrm{Q}_0) < \epsilon$.*

The following are regularity conditions on moments.

Assumption 5 (Tail Properties of Moments). *We assume that the moment function $m(\cdot, \theta)$, defined on a compact set Θ, satisfies the following conditions:*

(i) $\mathrm{E}^{\mathrm{Q}_0}\left[|m_t(\theta_0)|^{2+\nu}\right] < \infty$ *for some* $\nu > 0$*, and* $\mathrm{E}^{\mathrm{Q}_0}\left[\sup_{\theta \in \Theta} ||\nabla_\theta m_t(\theta)||_{\mathcal{S}}^2\right] < \infty$*,*

(ii) $n^{-1/2} \max_{1 \leq t \leq n} |m_t(\theta_0)| = o_p(1)$*,*

(iii) $\sum_{t=1}^{\infty} \sqrt{\mathrm{E}^{\mathrm{Q}_0}\left[|\gamma_t|^2\right]} < \infty$*, with* $\gamma_t \equiv \mathrm{E}^{\mathrm{Q}_0}\left[m_t(\theta_0)|\mathcal{F}_1\right] - \mathrm{E}^{\mathrm{Q}_0}\left[m_t(\theta_0)|\mathcal{F}_0\right]$*,*

where $|| \cdot ||_{\mathcal{S}}$ is the spectral norm of matrices, and the information set $\mathcal{F}_t$ is the sigma-field generated by $\{\mathbf{y}_{t-j}\}_{j=0}^{\infty}$.

Conditions (i) and (ii) of Assumption 5 are needed to establish the functional central limit theorem (invariance principle) of McLeish (1975b) and Phillips and Durlauf (1986). Condition (i) imposes restrictions on the amount of heteroskedasticity allowed in the observed moment series and their gradients, which also ensures the uniform square integrability of the moment function. This condition is commonly adopted in the literature (e.g., Newey, 1985; Andrews, 1993; Sowell, 1996; Li and Müller, 2009, for similar regularity conditions). Condition (iii) states that the incremental information about the current moments between two consecutive information sets eventually becomes negligible as the information sets recede in history from the current observation. This condition ensures the martingale difference approximation for the temporal-dependent moment function as in Hansen (1985), which plays a key role in analyzing the semiparametric efficiency bound based on unconditional moment restrictions (see Proposition 5 in Online Appendix 2.2 and Theorem 1 in Online Appendix 3).

Assumption 6 (Correct Baseline Structural Model). *We assume that the true local DGP with a joint distribution $\mathbb{P}_{1/\sqrt{n}, g^*, b^*}$ is such that $g^* \in \mathcal{G}_B(\mathrm{Q}_0)$, where*

$$\mathcal{G}_B(\mathrm{Q}_0) \equiv \left\{ g \in \mathcal{G}(\mathrm{Q}_0) : \lambda^{(1)}(g_1) = 0 \text{ and } \lambda^{(1)}(g_2) \in \mathit{lin}(D_{11}) \right\}. \tag{1}$$

Linear operator $\lambda^{(1)}(\cdot)$ is defined in (79).

Assumption 6 ensures that the baseline structural model is correctly specified because $\lambda^{(1)}(f_{n,t}) = \lambda^{(1)}(g_2)b(t/n) \in \mathsf{lin}(D_{11})$ for every $t \in \{1, \cdots, n\}$. We can replace (1) with a seemingly weaker assumption $\lambda^{(1)}(g_1^*), \lambda^{(1)}(g_2^*) \in \mathsf{lin}(D_{11})$. However, this does not add generality because we can always replace θ_0 with a sequence of new reference points (reparametrization) to ensure that (1) is satisfied.

2 Auxiliary Results

2.1 Auxiliary Results on Data-Generating Processes

In this section, we introduce auxiliary propositions that characterize the useful properties of the data-generating processes under the regularity conditions. Proposition 1 derives the corresponding scores (or local perturbations) of the univariate marginal distribution $\mu_{s,f}$ and the Markov transition kernel $K_{s,f}$ when we perturb the bivariate distribution from Q_0 to $\mathrm{Q}_{s,f}$. Proposition 2 considers local data-generating processes characterized by scores $f_{n,t}$ and shows that the scores $f_{n,t}$ satisfy the law of large numbers and the central limit theorem. Proposition 2, together with Hellinger-differentiability, is needed to ensure the local asymptotic normality of the local data-generating processes, as established in Proposition 3. The LAN property is needed to establish the contiguity property of the locally unstable data-generating process $\mathbb{P}_{1/\sqrt{n},g,b}$ as a local perturbation with respect to the reference process $\mathbb{P}_0$ for asymptotic equivalence arguments. We denote $\sum_{t=1}^{\lfloor \pi n \rfloor}$ by $\sum_{t \leq \pi n}$ and $\sum_{t=\lfloor \pi n \rfloor+1}^{n}$ by $\sum_{t>\pi n}$ for notational simplicity.

Proposition 1 (Implied Scores of Marginal and Transition Distributions)**.** *Suppose $\mathrm{Q}_{s,f} \in \mathcal{N}(\mathrm{Q}_0)$ for some $\mathrm{Q}_0 \in \mathcal{H}$. Let μ and K be the univariate marginal distribution and the Markov transition kernel of Q_0, respectively. Then, the marginal distribution $\mu_{s,f}$ and Markov transition kernel $K_{s,f}$ of $\mathrm{Q}_{s,f}$ satisfy the Hellinger differentiability conditions:*

$$\frac{d\mu_{s,f}}{d\mu_0} = 1 + s\bar{f} + s\Delta_\mu(s) \quad \textit{and} \quad \frac{dK_{s,f}(\cdot|y)}{dK_0(\cdot|y)} = 1 + s\tilde{f}(y,\cdot) + s\Delta_K(y,s) \ \forall\ y \in \mathcal{Y}, \tag{2}$$

where $\Delta_\mu(s)$ and $\Delta_K(y,s)$ converge to 0 in $L^2(\mathrm{Q}_0)$ for all $y \in \mathcal{Y}$ as $s \to 0$, and the marginal score and the conditional score are

$$\bar{f}(\mathbf{y}) \equiv \mathrm{E}^{\mathrm{Q}_0}\left[f(\mathbf{y},\mathbf{y}')|\mathbf{y}\right] = \mathrm{E}^{\mathrm{Q}_0}\left[f(\mathbf{y}',\mathbf{y})|\mathbf{y}\right] \quad \textit{and} \quad \tilde{f}(\mathbf{y},\mathbf{y}') \equiv f(\mathbf{y},\mathbf{y}') - \bar{f}(\mathbf{y}). \tag{3}$$

Proposition 2. *Suppose Assumption 3 holds. Let $\tilde{f}_{n,t} \equiv f_{n,t} - \mathrm{E}_{t-1}^{\mathrm{Q}_0}\left[f_{n,t}\right]$ and $\tilde{g}(\mathbf{y}_{t-1},\mathbf{y}_t) \equiv g(\mathbf{y}_{t-1},\mathbf{y}_t) - \mathrm{E}_{t-1}^{\mathrm{Q}_0}\left[g(\mathbf{y}_{t-1},\mathbf{y}_t)\right]$. Then it holds that under Q_0,*

$$n^{-1}\sum_{t\leq \pi n} \tilde{f}_{n,t}^2 \xrightarrow{p} \Upsilon(\pi) \quad \textit{and} \quad n^{-1}\sum_{t\leq \pi n} \mathrm{E}_{t-1}^{\mathrm{Q}}\left[\tilde{f}_{n,t}^2\right] \xrightarrow{p} \Upsilon(\pi), \ \textit{where} \tag{4}$$

$$\Upsilon(\pi) \equiv \mathrm{E}^{\mathrm{Q}}\left[\tilde{g}^T B_\pi \tilde{g}\right] \ \textit{with}\ B_\pi \equiv \begin{bmatrix} \pi & \int_0^\pi b(u)du \\ \int_0^\pi b(u)du & \int_0^\pi b(u)^2du \end{bmatrix}. \tag{5}$$

Further, the asymptotic normality result follows:

$$\operatorname*{wlim}_{n\to\infty} n^{-1/2} \sum_{t\le \pi n} \tilde{f}_{n,t} = N\left(0, \Upsilon(\pi)\right). \tag{6}$$

Proposition 3 (LAN of Unstable Parametric Submodels)**.** *Suppose Assumption 3 holds. For any $g \in \mathcal{G}(\mathrm{Q}_0)$ and $b \in \mathcal{B}$, the corresponding locally unstable data-generating process with distribution $\mathbb{P}_{1/\sqrt{n},g,b}$ for $\mathbf{y}^n = \{\mathbf{y}_0, \cdots, \mathbf{y}_n\}$ satisfies*

$$\ln \frac{d\mathbb{P}_{1/\sqrt{n},g,b}}{d\mathbb{P}_0} = \frac{1}{\sqrt{n}} \sum_{t\le n} \tilde{g}(\mathbf{y}_{t-1}, \mathbf{y}_t)^T \begin{bmatrix} 1 \\ b(t/n) \end{bmatrix} - \frac{1}{2}\Upsilon(1) + o_p(1),$$

where $\tilde{g}$ and $\Upsilon(\cdot)$ are defined in Proposition 2, and $o_p(1)$ denotes a sequence of random variables that converge to zero in probability $\mathbb{P}_0$.

Corollary 1 (Contiguity)**.** *Suppose Assumption 3 holds. The locally unstable data-generating process with distribution $\mathbb{P}_{1/\sqrt{n},g,b}$ is contiguous to the stable data-generating process with distribution $\mathbb{P}_0$. More precisely, $X_n \xrightarrow{p} 0$ under $\mathbb{P}_0$ implies $X_n \xrightarrow{p} 0$ under $\mathbb{P}_{1/\sqrt{n},g,b}$ for all $\mathcal{F}^n$-measurable random variables $X_n : \mathcal{Y}^n \to \mathbb{R}$.*

2.2 Auxiliary Results on Moment Functions

In this section, we introduce the basic results (Proposition 4) extending the standard moment function approximations (Hansen, 1982). Similar results on the (functional) central limit theorem with local instability are developed and used in Andrews (1993), Sowell (1996), and Li and Müller (2009).

Define $\lambda(g^T) \equiv [\lambda(g_1), \lambda(g_2)]$ for all $g = [g_1, g_2]^T$ with $g \in \mathcal{G}(\mathrm{Q}_0)$. We denote

$$\nu_e(g, b, \pi) \equiv \frac{\lambda(g^T)}{\sqrt{\pi}} \begin{bmatrix} \pi \\ \int_0^{\pi} b(u)\mathrm{d}u \end{bmatrix} \text{ and } \nu_o(g, b, \pi) \equiv \frac{\lambda(g^T)}{\sqrt{1-\pi}} \begin{bmatrix} 1-\pi \\ \int_{\pi}^{1} b(u)\mathrm{d}u \end{bmatrix}. \tag{7}$$

Proposition 4. *Suppose Assumptions 1 – 5 hold. Then, under $\mathbb{P}_{1/\sqrt{n},g,b}$,*

(i) $\operatorname*{wlim}_{n\to\infty} \begin{bmatrix} \frac{1}{\sqrt{\pi n}} \sum_{t\le \pi n} m_t(\theta_0) \\ \frac{1}{\sqrt{(1-\pi)n}} \sum_{t>\pi n} m_t(\theta_0) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{\pi}} W(\pi) \\ \frac{1}{\sqrt{1-\pi}} (W(1) - W(\pi)) \end{bmatrix} + \begin{bmatrix} \nu_e(g, b, \pi) \\ \nu_o(g, b, \pi) \end{bmatrix}$ *on $\mathcal{D}([0,1])$ for all split point $\pi \in [0,1]$, where $W(\pi)$ is a d_m-dimensional Wiener process and $\mathcal{D}([0,1])$ is the space of right continuous functions on $[0,1]$ endowed with the Skorohod J_1 topology;*

(ii) $\begin{bmatrix} \frac{1}{\sqrt{\pi n}} \sum_{t\le \pi n} m_t(\theta_{n,t}) \\ \frac{1}{\sqrt{(1-\pi)n}} \sum_{t>\pi n} m_t(\theta_{n,t}) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{\pi n}} \sum_{t\le \pi n} m_t(\theta_0) \\ \frac{1}{\sqrt{(1-\pi)n}} \sum_{t>\pi n} m_t(\theta_0) \end{bmatrix} - \begin{bmatrix} \nu_e(g, b, \pi) \\ \nu_o(g, b, \pi) \end{bmatrix} + o_p(1)$*, for all random variables $g_1, g_2 \in \mathcal{T}(\mathrm{Q}_0)$;*

(iii) $\begin{bmatrix} \frac{1}{\sqrt{\pi n}} \sum_{t\le \pi n} m_t(\hat{\theta}_{e,n}) \\ \frac{1}{\sqrt{(1-\pi)n}} \sum_{t>\pi n} m_t(\hat{\theta}_{e,n}) \end{bmatrix} = \begin{bmatrix} \left[I - D(D^TD)^{-1}D^T\right] \frac{1}{\sqrt{\pi n}} \sum_{t\le \pi n} m_t(\theta_0) \\ \frac{1}{\sqrt{(1-\pi)n}} \sum_{t>\pi n} m_t(\theta_0) - D(D^TD)^{-1}D^T \frac{1}{\sqrt{\pi n}} \sum_{t\le \pi n} m_t(\theta_0) \end{bmatrix} + o_p(1),$
where $\hat{\theta}_{e,n}$ is the efficient GMM estimator based on estimation sample $\mathbf{y}_e^n$ and D is the Jacobian matrix

evaluated at θ_0;

(iv)
$$\begin{bmatrix} \frac{1}{\sqrt{\pi n}} \sum_{t\leq \pi n} m_t(\tilde{\theta}_{e,n}) \\ \frac{1}{\sqrt{(1-\pi)n}} \sum_{t>\pi n} m_t(\tilde{\theta}_{e,n}) \end{bmatrix} = \begin{bmatrix} \left[I - D(A^TD)^{-1}A^T\right] \frac{1}{\sqrt{\pi n}} \sum_{t\leq \pi n} m_t(\theta_0) \\ \frac{1}{\sqrt{(1-\pi)n}} \sum_{t>\pi n} m_t(\theta_0) - D(A^TD)^{-1}A^T \frac{1}{\sqrt{\pi n}} \sum_{t\leq \pi n} m_t(\theta_0) \end{bmatrix} + o_p(1),$$
where $\tilde{\theta}_{e,n}$ is the recursive GMM estimator based on estimation sample $\mathbf{y}_e^n$, D is the Jacobian matrix evaluated at θ_0, and A is defined by $A \equiv \begin{bmatrix} D_{11} & 0 \\ 0 & A_{22} \end{bmatrix}$ and $A_{22} = \left[D_{21}(D_{11}^T D_{11})^{-1} D_{21}^T + I\right]^{-1} D_{22}$.

We construct the martingale difference array $h(\mathbf{y}, \mathbf{y}', \theta_0)$ inspired by the martingale difference approximation for the temporal-dependent moment function in Hansen (1985). The martingale difference approximation plays a key role in analyzing the semiparametric efficiency bound of estimation based on moment restrictions. To guarantee that $h(\mathbf{y}, \mathbf{y}', \theta_0)$ is well defined in (8), we postulate the condition of asymptotic negligibility of innovations (Assumption 5 (iii)), which has been used to establish Gordin's CLT (Gordin, 1969).

Proposition 5. *Suppose Assumptions 1 – 5 hold. Then $h(\cdot, \theta_0)$ is defined as follows:*

$$\begin{aligned} h(\mathbf{y}, \mathbf{y}', \theta_0) &= m(\mathbf{y}, \mathbf{y}', \theta_0) - \mathrm{E}^{\mathrm{Q}_0}\left[m_1(\theta_0) | \mathbf{y}_0 = \mathbf{y}\right] \\ &\quad + \sum_{t=1}^{\infty} \left\{ \mathrm{E}^{\mathrm{Q}_0}\left[m_{t+1}(\theta_0) | \mathbf{y}_1 = \mathbf{y}'\right] - \mathrm{E}^{\mathrm{Q}_0}\left[m_{t+1}(\theta_0) | \mathbf{y}_0 = \mathbf{y}\right] \right\}. \end{aligned} \tag{8}$$

Moreover, $h(\cdot, \theta_0)$ satisfies $\mathrm{E}^{\mathrm{Q}_0}\left[h(\mathbf{y}, \mathbf{y}', \theta_0) | \mathbf{y}\right] = 0$ and $\mathrm{E}^{\mathrm{Q}_0}\left[h(\mathbf{y}, \mathbf{y}', \theta_0) h(\mathbf{y}, \mathbf{y}', \theta_0)^T\right] = I$ and

$$\mathrm{E}^{\mathrm{Q}_0}\left[m(\cdot, \theta_0) f\right] = \mathrm{E}^{\mathrm{Q}_0}\left[h(\cdot, \theta_0) f\right] \text{ for all } f \in L_0^2(\mathrm{Q}_0). \tag{9}$$

Therefore, the tangent set of $\mathcal{Q}$ at the distribution Q_0 can be represented by

$$\mathcal{T}(\mathrm{Q}_0) = \left\{ f \in L_0^2(\mathrm{Q}_0) : \ \lambda(f) \in \mathsf{lin}(D) \right\}, \tag{10}$$

where the operator $\lambda(f) \equiv \mathrm{E}^{\mathrm{Q}_0}\left[h(\cdot, \theta_0) f\right]$ is a linear operator on $L_0^2(\mathrm{Q}_0)$, and the linear space $\mathsf{lin}(D)$ is spanned by columns of D, the Jacobian matrix evaluated at θ_0.

2.3 Auxiliary Results on GMM Estimators Based on the Estimation Sample

We now introduce the basic results that extend the standard GMM approximations (Hansen, 1982) in Proposition 6. Then, we introduce a new set of GMM approximations in Proposition 7, which are novel contributions of this paper.

Proposition 6. *Suppose Assumptions 1 – 5 hold. Let $\tilde{\theta}_{e,n}$ and $\hat{\theta}_{e,n}$ be the recursive GMM and the efficient GMM estimators based on the estimation sample $\mathbf{y}_e^n = \{\mathbf{y}_1, \cdots, \mathbf{y}_{\lfloor \pi n \rfloor}\}$, respectively. Then, under $\mathbb{P}_{1/\sqrt{n}, g, b}$,*

(i) $\sqrt{\pi n}\left(\tilde{\theta}_{e,n} - \theta_0\right) = -(A^TD)^{-1}A^T \left[\frac{1}{\sqrt{\pi n}} \sum_{t\leq \pi n} m_t(\theta_0)\right] + o_p(1),$

with $A = \begin{bmatrix} D_{11} & 0 \\ 0 & A_{22} \end{bmatrix}$ and $A_{22} = \left[D_{21}(D_{11}^T D_{11})^{-1} D_{21}^T + I\right]^{-1} D_{22}$;

(ii) $\sqrt{\pi n}\left(\hat{\theta}_{e,n}-\theta_0\right)=-(D^TD)^{-1}D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\leq\pi n}m_t(\theta_0)\right]+o_p(1).$

Proposition 7. *Suppose Assumptions 1 – 6 hold and $g\in\mathcal{G}_B(\mathrm{Q}_0)$. Let $\tilde{\theta}_{e,n}$ and $\hat{\theta}_{e,n}$ be the recursive GMM estimator and efficient GMM estimator based on the estimation sample $\mathbf{y}_e^n=\{\mathbf{y}_1,\cdots,\mathbf{y}_{\lfloor\pi n\rfloor}\}$, respectively. Then, under $\mathbb{P}_{1/\sqrt{n},g,b}$,*

$$(i)\quad \sqrt{\pi n}\begin{bmatrix}\tilde{\theta}^{(1)}_{e,n}-\theta^{(1)}_{e,n}\\ \psi_s(\tilde{\theta}^{(1)}_{e,n})-\psi_s(\theta^{(1)}_{e,n})\end{bmatrix}=-\mathbf{I}_{\mathrm{Q}}^{-1}\Gamma_{\theta,1}^T\mathbf{I}_F\mathbf{I}_B^{-1}D_{11}^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\leq\pi n}m_t^{(1)}(\theta^{(1)}_{n,t})\right]+o_p(1);$$

$$(ii)\quad \sqrt{\pi n}\begin{bmatrix}\hat{\theta}^{(1)}_{e,n}-\theta^{(1)}_{e,n}\\ \psi_s(\hat{\theta}^{(1)}_{e,n})-\psi_s(\theta^{(1)}_{e,n})\end{bmatrix}=-\mathbf{I}_{\mathrm{Q}}^{-1}\Gamma_{\theta,1}^T\mathbf{I}_F\left\{L_F\left[\frac{1}{\sqrt{\pi n}}\sum_{t\leq\pi n}m_t(\theta_0)\right]-L_B\nu_e(g,b,\pi)\right\}+o_p(1).$$

Here the matrices L_B and L_F are

$$L_B\equiv\mathbf{I}_B^{-1}D_{11}^T\Gamma_{m,1}\quad\text{and}\quad L_F\equiv\Gamma_{\theta,1}\mathbf{I}_{\mathrm{Q}}^{-1}D^T,\tag{11}$$

and D_{11} and D are the respective Jacobian matrices for the baseline and full model, $\mathbf{I}_B$ and $\mathbf{I}_{\mathrm{Q}}$ are the respective information matrices for the baseline and full model, and the selection matrices $\Gamma_{m,1}$ and $\Gamma_{\theta,1}$ are defined by $\Gamma_{m,1}\equiv\left[I,0_{d_{m,1}\times(d_m-d_{m,1})}\right]$ and $\Gamma_{\theta,1}\equiv\left[I,0_{d_{\theta,1}\times(d_\theta-d_{\theta,1})}\right]$.

Proposition 8. *Suppose Assumptions 1 – 6 hold and $g\in\mathcal{G}_B(\mathrm{Q}_0)$. Let $\mathcal{L}(\theta^{(1)},\cdot)$ be the loss function for assessing the goodness of fit of the baseline parameter $\theta^{(1)}$ to the data as defined in Chen, Dou, and Kogan (2021). Let $\tilde{\theta}_{e,n}$ and $\hat{\theta}_{e,n}$ be the recursive GMM estimator and efficient GMM estimator based on the estimation sample $\mathbf{y}_e^n=\{\mathbf{y}_1,\cdots,\mathbf{y}_{\lfloor\pi n\rfloor}\}$, respectively. Let $\mathbf{y}_o^n=\{\mathbf{y}_{\lfloor\pi n\rfloor+1},\cdots,\mathbf{y}_n\}$ be the holdout sample. Then, under $\mathbb{P}_{1/\sqrt{n},g,b}$,*

$$(i)\quad \begin{bmatrix}\mathcal{L}(\tilde{\theta}^{(1)}_{e,n};\mathbf{y}_e^n)\\ \mathcal{L}(\tilde{\theta}^{(1)}_{e,n};\mathbf{y}_o^n)\end{bmatrix}=\begin{bmatrix}((L_B-2L_F)\zeta_{e,n}-2L_\Delta\nu_e)^T\mathbf{I}_F(L_B\zeta_{e,n})\\ (L_B\zeta_{e,n}-2L_F\zeta_{o,n}-2L_\Delta\nu_o)^T\mathbf{I}_F(L_B\zeta_{e,n})\end{bmatrix}+o_p(1),\ \text{and}$$

$$(ii)\quad \begin{bmatrix}\mathcal{L}(\hat{\theta}^{(1)}_{e,n};\mathbf{y}_e^n)\\ \mathcal{L}(\hat{\theta}^{(1)}_{e,n};\mathbf{y}_o^n)\end{bmatrix}=\begin{bmatrix}-(L_F\zeta_{e,n}+L_\Delta\nu_e)^T\mathbf{I}_F(L_F\zeta_{e,n}+L_\Delta\nu_e)\\ (L_F(\zeta_{e,n}-2\zeta_{o,n})+L_\Delta(\nu_e-2\nu_o))^T\mathbf{I}_F(L_F\zeta_{e,n}+L_\Delta\nu_e)\end{bmatrix}+o_p(1),$$

where $\nu_e(g,b,\pi)$ and $\nu_o(g,b,\pi)$ are defined in (7), and the random vectors $\zeta_{e,n}$ and $\zeta_{o,n}$ are

$$\zeta_{e,n}\equiv\frac{1}{\sqrt{\pi n}}\sum_{t\leq\pi n}m_t(\theta_0)-\nu_e(g,b,\pi)\ \text{and}\ \zeta_{o,n}\equiv\frac{1}{\sqrt{\pi n}}\sum_{t>\pi n}m_t(\theta_0)-\nu_o(g,b,\pi),\tag{12}$$

and the matrices L_F, L_B are defined in (11) and $L_\Delta\equiv L_F-L_B$. Further, using Proposition 4,

$$\underset{n\to\infty}{\mathrm{wlim}}\begin{bmatrix}\zeta_{e,n}\\ \zeta_{o,n}\end{bmatrix}=\begin{bmatrix}\frac{1}{\sqrt{\pi}}W(\pi)\\ \frac{1}{\sqrt{1-\pi}}(W(1)-W(\pi))\end{bmatrix}.\tag{13}$$

3 Semiparametric Minimax Efficiency Bounds

Given the LAN for the Markov processes with potential local instability, the local asymptotic minimax (LAM) justification for the efficiency bounds can be established using the asymptotic equivalence argument.[1] For the local data-generating process that is described by a locally unstable distribution $\mathbb{P}_{1/\sqrt{n},g,b}$, the goal is to estimate the average model parameter value:

$$\vartheta(\mathbb{P}_{1/\sqrt{n},g,b}) \equiv \frac{1}{n}\sum_{t=1}^{n}\vartheta(\mathrm{Q}_{1/\sqrt{n},f_{n,t}}), \text{ with } f_{n,t} = g_1(\mathbf{y}_{t-1},\mathbf{y}_t) + g_2(\mathbf{y}_{t-1},\mathbf{y}_t)b(t/n). \tag{14}$$

We formalize the precise meaning of semiparametric efficiency bounds based on local asymptotic minimax risk, which is stated in the following theorem.

Theorem 1 (LAM Lower Bounds)**.** *Suppose assumptions 1 – 5 hold and $\vartheta(\mathbb{P}_{1/\sqrt{n},g,b})$ exists . Thus, for any $v \in \mathbb{R}^{d_\theta}$, any arbitrary estimator sequence $\check{\theta}_n$ satisfies*

$$\lim_{l\to\infty}\liminf_{n\to\infty}\sup_{g\in\mathcal{G}(\mathrm{Q}_0),b\in\mathcal{B}}\int l\wedge\left[\sqrt{n}v^T\left(\check{\theta}_n-\vartheta(\mathbb{P}_{1/\sqrt{n},g,b})\right)\right]^2 d\mathbb{P}_{1/\sqrt{n},g,b} \geq v^T(D^TD)^{-1}v.$$

The method of first calculating the truncated mean squared error (MSE), then letting the ceiling l increase to infinity, is widely adopted in the literature (e.g., Bickel, 1981; Le Cam and Yang, 2000; Kitamura, Otsu, and Evdokimov, 2013).

Theorem 2 (LAM Upper Bounds)**.** *Suppose assumptions 1 – 5 hold and $\vartheta(\mathbb{P}_{1/\sqrt{n},g,b})$ exists. Then, for any $v \in \mathbb{R}^{d_\theta}$, there exists an estimator sequence $\hat{\theta}_n$ such that*

$$\lim_{l\to\infty}\liminf_{n\to\infty}\sup_{g\in\mathcal{G}(Q_0),b\in\mathcal{B}}\int l\wedge\left[\sqrt{n}v^T\left(\hat{\theta}_n-\vartheta(\mathbb{P}_{1/\sqrt{n},g,b})\right)\right]^2 d\mathbb{P}_{1/\sqrt{n},g,b} \leq v^T(D^TD)^{-1}v.$$

In our proof, we show that the efficient GMM estimator (Hansen, 1982) can achieve the semiparametric efficiency bound. Importantly, the proof is similar to that of Theorem 1 in Li and Müller (2009) through using Le Cam's theory of asymptotic equivalence. Therefore, Theorems 1 and 2 of the Online Appendix extend the results on the minimax efficiency bounds for unconditional moment restrictions developed in Levit (1976), Nevelson (1977), and Chamberlain (1987, Theorem 2) to general Markov processes with local instability.

Proof of Theorem 1 of the Online Appendix. The local asymptotic normality (LAN) (see Proposition 3), as well as the implied contiguity, and Le Cam's first and third lemmas play crucial roles in the proof as in the standard proof of semiparametric minimax lower bounds (e.g. van der Vaart, 1998, Theorem 8.11 and Theorem 25.21). Our results are new in the sense that they apply to Markov processes with local instability, which is more general than the i.i.d. case.

Following the literature (e.g. Bickel, Klaassen, Ritov, and Wellner, 1993; van der Vaart, 1998), we define the functional $\vartheta(\mathrm{Q})$ to be pathwise differentiable at Q_0 relative to the parametric submodels $s \mapsto \mathrm{Q}_{s,f}$, if

[1] Dou, Pollard, and Zhou (2010) also appeal to the asymptotic equivalence argument to establish the global minimax upper bound for a non-parametric estimation problem.

there exists a measurable function $\dot{\vartheta}$: $\mathcal{Y} \times \mathcal{Y} \to \mathbb{R}^{d_\theta}$ with $\dot{\vartheta} \in L_0^2(\mathrm{Q}_0)$ such that

$$\lim_{s \to 0} \frac{1}{s} \left[\vartheta(\mathrm{Q}_{s,f}) - \vartheta(\mathrm{Q}_0) \right] = \mathrm{E}^{\mathrm{Q}_0} \left[\dot{\vartheta} f \right], \tag{15}$$

where $\dot{\vartheta}(\mathbf{y}_{t-1}, \mathbf{y}_t) \equiv (D^T D)^{-1} D^T h(\mathbf{y}_{t-1}, \mathbf{y}_t, \theta_0)$ with $h(\mathbf{y}_{t-1}, \mathbf{y}_t, \theta_0)$ defined in Proposition 5 (e.g., Greenwood and Wefelmeyer, 1995). According to Proposition 5, $h(\cdot, \theta_0)$ satisfies the conditions: $\mathrm{E}^{\mathrm{Q}_0} \left[h(\mathbf{y}, \mathbf{y}', \theta_0) | \mathbf{y} \right] = 0$ and $\mathrm{E}^{\mathrm{Q}_0} \left[h(\mathbf{y}, \mathbf{y}', \theta_0) h(\mathbf{y}, \mathbf{y}', \theta_0)^T \right] = I$.

First, we only need to consider the case $g_1(\mathbf{y}, \mathbf{y}') = v^T \dot{\vartheta}(\mathbf{y}, \mathbf{y}')$, $g_2(\mathbf{y}_{t-1}, \mathbf{y}_t) \equiv 0$, and $b(u) \equiv 0$ for establishing the lower bound. In such case, $f(\mathbf{y}_{t-1}, \mathbf{y}_t) \equiv g_1(\mathbf{y}_{t-1}, \mathbf{y}_t)$ for all $1 \leq t \leq n$. Second, we further focus on the estimators $\check{\theta}_n$ such that $\sqrt{n} \left(\check{\theta}_n - \theta_0 \right)$ is uniformly tight under the distribution $\mathbb{P}_0$, similar to van der Vaart (1998). The tightness assumption can be dropped by a compactification argument (e.g. van der Vaart, 1988; van der Vaart and Wellner, 1996, Chapter 3.11). Moreover, without loss of generality, due to Prohorov's theorem, we can assume that

$$\operatorname*{wlim}_{n \to \infty} \left(\sqrt{n} \left(v^T \check{\theta}_n - v^T \theta_0 \right), \ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} g_1(\mathbf{y}_{t-1}, \mathbf{y}_t) \right) = (\Xi_0, U_0), \tag{16}$$

where $U_0 \sim N(0, v^T (D^T D)^{-1} v)$ (see Proposition 2). Using the contiguity between $\mathbb{P}_{1/\sqrt{n}, g, 0}$ and $\mathbb{P}_0$, Le Cam's third lemma (e.g. van der Vaart, 1998, Theorem 6.6), and differentiability of $\vartheta(\mathrm{Q}_{s,f})$ with respect to s, we know that under the sequence of distributions $\mathbb{P}_{1/\sqrt{n}, g, 0}$,

$$\operatorname*{wlim}_{n \to \infty} \sqrt{n} \left(v^T \check{\theta}_n - v^T \vartheta(\mathbb{P}_{1/\sqrt{n}, g, 0}) \right) = \Xi_g, \tag{17}$$

where, appealing to Theorem 8.3 of van der Vaart (1998), the limiting random variable Ξ_g has the following representation with a certain measurable function $\tau: \ \mathbb{R}^{d_\theta} \to \mathbb{R}$:

$$\begin{aligned} \Xi_g &= \tau(X_g) - v^T \xi \\ &= \tau(X_g) - \mathrm{E}^{\mathrm{Q}_0} \left[v^T \dot{\vartheta} f \right] \\ &= \tau(X_g) - \left[v^T (D^T D)^{-1} v \right]. \end{aligned} \tag{18}$$

Here, the local estimation bias is $\xi \equiv (D^T D)^{-1} D^T \lambda(g_1) = (D^T D)^{-1} v$ (similar to Corollary 1 or the proof of Proposition 4 (ii)) and $X_g \sim N(\xi, (D^T D)^{-1})$. Based on Theorem 8.6 of van der Vaart (1998) for estimating normal means, it holds that for all measurable function τ,

$$\mathrm{E}^{\mathrm{Q}_{1/\sqrt{n}, f}} \left[\Xi_g^2 \right] \geq \mathrm{E}^{\mathrm{Q}_0} \left[\left(v^T X_0 \right)^2 \right] = v^T (D^T D)^{-1} v. \tag{19}$$

The key idea of (16) – (18) is a change-of-measure argument, inspired by Le Cam's theory of asymptotic equivalence, whose stronger form has also been developed and used in the minimax inference of Dou, Pollard, and Zhou (2010).

Consequently, it suffices to show that the left-hand side of (19) is a lower bound for the minimax risk R:

$$R \equiv \lim_{l \to \infty} \liminf_{n \to \infty} \int l \wedge \left[\sqrt{n} v^T \left(\check{\theta}_n - \vartheta(\mathbb{P}_{1/\sqrt{n}, g, 0}) \right) \right]^2 d\mathbb{P}_{1/\sqrt{n}, g, 0}. \tag{20}$$

In fact, it holds that

$$
\begin{aligned}
&\liminf_{n\to\infty} \int l \wedge \left[\sqrt{n} v^T \left(\check{\theta}_n - \vartheta(\mathbb{P}_{1/\sqrt{n},g,0})\right)\right]^2 d\mathbb{P}_{1/\sqrt{n},g,0} \\
&\quad \geq \liminf_{n\to\infty} \int l \wedge \left[\sqrt{n} v^T \left(\check{\theta}_n - \vartheta(\mathbb{P}_{1/\sqrt{n},g,0})\right)\right]^2 d\mathbb{P}_{1/\sqrt{n},g,0} \\
&\quad = \mathrm{E}^{\mathrm{Q}_{1/\sqrt{n},g,0}} \left[l \wedge \Xi_g^2\right].
\end{aligned}
$$

Thus, the minimax risk can be bounded from below by

$$
R \geq \lim_{l\to\infty} \mathrm{E}^{\mathrm{Q}_{1/\sqrt{n},f}} \left[l \wedge \Xi_g^2\right] \geq \lim_{l\to\infty} \mathrm{E}^{\mathrm{Q}_{1/\sqrt{n},f}} \left[l \wedge \Xi_g^2\right]. \tag{21}
$$

According to the monotone convergence theorem, it follow that

$$
R \geq \mathrm{E}^{\mathrm{Q}_{1/\sqrt{n},f}} \left[\Xi_g^2\right]. \tag{22}
$$

Combining (19) and (22), the local asymptotic minimax lower bound result holds: $R \geq v^T(D^TD)^{-1}v$.

Proof of Theorem 2 of the Online Appendix. We start with

$$
\sqrt{n}\left[\hat{\theta}_n - \vartheta(\mathbb{P}_{1/\sqrt{n},g,b})\right] = \sqrt{n}\left(\hat{\theta}_n - \theta_0\right) - \sqrt{n}\left[\vartheta(\mathbb{P}_{1/\sqrt{n},g,b}) - \theta_0\right]. \tag{23}
$$

According to Proposition 6 (ii), it follows that

$$
\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) = -(D^TD)^{-1}D^T\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} m_t(\theta_0)\right] + o_p(1). \tag{24}
$$

Consequently, similar to Corollary 1 or the proof of Proposition 4 (ii),

$$
\sqrt{n}\left[\vartheta(\mathbb{P}_{1/\sqrt{n},g,b}) - \theta_0\right] = -(D^TD)^{-1}D^T\lambda(g_1) + o(1). \tag{25}
$$

Thus, appealing to Proposition 4 (i), we can show that

$$
\operatorname*{wlim}_{n\to\infty} \sqrt{n}\left[\hat{\theta}_n - \vartheta(\mathbb{P}_{1/\sqrt{n},g,b})\right] = -(D^TD)^{-1}D^TW(1), \tag{26}
$$

where $W(\cdot)$ is a d_m-dimensional Wiener process. Therefore, for any $v \in \mathbb{R}^{d_\theta}$,

$$
\liminf_{n\to\infty} \int l \wedge \left[\sqrt{n} v^T \left(\hat{\theta}_n - \vartheta(\mathbb{P}_{1/\sqrt{n},g,b})\right)\right]^2 d\mathbb{P}_{1/\sqrt{n},g,b} = \mathrm{E}\left[l \wedge X^2\right], \quad \text{with } X \sim N(0, v^T(D^TD)^{-1}v). \tag{27}
$$

Let l increase monotonically to infinity, and using the monotonic convergence theorem, we obtain

$$
\lim_{l\to\infty} \liminf_{n\to\infty} \int l \wedge \left[\sqrt{n} v^T \left(\hat{\theta}_n - \vartheta(\mathbb{P}_{1/\sqrt{n},g,b})\right)\right]^2 d\mathbb{P}_{1/\sqrt{n},g,b} = \mathrm{E}\left[X^2\right] = v^T(D^TD)^{-1}v. \tag{28}
$$

4 Proofs of the Main Theorems, Propositions, and Corollaries

4.1 Proofs of the Main Theorems

Proof of Theorem 1 of Chen, Dou, and Kogan (2021). The test statistic based on the C statistic is $\hat{\varphi}_n \equiv \mathbf{1}_{\{C_n > c_{1-\alpha}\}}$, where $c_{1-\alpha}$ is the $(1-\alpha)$ quantile of a chi-square distribution with $d_{m,2} - d_{\theta,2}$ degrees of freedom. From Proposition 4, we know that Assumption 3.1 of Chen and Santos (2018) is satisfied. Thus, by Lemma 3.2 of Chen and Santos (2018) and the results of Newey (1985), it follows that for any GMM specification test $\check{\varphi}_n$ with an asymptotic level α and an asymptotic local power function ($\forall\ \check{\varphi}_n \in \Phi_\alpha(\mathrm{Q}_0)$),

$$\inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} \lim_{n\to\infty} \int \check{\varphi}_n d\mathbb{P}_{1/\sqrt{n},g,0} \le \inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} \lim_{n\to\infty} \int \hat{\varphi}_n d\mathbb{P}_{1/\sqrt{n},g,0} \quad \text{(i.e., } C \text{ test is asymptotically optimal)} \tag{29}$$

$$= \inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} \lim_{n\to\infty} \mathbb{P}_{1/\sqrt{n},g,0} \left\{ \left| \widehat{\mathbb{G}}_n \right|^2 > c_{1-\alpha} \right\}, \tag{30}$$

where $\mathcal{A}_\kappa(\mathrm{Q}_0) \equiv \left\{ g \in \mathcal{G}_{\mathrm{B}}(\mathrm{Q}_0) : |\lambda^{(2)}(g_1)| \ge \kappa \text{ and } \lambda^{(2)}(g_1) \perp \mathsf{lin}(D_{22}) \right\}$, and

$$\widehat{\mathbb{G}}_n = \left(\Lambda_2 - \Lambda_2 D_{21} \mathbf{I}_{\mathrm{F}}^{-1} D_{21}^T \Lambda_2 \right)^{-1/2} \left[\frac{1}{\sqrt{n}} \sum_{t=1}^n m_t^{(2)}(\hat{\theta}_n) \right]; \tag{31}$$

see page 243 of Newey (1985) and Online Appendix 5.4. Here $\Lambda_2 = I - D_{22}(D_{22}^T D_{22})^{-1} D_{22}^T$.

Now, we obtain (e.g., Newey, 1985; Chen and Santos, 2018, or Proposition 4 of this onine appendix)

$$\operatorname*{wlim}_{n\to\infty} |\widehat{\mathbb{G}}_n|^2 = \chi^2_{d_{m,2}-d_{\theta,2}}(\mu_g), \tag{32}$$

where $\chi^2_{d_{m,2}-d_{\theta,2}}(\mu_g)$ is a noncentral chi-squared random variable with degrees of freedom $d_{m,2} - d_{\theta,2}$ and the noncentrality parameter $\mu_g = \lambda^{(2)}(g_1)^T \left(\Lambda_2 - \Lambda_2 D_{21} \mathbf{I}_{\mathrm{F}}^{-1} D_{21}^T \Lambda_2 \right) \lambda^{(2)}(g_1)$.

Using (29) and (30), we conclude that

$$\inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} q(g, \check{\varphi}) \le \inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} \lim_{n\to\infty} \mathbb{P}_{1/\sqrt{n},g} \left\{ \left| \widehat{\mathbb{G}}_n \right|^2 > c_{1-\alpha} \right\} = \inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} P\left\{ \chi^2_{d_m - d_\theta}(\mu_g) > c_{1-\alpha} \right\}. \tag{33}$$

Note that $\mu_g > 0$ for all $g \in \mathcal{A}_\kappa(\mathrm{Q}_0)$, since $\Lambda_2 D_{21} \mathbf{I}_{\mathrm{F}}^{-1} D_{21}^T \Lambda_2$ does not have unit eigenvalues. The local asymptotic maximin power is then bounded from above by

$$\inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} q(g, \check{\varphi}) \le \inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} M_{\frac{d_{m,2}-d_{\theta,2}}{2}} \left(\sqrt{\mu_g}, \sqrt{c_{1-\alpha}} \right) = M_{\frac{d_{m,2}-d_{\theta,2}}{2}} \left(\inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} \sqrt{\mu_g}, \sqrt{c_{1-\alpha}} \right), \tag{34}$$

where the equality above is due to the continuity and monotonicity of the Marcum Q-function $M_\gamma(x_1, x_2)$.

Following the definition of μ_g and the fact that $\Lambda_2^2 = \Lambda_2$ as a projection matrix onto the linear space spanned by the column vectors of D_{22}, it holds that

$$\begin{aligned}\inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)}\mu_g &= \inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)}\lambda^{(2)}(g_1)^T\Lambda_2\left(I-\Lambda_2 D_{21}\mathbf{I}_{\mathrm{F}}^{-1}D_{21}^T\Lambda_2\right)\Lambda_2\lambda^{(2)}(g_1)\\ &= \inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)}|\lambda^{(2)}(g_1)^T\Lambda_2\lambda^{(2)}(g_1)|\times\text{the smallest eigenvalue of } I-\Lambda_2 D_{21}\mathbf{I}_{\mathrm{F}}^{-1}D_{21}^T\Lambda_2\\ &= \kappa^2\times\text{the smallest eigenvalue of } I-\Lambda_2 D_{21}\mathbf{I}_{\mathrm{F}}^{-1}D_{21}^T\Lambda_2,\end{aligned}$$

where the last equality is due to the definition of the set $\mathcal{A}_\kappa(\mathrm{Q}_0)$, in which $|\lambda^{(2)}(g_1)|\geq\kappa$ and $\lambda^{(2)}(g_1)\perp\mathsf{lin}(D_{22})$.

We shall now show that $1/(1+\varrho(\theta_0))$ is an eigenvalue of $I-\Lambda_2 D_{21}\mathbf{I}_{\mathrm{F}}^{-1}D_{21}^T\Lambda_2$, and thus $\inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)}\sqrt{\mu_g}\leq\sqrt{\kappa^2/(1+\varrho(\theta_0))}$. In fact, $1-1/(1+\varrho(\theta_0))$ is an eigenvalue of $\mathbf{I}_{\mathrm{F}}^{-1/2}\left(\mathbf{I}_{\mathrm{F}}-\mathbf{I}_{\mathrm{B}}\right)\mathbf{I}_{\mathrm{F}}^{-1/2}=\mathbf{I}_{\mathrm{F}}^{-1/2}(D_{21}^T\Lambda_2 D_{21})\mathbf{I}_{\mathrm{F}}^{-1/2}$, and thus an eigenvalue of $\Lambda_2 D_{21}\mathbf{I}_{\mathrm{F}}^{-1}D_{21}^T\Lambda_2$. Therefore, $1/(1+\varrho(\theta_0))$ is an eigenvalue of $I-\Lambda_2 D_{21}\mathbf{I}_{\mathrm{F}}^{-1}D_{21}^T\Lambda_2$.

Due to the monotonicity of the generalized Marcum Q-function, the local asymptotic maximin power is upper bounded by

$$\inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)}q(g,\check{\varphi})\leq M_{\frac{d_{m,2}-d_{\theta,2}}{2}}\left(\sqrt{\frac{\kappa^2}{1+\varrho(\theta_0)}},\sqrt{c_{1-\alpha}}\right). \tag{35}$$

Proof of Theorem 2 of Chen, Dou, and Kogan (2021). According to Proposition 8 (ii), it follows that

$$\begin{aligned}\mathrm{E}\left[\operatorname*{wlim}_{n\to\infty}\frac{1}{2}\left(\mathcal{L}(\hat{\theta}_{\mathrm{e,n}};\mathbf{y}_{\mathrm{o}}^n)-\mathcal{L}(\hat{\theta}_{\mathrm{e,n}};\mathbf{y}_{\mathrm{e}}^n)\right)\right] &= \pi^{-1}\mathrm{E}\left[W(\pi)^T L_F^T\mathbf{I}_{\mathrm{F}}L_F W(\pi)\right]\\ &\quad+\left[\nu_e(g,b,\pi)-\nu_o(g,b,\pi)\right]^T L_\Delta^T\mathbf{I}_{\mathrm{F}}L_\Delta\nu_e(g,b,\pi),\end{aligned} \tag{36}$$

where $\mathrm{wlim}_{n\to\infty}$ is the weak convergence limit and $W(\cdot)$ is a d_m-dimensional Wiener process, and $L_{\mathrm{B}}\equiv\mathbf{I}_{\mathrm{B}}^{-1}D_{11}^T\Gamma_{m,1}$, $L_{\mathrm{F}}\equiv\Gamma_{\theta,1}\mathbf{I}_{\mathrm{Q}}^{-1}D^T$, and $L_\Delta\equiv L_{\mathrm{F}}-L_{\mathrm{B}}$. The first term above is

$$\pi^{-1}\mathrm{E}\left[W(\pi)^T L_{\mathrm{F}}^T\mathbf{I}_{\mathrm{F}}L_{\mathrm{F}}W(\pi)\right]=\pi^{-1}\mathrm{E}\left[\mathbf{tr}\left(\mathbf{I}_{\mathrm{F}}^{1/2}L_{\mathrm{F}}W(\pi)W(\pi)^T L_{\mathrm{F}}^T\mathbf{I}_{\mathrm{F}}^{1/2}\right)\right] \tag{37}$$

$$=\mathbf{tr}\left(\mathbf{I}_{\mathrm{F}}^{1/2}L_{\mathrm{F}}L_{\mathrm{F}}^T\mathbf{I}_{\mathrm{F}}^{1/2}\right). \tag{38}$$

According to the definition of L_{F} in (11),

$$L_{\mathrm{F}}L_{\mathrm{F}}^T=\Gamma_{\theta,1}\mathbf{I}_{\mathrm{Q}}^{-1}\Gamma_{\theta,1}^T=\mathbf{I}_{\mathrm{F}}^{-1}. \tag{39}$$

Combining (38) and (39) yields

$$\pi^{-1}\mathrm{E}\left[W(\pi)^T L_{\mathrm{F}}^T\mathbf{I}_{\mathrm{F}}L_{\mathrm{F}}W(\pi)\right]=d_{\theta,1}. \tag{40}$$

Because $\lambda(g_1)\in\mathsf{lin}(D)$, it holds that $L_\Delta\lambda(g_1)=0$, and thus

$$\left[\nu_e-\nu_o\right]^T L_\Delta^T\mathbf{I}_{\mathrm{F}}L_\Delta\nu_e=\frac{1}{\sqrt{\pi}}\left(\frac{1}{\sqrt{\pi}}+\frac{1}{\sqrt{1-\pi}}\right)\left(\int_0^\pi b(u)\mathrm{d}u\right)^2\lambda(g_2)^T L_\Delta^T\mathbf{I}_{\mathrm{F}}L_\Delta\lambda(g_2). \tag{41}$$

The left-hand side of (41) is bounded from above by

$$\frac{1}{\sqrt{\pi}}\left(\frac{1}{\sqrt{\pi}}+\frac{1}{\sqrt{1-\pi}}\right)\left(\int_0^{\pi} b(u)\mathrm{d}u\right)^2 \lambda(g_2)^T L_\Delta^T \mathbf{I}_{\mathrm{F}} L_\Delta \lambda(g_2) \tag{42}$$

$$\leq \pi\left(1+\sqrt{\frac{\pi}{1-\pi}}\right)|\lambda(g_2)|^2 \times \text{the largest eigenvalue of } L_\Delta^T \mathbf{I}_{\mathrm{F}} L_\Delta. \tag{43}$$

The largest eigenvalue of $L_\Delta^T \mathbf{I}_{\mathrm{F}} L_\Delta$ is that of $\Pi = \mathbf{I}_{\mathrm{F}}^{1/2} L_\Delta L_\Delta^T \mathbf{I}_{\mathrm{F}}^{1/2}$, which is the dark matter measure $\varrho(\theta_0)$.

Proof of Theorem 3 of Chen, Dou, and Kogan (2021). According to Proposition 8 (i), it follows that

$$\mathrm{E}\left[\operatorname*{wlim}_{n\to\infty} \frac{1}{2}\left(\mathcal{L}(\tilde{\theta}_{\mathrm{e,n}};\mathbf{y}_{\mathrm{o}}^n)-\mathcal{L}(\tilde{\theta}_{\mathrm{e,n}};\mathbf{y}_{\mathrm{e}}^n)\right)\right] = \pi^{-1}\mathrm{E}\left[W(\pi)^T L_{\mathrm{F}}^T \mathbf{I}_{\mathrm{F}} L_{\mathrm{B}} W(\pi)\right], \tag{44}$$

where $\operatorname{wlim}_{n\to\infty}$ is the weak convergence limit and $W(\cdot)$ is a d_m-dimensional Wiener process. Further,

$$\pi^{-1}\mathrm{E}\left[W(\pi)^T L_{\mathrm{F}}^T \mathbf{I}_{\mathrm{F}} L_{\mathrm{B}} W(\pi)\right] = \mathbf{tr}(\mathbf{I}_{\mathrm{F}}^{1/2} L_{\mathrm{B}} L_{\mathrm{F}}^T \mathbf{I}_{\mathrm{F}}^{1/2}). \tag{45}$$

Because $L_{\mathrm{B}} L_{\mathrm{F}}^T = \mathbf{I}_{\mathrm{B}}^{-1} D_{11}^T \left[D_{11}, 0_{d_{m,1}\times(d_\theta - d_{\theta,1})}\right] \mathbf{I}_{\mathrm{Q}}^{-1}\Gamma_{\theta,1}^T = \Gamma_{\theta,1}\mathbf{I}_{\mathrm{Q}}^{-1}\Gamma_{\theta,1}^T = \mathbf{I}_{\mathrm{F}}^{-1}$, the equality (45) can further be rewritten as

$$\pi^{-1}\mathrm{E}\left[W(\pi)^T L_{\mathrm{F}}^T \mathbf{I}_{\mathrm{F}} L_{\mathrm{B}} W(\pi)\right] = d_{\theta,1}. \tag{46}$$

4.2 Proofs of Propositions

Proof of Proposition 1 of Chen, Dou, and Kogan (2021). Following the standard argument such as in the proof of Theorem 7.2 of van der Vaart (1998), we can show that $\mathrm{E}^{\mathrm{Q}_0}[f] = 0$. Thus,

$$\mathrm{E}^{\mathrm{Q}_0}[\Delta(s)] = \mathrm{E}^{\mathrm{Q}_0}\left[\frac{\mathrm{d}Q_{s,f}}{\mathrm{d}Q_0}-1\right] = \int \mathrm{d}Q_{s,f} - \int \mathrm{d}Q_0 = 0. \tag{47}$$

According to Proposition 1, the conditional expectations denoted by $\bar{f}(\mathbf{y}_{t-1}) = \mathrm{E}^{\mathrm{Q}_0}[f(\mathbf{y}_{t-1},\mathbf{y}_t)|\mathbf{y}_{t-1}]$ and $\bar{f}(\mathbf{y}_t) = \mathrm{E}^{\mathrm{Q}_0}[f(\mathbf{y}_{t-1},\mathbf{y}_t)|\mathbf{y}_t]$ are the scores for the marginal distributions of $\mathbf{y}_{t-1}$ and $\mathbf{y}_t$, respectively. Because the marginal distributions are constant over time,

$$\mathrm{E}^{\mathrm{Q}_0}\left[f(\mathbf{y},\mathbf{y}')|\mathbf{y}\right] = \mathrm{E}^{\mathrm{Q}_0}\left[f(\mathbf{y}',\mathbf{y})|\mathbf{y}\right]. \tag{48}$$

Proof of Proposition 2 of Chen, Dou, and Kogan (2021). According to Definition 4, it follows that

$$\mathrm{E}^{\mathrm{Q}_{1/\sqrt{n},f_{n,t}}}[m_t(\theta_0)] = \int m_t(\theta_0)\left[1+f_{n,t}/\sqrt{n}+\Delta_n\right]\mathrm{dQ}_0. \tag{49}$$

Because $\mathrm{E}^{\mathrm{Q}_0}[m_t(\theta_0)] = 0$, the equality (49) above leads to

$$\mathrm{E}^{\mathrm{Q}_{1/\sqrt{n},f_{n,t}}}[m_t(\theta_0)] = \frac{\lambda(g_1)+\lambda(g_2)b(t/n)}{\sqrt{n}} + \int m_t(\theta_0)\Delta_n \mathrm{dQ}_0. \tag{50}$$

Based on Assumption 5 and Definition 4, the Cauchy-Schwarz inequality leads to

$$|\int m_t(\theta_0)\Delta_n \mathrm{dQ}_0| \leq \mathrm{E}^{\mathrm{Q}_0}\left[|m_t(\theta_0)|^2\right]^{1/2}\mathrm{E}^{\mathrm{Q}_0}\left[|\Delta_n|^2\right]^{1/2} = o\left(\frac{1}{\sqrt{n}}\right). \tag{51}$$

Proof of Proposition 1 of the Online Appendix. By the definition of a marginal distribution,

$$\begin{aligned}
\mathrm{d}\mu_{s,f}(\mathbf{y}) &= \int_{\mathbf{y}'\in\mathcal{Y}} \mathrm{dQ}_{s,f}(\mathbf{y},\mathbf{y}') = \int_{\mathbf{y}'\in\mathcal{Y}}\left[1 + sf(\mathbf{y},\mathbf{y}') + s\Delta_{\mathrm{Q}}(s)\right]\mathrm{dQ}(\mathbf{y},\mathbf{y}')\\
&= \left[1 + s\int_{\mathbf{y}'\in\mathcal{Y}} f(\mathbf{y},\mathbf{y}')\mathrm{d}K_{s,f}(\mathbf{y}'|\mathbf{y}) + s\int_{\mathbf{y}'\in\mathcal{Y}}\Delta_{\mathrm{Q}}(s)\mathrm{d}K_{s,f}(\mathbf{y}'|\mathbf{y})\right]\mathrm{d}\mu(\mathbf{y}).
\end{aligned}$$

By the definition of $\bar{f}(\mathbf{y})$, we know that

$$\mathrm{d}\mu_{s,f}(\mathbf{y}) = \left[1 + s\bar{f}(\mathbf{y}) + s\Delta_\mu(s)\right]\mathrm{d}\mu(\mathbf{y}), \tag{52}$$

where $\Delta_\mu(s) \equiv \mathrm{E}^{\mathrm{Q}}\left[\Delta_{\mathrm{Q}}(s)|\mathbf{y}\right]$ and it converges to zero in quadratic mean under μ as $s \to 0$. Further, by definition, it holds that

$$\begin{aligned}
\mathrm{d}K_{s,f}(\mathbf{y}'|\mathbf{y}) &= \frac{\mathrm{dQ}_{s,f}(\mathbf{y},\mathbf{y}')}{\mathrm{d}\mu_{s,f}(\mathbf{y})} = \frac{1 + sf(\mathbf{y},\mathbf{y}') + s\Delta_{\mathrm{Q}}(s)}{1 + s\bar{f}(\mathbf{y}) + s\Delta_\mu(s)}\frac{\mathrm{dQ}(\mathbf{y},\mathbf{y}')}{\mathrm{d}\mu(\mathbf{y})}\\
&= \frac{1 + sf(\mathbf{y},\mathbf{y}') + s\Delta_{\mathrm{Q}}(s)}{1 + s\bar{f}(\mathbf{y}) + s\Delta_\mu(s)}\mathrm{d}K(\mathbf{y}'|\mathbf{y}).
\end{aligned}$$

Rearranging and combining terms leads to

$$\mathrm{d}K_{s,f}(\mathbf{y}'|\mathbf{y}) = \left\{1 + s\left[f(\mathbf{y},\mathbf{y}') - \bar{f}(\mathbf{y})\right] + s\Delta_K(\mathbf{y},s)\right\}\mathrm{d}K(\mathbf{y}'|\mathbf{y}), \tag{53}$$

where $\Delta_K(\mathbf{y},s)$ converges to zero in quadratic mean under $K(\mathbf{y}'|\mathbf{y})$ as $s \to 0$ for all $\mathbf{y} \in \mathcal{Y}$. By definition of $\tilde{f}(\mathbf{y},\mathbf{y}')$, it follows that $\mathrm{E}^{\mathrm{Q}}\left[\tilde{f}(\mathbf{y},\mathbf{y}')|\mathbf{y}\right] = 0$. Thus, similar to the proof of Proposition 1, we can show that $\mathrm{E}^{\mathrm{Q}}\left[\Delta_K(\mathbf{y},s)|\mathbf{y}\right] = 0$.

Proof of Proposition 2 of the Online Appendix. According to Assumption 3 (i),

$$n^{-1}\max_{1\leq t\leq n}|g(\mathbf{y}_{t-1},\mathbf{y}_t)|^2 \xrightarrow{p} 0. \tag{54}$$

According to simple algebra, we can show that

$$n^{-1}\sum_{t\leq\pi n}\tilde{f}_{n,t}^2 = n^{-1}\sum_{t\leq\pi n}\left[\tilde{g}_1(\mathbf{y}_{t-1},\mathbf{y}_t)^2 + 2\tilde{g}_1(\mathbf{y}_{t-1},\mathbf{y}_t)\tilde{g}_2(\mathbf{y}_{t-1},\mathbf{y}_t)b(t/n) + \tilde{g}_2(\mathbf{y}_{t-1},\mathbf{y}_t)^2 b(t/n)^2\right]. \tag{55}$$

Therefore, by Lemma 4 of Li and Müller (2009), it follows that

$$n^{-1}\sum_{t\leq\pi n}\tilde{f}_{n,t}^2 \xrightarrow{p} \mathrm{E}^{\mathrm{Q}_0}\left[\tilde{g}_1^2\right]\pi + 2\mathrm{E}^{\mathrm{Q}_0}\left[\tilde{g}_1\tilde{g}_2\right]\int_0^\pi b(u)\mathrm{d}u + \mathrm{E}^{\mathrm{Q}_0}\left[\tilde{g}_1^2\right]\int_0^\pi b(u)^2\mathrm{d}u,$$

and hence

$$n^{-1} \sum_{t \leq \pi n} \tilde{f}_{n,t}^2 \to \Upsilon(\pi) \equiv \mathrm{E}^{\mathrm{Q}_0} \left[\tilde{g}^T B_\pi \tilde{g} \right]. \tag{56}$$

Using the same argument, we can show that

$$n^{-1} \sum_{t \leq \pi n} \mathrm{E}_{t-1}^{\mathrm{Q}} \left[\tilde{f}_{n,t}^2 \right] \xrightarrow{p} \Upsilon(\pi) \equiv \mathrm{E}^{\mathrm{Q}_0} \left[\tilde{g}^T B_\pi \tilde{g} \right]. \tag{57}$$

The results above and Assumption 3 (i) together lead to a Lindeberg-type condition. Thus, according to the mixing condition implied by the Doeblin condition for the Markov process, we can obtain the following CLT result for martingale difference sequences:

$$\operatorname*{wlim}_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{t \leq \pi n} \tilde{f}_{n,t} = N(0, \Upsilon(\pi)). \tag{58}$$

Proof of Proposition 3 of the Online Appendix. The proof is similar to that of Theorem 7.2 in van der Vaart (1998), except that we allow for non-IID time series and local instability. For brevity, we denote $K_{n,t} \equiv K_{1/\sqrt{n}, g_{n,t}}$. The random variable $W_{n,t} \equiv \frac{\mathrm{d}K_{n,t}}{\mathrm{d}K_0} - 1$ is well defined with probability one. According to (53), it follows that

$$\sum_{t \leq n} W_{n,t} = \frac{1}{\sqrt{n}} \sum_{t \leq n} \tilde{f}_{n,t} + \frac{1}{\sqrt{n}} \sum_{t \leq n} \tilde{\Delta}_{n,t}. \tag{59}$$

where $\tilde{f}_{n,t} \equiv f_{n,t} - \mathrm{E}_{t-1}^{\mathrm{Q}_0} [f_{n,t}]$. Because $\mathrm{E}_{t-1}^{\mathrm{Q}_0} \left[\tilde{\Delta}_{n,t} \right] = 0$ and $\mathrm{E}^{\mathrm{Q}_0} \left[\tilde{\Delta}_{n,t}^2 \right] \to 0$ as $n \to \infty$ for all $t = 1, \cdots, n$, it follows that

$$\mathrm{E}^{\mathrm{Q}_0} \left[\frac{1}{\sqrt{n}} \sum_{t \leq n} \tilde{\Delta}_{n,t} \right] = 0 \quad \text{and} \quad \mathrm{var}^{\mathrm{Q}_0} \left[\frac{1}{\sqrt{n}} \sum_{t \leq n} \tilde{\Delta}_{n,t} \right] \leq \frac{1}{n} \sum_{t \leq n} \mathrm{E}^{\mathrm{Q}_0} \left[\tilde{\Delta}_{n,t}^2 \right] \to 0. \tag{60}$$

Thus, $\frac{1}{\sqrt{n}} \sum_{t \leq n} \tilde{\Delta}_{n,t} = o_p(1)$ under Q_0. And hence, the following approximation holds:

$$\sum_{t \leq n} W_{n,t} = \frac{1}{\sqrt{n}} \sum_{t \leq n} \tilde{f}_{n,t} + o_p(1). \tag{61}$$

By Taylor expansion, we have

$$\ln(1+x) = x - \frac{1}{2} x^2 + x^2 R(x), \tag{62}$$

where $R(x)$ is a continuous function such that $R(x) \to 0$ as $x \to 0$. Therefore, it follows that

$$\ln \prod_{t \leq n} \frac{\mathrm{d}K_{n,t}}{\mathrm{d}K_0} = \sum_{t \leq n} \ln(1 + W_{n,t}) = \sum_{t \leq n} \left[W_{n,t} - \frac{1}{2} W_{n,t}^2 + W_{n,t}^2 R(W_{n,t}) \right] \tag{63}$$

$$= \sum_{t \leq n} W_{n,t} - \frac{1}{2} \sum_{t \leq n} W_{n,t}^2 + \sum_{t \leq n} W_{n,t}^2 R(W_{n,t}). \tag{64}$$

Combining (61) and (64) yields

$$\ln \prod_{t\leq n} \frac{\mathrm{d}K_{n,t}}{\mathrm{d}K_0} = \frac{1}{\sqrt{n}} \sum_{t\leq n} \tilde{f}_{n,t} - \frac{1}{2} \sum_{t\leq n} W_{n,t}^2 + \sum_{t\leq n} W_{n,t}^2 R(W_{n,t}) + o_p(1). \tag{65}$$

We shall first show that

$$\sum_{t\leq n} W_{n,t}^2 = \frac{1}{n} \sum_{t\leq n} \tilde{f}_{n,t}^2 + o_p(1). \tag{66}$$

In fact, by the triangular inequality and the Cauchy-Schwarz inequality, it follows that

$$\left| \sum_{t\leq n} W_{n,t}^2 - \frac{1}{n} \sum_{t\leq n} \tilde{f}_{n,t}^2 \right| \leq \sum_{t\leq n} \left| \frac{1}{\sqrt{n}} \tilde{\Delta}_{n,t} \left(\frac{2}{\sqrt{n}} \tilde{f}_{n,t} + \frac{1}{\sqrt{n}} \tilde{\Delta}_{n,t} \right) \right| \tag{67}$$

$$\leq \left(\frac{1}{n} \sum_{t\leq n} \tilde{\Delta}_{n,t}^2 \right)^{1/2} \left[\frac{1}{n} \sum_{t\leq n} \left(2\tilde{f}_{n,t} + \tilde{\Delta}_{n,t} \right)^2 \right]^{1/2}. \tag{68}$$

Based on (53), it is straightforward to show that $\frac{1}{n} \sum_{t\leq n} \tilde{\Delta}_{n,t}^2 = o_p(1)$. Further, according to Assumption 3 (ii), it follows that $\frac{1}{n} \sum_{t\leq n} \left(2\tilde{f}_{n,t} + \tilde{\Delta}_{n,t} \right)^2 \leq \frac{1}{n} \sum_{t\leq n} 4\tilde{f}_{n,t}^2 + 2\tilde{\Delta}_{n,t}^2 = O_p(1)$. Substituting them into (68) leads to $\sum_{t\leq n} W_{n,t}^2 - \frac{1}{n} \sum_{t\leq n} \tilde{f}_{n,t}^2 = o_p(1)$. Therefore, the equality (65) can be rewritten as

$$\ln \prod_{t\leq n} \frac{\mathrm{d}K_{n,t}}{\mathrm{d}K_0} = \frac{1}{\sqrt{n}} \sum_{t\leq n} \tilde{f}_{n,t} - \frac{1}{2n} \sum_{t\leq n} \tilde{f}_{n,t}^2 + \sum_{t\leq n} W_{n,t}^2 R(W_{n,t}) + o_p(1) \tag{69}$$

$$= \frac{1}{\sqrt{n}} \sum_{t\leq n} \tilde{f}_{n,t} - \frac{1}{2} \int_0^1 \Upsilon(u) \mathrm{d}u + \sum_{t\leq n} W_{n,t}^2 R(W_{n,t}) + o_p(1). \tag{70}$$

Finally, we show that $\sum_{t\leq n} W_{n,t}^2 R(W_{n,t}) = o_p(1)$. Because we have shown that $\sum_{t\leq n} W_{n,t}^2 = O_p(1)$, and

$$\sum_{t\leq n} W_{n,t}^2 |R(W_{n,t})| \leq \max_{1\leq t\leq n} |R(W_{n,t})| \sum_{t\leq n} W_{n,t}^2, \tag{71}$$

it suffices to show that $\max_{1\leq t\leq n} |R(W_{n,t})| = o_p(1)$.

For any $\epsilon > 0$, there exists $\epsilon_R > 0$ such that

$$\mathbb{P}_0 \left(\max_{1\leq t\leq n} |R(W_{n,t})| > \epsilon \right) \leq \sum_{t\leq n} \mathbb{P}_0 \left(|R(W_{n,t})| > \epsilon \right) \leq \sum_{t\leq n} \mathbb{P}_0 \left(W_{n,t}^2 > \epsilon_R \right) \tag{72}$$

$$\leq \sum_{t\leq n} \mathbb{P}_0 \left(\tilde{f}_{n,t}^2 > n\epsilon_R/4 \right) + \sum_{t\leq n} \mathbb{P}_0 \left(\tilde{\Delta}_{n,t}^2 > n\epsilon_R/4 \right). \tag{73}$$

By Markov's inequality, we can further show that

$$\mathbb{P}_0 \left(\max_{1\leq t\leq n} |R(W_{n,t})| > \epsilon \right) \leq \frac{4}{n\epsilon_R} \sum_{t\leq n} \mathrm{E}^{\mathrm{Q}_0} \left[\tilde{f}_{n,t}^2 \mathbf{1}\{\tilde{f}_{n,t}^2 > n\epsilon_R/4\} \right] + \frac{4}{n\epsilon_R} \sum_{t\leq n} \mathrm{E}^{\mathrm{Q}_0} \left[\tilde{\Delta}_{n,t}^2 \right]. \tag{74}$$

According to Assumption 3 (ii), the squared conditional scores $\tilde{f}_{n,t}^2$ are uniformly integrable, and thus

$$\frac{1}{n}\sum_{t\leq n}\mathrm{E}^{\mathrm{Q}_0}\left[\tilde{f}_{n,t}^2\mathbf{1}\{\tilde{f}_{n,t}^2 > n\epsilon_R/4\}\right] \to 0 \text{ as } n\to\infty. \tag{75}$$

Further, according to (53), it holds that

$$\frac{1}{n}\sum_{t\leq n}\mathrm{E}^{\mathrm{Q}_0}\left[\tilde{\Delta}_{n,t}^2\right] \to 0 \text{ as } n\to\infty. \tag{76}$$

Therefore, $\mathbb{P}_0\left(\max_{1\leq t\leq n}|R(W_{n,t})| > \epsilon\right) \to 0$ as $n\to\infty$.

Proof of Proposition 4 of the Online Appendix. We first prove part (i). According to Proposition 2, if defining $\tilde{m}_t(\theta_0) \equiv m_t(\theta_0) - \frac{1}{\sqrt{n}}\lambda(g^T)\begin{bmatrix}1\\ b(t/n)\end{bmatrix}$ for $t = 1,\cdots,n$, we have

$$\mathrm{E}^{\mathrm{Q}_{1/\sqrt{n},f_{n,t}}}\left[\tilde{m}_t(\theta_0)\right] = o\left(\frac{1}{\sqrt{n}}\right), \text{ with } f_{n,t} = g(\mathbf{y}_{t-1},\mathbf{y}_t)^T\begin{bmatrix}1\\ b(t/n)\end{bmatrix}. \tag{77}$$

Further, for $m_t(\theta_0)$ which satisfies Assumption 5, we know that the corresponding $\tilde{m}_t(\theta_0)$ also satisfies Assumption 5. Therefore, according to the functional central limit theorem (invariance principle) of McLeish (1975a) and Phillips and Durlauf (1986), we know that

$$\operatorname*{wlim}_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{t\leq \pi n}\tilde{m}(\theta_0) = W(\pi), \quad \text{for all } \pi\in[0,1]. \tag{78}$$

Thus,

$$\frac{1}{\sqrt{\pi n}}\sum_{t\leq \pi n}m_t(\theta_0) = \frac{1}{\sqrt{\pi n}}\sum_{t\leq \pi n}\tilde{m}_t(\theta_0) + \frac{1}{n}\sum_{t\leq \pi n}\frac{\lambda(g^T)}{\sqrt{\pi}}\begin{bmatrix}1\\ b(t/n)\end{bmatrix}, \tag{79}$$

and hence,

$$\frac{1}{\sqrt{\pi n}}\sum_{t\leq \pi n}m_t(\theta_0) = \frac{W(\pi)}{\sqrt{\pi}} + \frac{\lambda(g^T)}{\sqrt{\pi}}\begin{bmatrix}\pi\\ \int_0^\pi b(u)\mathrm{d}u\end{bmatrix}. \tag{80}$$

Similarly, we can show that

$$\operatorname*{wlim}_{n\to\infty}\frac{1}{\sqrt{(1-\pi)n}}\sum_{t>\pi n}m_t(\theta_0) = \frac{W(1)-W(\pi)}{\sqrt{1-\pi}} + \frac{\lambda(g^T)}{\sqrt{1-\pi}}\begin{bmatrix}1-\pi\\ \int_\pi^1 b(u)\mathrm{d}u\end{bmatrix}. \tag{81}$$

Now, we prove part (ii). Because $g_1, g_2 \in \mathcal{T}(\mathrm{Q}_0)$, by the definition of $\theta_{n,t}$, we know that

$$0 = \int m_t(\theta_{n,t})\mathrm{d}\mathrm{Q}_{1/\sqrt{n},f_{n,t}}, \text{ for all } t,n. \tag{82}$$

Using the Taylor expansion, we obtain

$$0 = \int\left[m_t(\theta_0) + \nabla_\theta m_t(\dot{\theta}_{n,t})(\theta_{n,t}-\theta_0)\right]\left[1 + f_{n,t}/\sqrt{n} + \Delta_{n,t}/\sqrt{n}\right]\mathrm{d}\mathrm{Q}_0, \text{ for all } t,n, \tag{83}$$

where $\dot{\theta}_{n,t}$ lies between θ_0 and $\theta_{n,t}$ for all t and n. Suppose $\theta_{n,t}$ converges θ_0 at the rate of $\sqrt{n}$ (as we verify later). According to Assumption 5, it follows that

$$0 = \frac{1}{\sqrt{n}}\lambda(g^T)\begin{bmatrix} 1 \\ b(t/n) \end{bmatrix} + D(\theta_{n,t} - \theta_0) + o\left(\frac{1}{\sqrt{n}}\right), \text{ for all } t, n. \tag{84}$$

Therefore, the parameter sequence $\theta_{n,t}$ can be specified as

$$\theta_{n,t} - \theta_0 = -(D^T D)^{-1} D^T \frac{1}{\sqrt{n}}\lambda(g^T)\begin{bmatrix} 1 \\ b(t/n) \end{bmatrix} + o\left(\frac{1}{\sqrt{n}}\right), \text{ for all } t, n. \tag{85}$$

Hence, using the Taylor expansion again leads to

$$\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_{n,t}) = \frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_0) - \frac{1}{n}\sum_{t\le\pi n}\nabla_\theta m_t(\dot{\theta}_{n,t})(D^T D)^{-1} D^T \frac{\lambda(g^T)}{\sqrt{\pi}}\begin{bmatrix} 1 \\ b(t/n) \end{bmatrix} + o\left(1\right). \tag{86}$$

Due to Assumption 5, according to Lemma 4 of Li and Müller (2009), it follows that

$$\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_{n,t}) = \frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_0) - D(D^T D)^{-1} D^T \frac{\lambda(g^T)}{\sqrt{\pi}}\begin{bmatrix} \pi \\ \int_0^\pi b(u)\mathrm{d}u \end{bmatrix} + o\left(1\right). \tag{87}$$

Because $g_1, g_2 \in \mathcal{T}(\mathrm{Q}_0)$, it holds that $\lambda(g_1), \lambda(g_2) \in \mathsf{lin}(D)$, and thus

$$\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_{n,t}) = \frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_0) - \frac{\lambda(g^T)}{\sqrt{\pi}}\begin{bmatrix} \pi \\ \int_0^\pi b(u)\mathrm{d}u \end{bmatrix} + o\left(1\right). \tag{88}$$

Similarly, we can show that

$$\frac{1}{\sqrt{(1-\pi)n}}\sum_{t>\pi n} m_t(\theta_{n,t}) = \frac{1}{\sqrt{(1-\pi)n}}\sum_{t>\pi n} m_t(\theta_0) - \frac{\lambda(g^T)}{\sqrt{1-\pi}}\begin{bmatrix} 1-\pi \\ \int_\pi^1 b(u)\mathrm{d}u \end{bmatrix} + o\left(1\right). \tag{89}$$

Finally, we prove parts (iii) and (iv). Using the Taylor expansion, we obtain the following approximation:

$$\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\hat{\theta}_{\mathrm{e,n}}) = \frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_0) + \frac{1}{\pi n}\sum_{t\le\pi n}\nabla_\theta m_t(\dot{\theta}_{\mathrm{e,n}})\left[\sqrt{\pi n}(\hat{\theta}_{\mathrm{e,n}} - \theta_0)\right] + o\left(1\right), \tag{90}$$

where $\dot{\theta}_{\mathrm{e,n}}$ lies between $\hat{\theta}_{\mathrm{e,n}}$ and θ_0. According to Proposition 6 (ii),

$$\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\hat{\theta}_{\mathrm{e,n}}) = \frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_0) - D(D^T D)^{-1} D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_0)\right] + o_p\left(1\right) \tag{91}$$

Further rearranging the terms on the right-hand side of (91) leads to

$$\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\hat{\theta}_{\mathrm{e,n}}) = \left[I - D(D^T D)^{-1} D^T\right]\left[\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_0)\right] + o_p\left(1\right). \tag{92}$$

Similarly,

$$\frac{1}{\sqrt{(1-\pi)n}}\sum_{t>\pi n} m_t(\hat{\theta}_{\mathrm{e,n}}) = \frac{1}{\sqrt{(1-\pi)n}}\sum_{t>\pi n} m_t(\theta_0) - D(D^TD)^{-1}D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\leq \pi n} m_t(\theta_0)\right] + o_p(1)\,. \tag{93}$$

Part (iv) can be proved using analogous steps, which we do not repeat.

Proof of Proposition 5 of the Online Appendix. Similar to the results in Severini and Tripathi (2013) and Chen and Santos (2018), the tangent set $\mathcal{T}(\mathrm{Q}_0)$ can be characterized as follows:

$$\mathcal{T}(\mathrm{Q}_0) = \left\{f \in L_0^2(\mathrm{Q}_0):\ \mathrm{E}^{\mathrm{Q}_0}\left[m(\cdot,\theta_0)f\right] \in \mathsf{lin}(D)\right\}, \tag{94}$$

where $\mathsf{lin}(D)$ is the linear space spanned by the column vectors of D. Therefore, it suffices to show that

$$\mathrm{E}^{\mathrm{Q}_0}\left[m(\cdot,\theta_0)f\right] = \mathrm{E}^{\mathrm{Q}_0}\left[h(\cdot,\theta_0)f\right] \text{ for all } f \in L_0^2(\mathrm{Q}_0). \tag{95}$$

Under the assumption, the following identity holds:

$$\mathrm{E}^{\mathrm{Q}_0}\left[h(\mathbf{y},\mathbf{y}',\theta_0)f(\mathbf{y},\mathbf{y}')\right] = \mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y},\mathbf{y}',\theta_0)f(\mathbf{y},\mathbf{y}')\right] - \sum_{k=1}^{\infty} A_k, \tag{96}$$

where

$$A_k = \mathrm{E}^{\mathrm{Q}_0}\left\{\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_{k-1},\mathbf{y}_k,\theta_0)|\mathbf{y}_0=\mathbf{y}\right]f(\mathbf{y},\mathbf{y}')\right\} - \mathrm{E}^{\mathrm{Q}_0}\left\{\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_k,\mathbf{y}_{k+1},\theta_0)|\mathbf{y}_1=\mathbf{y}'\right]f(\mathbf{y},\mathbf{y}')\right\}. \tag{97}$$

Further, for each $k \geq 1$, the Markov property implies that

$$\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_k,\mathbf{y}_{k+1},\theta_0)|\mathbf{y}_1=\mathbf{y}'\right]f(\mathbf{y},\mathbf{y}') = \mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_{k-1},\mathbf{y}_k,\theta_0)|\mathbf{y}_0=\mathbf{y}'\right]f(\mathbf{y},\mathbf{y}'). \tag{98}$$

Thus, the equation (97) can be rewritten as

$$A_k = \mathrm{E}^{\mathrm{Q}_0}\left\{\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_{k-1},\mathbf{y}_k,\theta_0)|\mathbf{y}_0=\mathbf{y}\right]f(\mathbf{y},\mathbf{y}')\right\} - \mathrm{E}^{\mathrm{Q}_0}\left\{\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_{k-1},\mathbf{y}_k,\theta_0)|\mathbf{y}_0=\mathbf{y}'\right]f(\mathbf{y},\mathbf{y}')\right\}. \tag{99}$$

It suffices to show that $A_k = 0$ for all k. In fact, the following equalities hold:

$$\begin{aligned}
&\mathrm{E}^{\mathrm{Q}_0}\left\{\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_{k-1},\mathbf{y}_k,\theta_0)|\mathbf{y}_0=\mathbf{y}'\right]f(\mathbf{y},\mathbf{y}')\right\}\\
&\quad= \mathrm{E}^{\mathrm{Q}_0}\left\{\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_{k-1},\mathbf{y}_k,\theta_0)|\mathbf{y}_0=\mathbf{y}'\right]\mathrm{E}^{\mathrm{Q}_0}\left[f(\mathbf{y},\mathbf{y}')|\mathbf{y}'\right]\right\} \text{ (Law of Iterated Projections)}\\
&\quad= \mathrm{E}^{\mathrm{Q}_0}\left\{\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_{k-1},\mathbf{y}_k,\theta_0)|\mathbf{y}_0=\mathbf{y}'\right]\mathrm{E}^{\mathrm{Q}_0}\left[f(\mathbf{y}',\mathbf{y})|\mathbf{y}'\right]\right\} \text{ (Proposition 1)}\\
&\quad= \mathrm{E}^{\mathrm{Q}_0}\left\{\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_{k-1},\mathbf{y}_k,\theta_0)|\mathbf{y}_0=\mathbf{y}'\right]f(\mathbf{y}',\mathbf{y})\right\} \text{ (Law of Iterated Projections)}
\end{aligned}$$

Therefore, $A_k = 0$ for all $k \geq 1$, and hence from (96), it follows that

$$\mathrm{E}^{\mathrm{Q}_0}\left[h(\mathbf{y},\mathbf{y}',\theta_0)f(\mathbf{y},\mathbf{y}')\right] = \mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y},\mathbf{y}',\theta_0)f(\mathbf{y},\mathbf{y}')\right]. \tag{100}$$

According to Greenwood and Wefelmeyer (1995), we know that

$$\mathrm{E}^{\mathrm{Q}_0}\left[h(\mathbf{y}_0,\mathbf{y}_1,\theta_0)h(\mathbf{y}_0,\mathbf{y}_1,\theta_0)^T\right] = \sum_{\tau=-\infty}^{\infty}\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_0,\mathbf{y}_1,\theta_0)m(\mathbf{y}_\tau,\mathbf{y}_{\tau+1}\theta_0)^T\right] = I. \tag{101}$$

By Markov's property and the law of iterated projections, for all $k \geq 0$,

$$\mathrm{E}^{\mathrm{Q}_0}\left\{\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_k,\mathbf{y}_{k+1},\theta_0)|\mathbf{y}_1\right]|\mathbf{y}_0\right\} = \mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_k,\mathbf{y}_{k+1},\theta_0)|\mathbf{y}_0\right]. \tag{102}$$

Therefore, $\mathrm{E}^{\mathrm{Q}_0}\left[h(\mathbf{y},\mathbf{y}',\theta_0)|\mathbf{y}\right] = 0$.

Proof of Proposition 6 of the Online Appendix. The proof follows the standard GMM approximations in Hansen (1982), Hansen (2007), and Hansen (2012).

Proof of Proposition 7 of the Online Appendix. The cases of ψ_s with $s \in \{e, o\}$ follow the same derivations, and so we only show the case $s = e$. We first prove part (i). Given the parameter value $\theta^{(1)}_{\mathrm{e,n}}$, the constrained efficient GMM estimator $(\theta^{(1)}_{\mathrm{e,n}}, \psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}}))^T$ for the full model satisfies the first-order condition

$$\nabla J(\theta^{(1)}_{\mathrm{e,n}}, \psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}}); \mathbf{y}^n_{\mathrm{e}}) = \Gamma^T_{\theta,1}\Lambda_{\mathrm{e,n}}, \quad \text{with } \Gamma_{\theta,1} = [I, 0_{d_{\theta,1}\times d_{\theta,2}}], \tag{103}$$

and $\Lambda_{\mathrm{e,n}}$ is a $d_{\theta,1}\times 1$ vector of Lagrangian multipliers for the constraints $\Gamma_{\theta,1}\theta = \theta^{(1)}_n$ in search of the constrained GMM estimator $(\theta^{(1)}_{\mathrm{e,n}}, \psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}}))^T$. The Taylor expansion of $\nabla J(\theta^{(1)}_{\mathrm{e,n}}, \psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}}); \mathbf{y}^n_{\mathrm{e}})$ around θ_0 leads to

$$\frac{1}{\sqrt{\pi n}}\Gamma^T_{\theta,1}\Lambda_{\mathrm{e,n}} = 2D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\leq \pi n} m_t(\theta_0)\right] + 2\mathbf{I}_{\mathrm{Q}}\sqrt{\pi n}\begin{bmatrix}\theta^{(1)}_{\mathrm{e,n}} - \theta^{(1)}_0 \\ \psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}}) - \theta^{(2)}_0\end{bmatrix} + o_p(1). \tag{104}$$

We first multiply both sides of (104) by $\Gamma_{\theta,1}\mathbf{I}^{-1}_{\mathrm{Q}}$, and then by $\left(\Gamma_{\theta,1}\mathbf{I}^{-1}_{\mathrm{Q}}\Gamma^T_{\theta,1}\right)^{-1}$. The optimal Lagrangian multipliers can be represented as

$$\begin{aligned}\frac{1}{\sqrt{\pi n}}\Lambda_{\mathrm{e,n}} = {} & 2\left(\Gamma_{\theta,1}\mathbf{I}^{-1}_{\mathrm{Q}}\Gamma^T_{\theta,1}\right)^{-1}\Gamma_{\theta,1}\mathbf{I}^{-1}_{\mathrm{Q}}D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\leq \pi n} m_t(\theta_0)\right] \\ & + 2\left(\Gamma_{\theta,1}\mathbf{I}^{-1}_{\mathrm{Q}}\Gamma^T_{\theta,1}\right)^{-1}\sqrt{\pi n}(\theta^{(1)}_{\mathrm{e,n}} - \theta^{(1)}_0) + o_p(1).\end{aligned} \tag{105}$$

Substituting (104) and (105) into (103) yields

$$\begin{aligned}\frac{1}{\sqrt{\pi n}}\nabla J(\theta^{(1)}_{\mathrm{e,n}}, \psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}}); \mathbf{y}^n_{\mathrm{e}}) = {} & 2\Gamma^T_{\theta,1}\left(\Gamma_{\theta,1}\mathbf{I}^{-1}_{\mathrm{Q}}\Gamma^T_{\theta,1}\right)^{-1}\Gamma_{\theta,1}\mathbf{I}^{-1}_{\mathrm{Q}}D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\leq \pi n} m_t(\theta_0)\right] \\ & + 2\Gamma^T_{\theta,1}(\Gamma_{\theta,1}\mathbf{I}^{-1}_{\mathrm{Q}}\Gamma^T_{\theta,1})^{-1}\sqrt{\pi n}(\theta^{(1)}_{\mathrm{e,n}} - \theta^{(1)}_0) + o_p(1).\end{aligned} \tag{106}$$

According to Proposition 1, we substitute (84) into (106) and obtain

$$\frac{1}{\sqrt{\pi n}}\nabla J(\theta^{(1)}_{\mathrm{e,n}}, \psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}}); \mathbf{y}^n_{\mathrm{e}}) = 2\Gamma^T_{\theta,1}\mathbf{I}_{\mathrm{F}}\Gamma_{\theta,1}\mathbf{I}^{-1}_{\Omega}D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\leq\pi n} m_t(\theta_0)\right] \\ - 2\Gamma^T_{\theta,1}\mathbf{I}_{\mathrm{F}}\mathbf{I}^{-1}_{\mathrm{B}}D^T_{11}\left[\sqrt{\pi}\lambda^{(1)}(g_1) + \lambda^{(1)}(g_2)\int_0^{\pi} b(u)\mathrm{d}u/\sqrt{\pi}\right] + o_p(1). \tag{107}$$

Based on (7), we have

$$\frac{1}{\sqrt{\pi n}}\nabla J(\theta^{(1)}_{\mathrm{e,n}}, \psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}}); \mathbf{y}^n_{\mathrm{e}}) = 2\Gamma^T_{\theta,1}\mathbf{I}_{\mathrm{F}}\left[\Gamma_{\theta,1}\mathbf{I}^{-1}_{\Omega}D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\leq\pi n} m_t(\theta_0)\right] - \mathbf{I}^{-1}_{\mathrm{B}}D^T_{11}\Gamma_{m,1}\nu(g,b,\pi)\right] + o_p(1). \tag{108}$$

Given the baseline efficient GMM estimator $\tilde{\theta}^{(1)}_{\mathrm{e,n}}$ based on the estimation sample, the constrained GMM estimator $(\tilde{\theta}^{(1)}_{\mathrm{e,n}}, \psi_{\mathrm{e}}(\tilde{\theta}^{(1)}_{\mathrm{e,n}}))^T$ for the full model satisfies the first-order condition

$$\nabla J(\tilde{\theta}^{(1)}_{\mathrm{e,n}}, \psi_{\mathrm{e}}(\tilde{\theta}^{(1)}_{\mathrm{e,n}}); \mathbf{y}^n_{\mathrm{e}}) = \Gamma^T_{\theta,1}\Lambda^{(1)}_{\mathrm{e,n}}, \quad \text{with } \Gamma_{\theta,1} = [I, 0_{d_{\theta,1}\times d_{\theta,2}}], \tag{109}$$

and $\Lambda^{(1)}_{\mathrm{e,n}}$ is a $d_{\theta,1}\times 1$ vector of Lagrangian multipliers for the constraints $\Gamma_{\theta,1}\theta = \tilde{\theta}^{(1)}_{\mathrm{e,n}}$ in search of the constrained GMM estimator $(\tilde{\theta}^{(1)}_{\mathrm{e,n}}, \psi_{\mathrm{e}}(\tilde{\theta}^{(1)}_{\mathrm{e,n}}))^T$. The Taylor expansion of $\nabla J(\tilde{\theta}^{(1)}_{\mathrm{e,n}}, \psi_{\mathrm{e}}(\tilde{\theta}^{(1)}_{\mathrm{e,n}}); \mathbf{y}^n_{\mathrm{e}})$ around $(\theta^{(1)}_{\mathrm{e,n}}, \psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}}))^T$, together with (109), leads to

$$\frac{1}{\sqrt{\pi n}}\Gamma^T_{\theta,1}\Lambda^{(1)}_{\mathrm{e,n}} = \frac{1}{\sqrt{\pi n}}\nabla J(\theta^{(1)}_{\mathrm{e,n}}, \psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}}); \mathbf{y}^n_{\mathrm{e}}) + 2\mathbf{I}_{\Omega}\sqrt{\pi n}\begin{bmatrix} \tilde{\theta}^{(1)}_{\mathrm{e,n}} - \theta^{(1)}_{\mathrm{e,n}} \\ \psi_{\mathrm{e}}(\tilde{\theta}^{(1)}_{\mathrm{e,n}}) - \psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}}) \end{bmatrix} + o_p(1). \tag{110}$$

We first multiply both sides of (110) by $\Gamma_{\theta,1}\mathbf{I}^{-1}_{\Omega}$, and then by $\left(\Gamma_{\theta,1}\mathbf{I}_{\Omega}\Gamma^T_{\theta,1}\right)^{-1}$. The optimal Lagrangian multipliers can be represented as

$$\frac{1}{\sqrt{\pi n}}\Lambda^{(1)}_{\mathrm{e,n}} = \mathbf{I}_{\mathrm{F}}\Gamma_{\theta,1}\mathbf{I}^{-1}_{\Omega}\frac{1}{\sqrt{\pi n}}\nabla J(\theta^{(1)}_{\mathrm{e,n}}, \psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}}); \mathbf{y}^n_{\mathrm{e}}) + 2\mathbf{I}_{\mathrm{F}}\sqrt{\pi n}(\tilde{\theta}^{(1)}_{\mathrm{e,n}} - \theta^{(1)}_{\mathrm{e,n}}) + o_p(1). \tag{111}$$

Further substituting (106) into equation (111) above yields

$$\frac{1}{\sqrt{\pi n}}\Lambda^{(1)}_{\mathrm{e,n}} = 2\mathbf{I}_{\mathrm{F}}\Gamma_{\theta,1}\mathbf{I}^{-1}_{\Omega}D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\leq\pi n} m_t(\theta_0)\right] + 2\mathbf{I}_{\mathrm{F}}\sqrt{\pi n}(\tilde{\theta}^{(1)}_{\mathrm{e,n}} - \theta^{(1)}_0) + o_p(1). \tag{112}$$

Based on Proposition 6 of the Online Appendix, we obtain

$$\sqrt{\pi n}(\tilde{\theta}^{(1)}_{\mathrm{e,n}} - \theta^{(1)}_0) = -\mathbf{I}^{-1}_{\mathrm{B}}D^T_{11}\left[\frac{1}{\sqrt{\pi n}}\sum_{t\leq\pi n} m^{(1)}_t(\theta^{(1)}_0)\right] + o_p(1). \tag{113}$$

Substituting (113) into (112) gives the following asymptotic representation of $\frac{1}{\sqrt{\pi n}}\Lambda_{\text{e,n}}^{(1)}$:

$$\frac{1}{\sqrt{\pi n}}\Lambda_{\text{e,n}}^{(1)} = 2\mathbf{I}_{\text{F}}\Gamma_{\theta,1}\mathbf{I}_{\text{Q}}^{-1}D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\leq\pi n} m_t(\theta_0)\right] - 2\mathbf{I}_{\text{F}}\mathbf{I}_{\text{B}}^{-1}D_{11}^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\leq\pi n} m_t^{(1)}(\theta_0^{(1)})\right] + o_p(1). \tag{114}$$

We substitute (106) and (114) into (110) and multiply the both sides by $\mathbf{I}_{\text{Q}}^{-1}/2$. The estimator can be represented by

$$\begin{aligned}\sqrt{\pi n}\begin{bmatrix}\tilde{\theta}_{\text{e,n}}^{(1)} - \theta_{\text{e,n}}^{(1)} \\ \psi_{\text{e}}(\tilde{\theta}_{\text{e,n}}^{(1)}) - \psi_{\text{e}}(\theta_{\text{e,n}}^{(1)})\end{bmatrix} &= -\mathbf{I}_{\text{Q}}^{-1}\Gamma_{\theta,1}^T\mathbf{I}_{\text{F}}\mathbf{I}_{\text{B}}^{-1}D_{11}^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\leq\pi n} m_t^{(1)}(\theta_0^{(1)}) - \nu_e^{(1)}(g,b,\pi)\right] + o_p(1) \qquad (115)\\ &= -\mathbf{I}_{\text{Q}}^{-1}\Gamma_{\theta,1}^T\mathbf{I}_{\text{F}}\mathbf{I}_{\text{B}}^{-1}D_{11}^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\leq\pi n} m_t^{(1)}(\theta_{n,t}^{(1)})\right] + o_p(1).\end{aligned}$$

Now we prove part (ii). The estimators $\psi_{\text{e}}(\theta_{\text{e,n}}^{(1)})$ and $\hat{\theta}_{\text{e,n}}^{(2)} = \psi_{\text{e}}(\hat{\theta}_{\text{e,n}}^{(1)})$ are the constrained efficient GMM estimators for the nuisance parameter $\theta^{(2)}$ when controlling for $\Gamma_{\theta,1}\theta = \theta_{\text{e,n}}^{(1)}$ and $\Gamma_{\theta,1}\theta = \hat{\theta}_{\text{e,n}}^{(1)}$, respectively. Due to the first order condition $\nabla J(\hat{\theta}_n^{(1)}, \psi_{\text{e}}(\hat{\theta}_n^{(1)}); \mathbf{y}_{\text{e}}^n) = 0$, the Taylor expansion of $\nabla J(\hat{\theta}_n^{(1)}, \psi_{\text{e}}(\hat{\theta}_n^{(1)}); \mathbf{y}_{\text{e}}^n)$ around $(\theta_{\text{e,n}}^{(1)}, \psi_{\text{e}}(\theta_{\text{e,n}}^{(1)}))^T$ leads to

$$0 = \nabla J(\theta_{\text{e,n}}^{(1)}, \psi_{\text{e}}(\theta_{\text{e,n}}^{(1)}); \mathbf{y}_{\text{e}}^n) + 2\mathbf{I}_{\text{Q}}\sqrt{\pi n}\begin{bmatrix}\hat{\theta}_{\text{e,n}}^{(1)} - \theta_{\text{e,n}}^{(1)} \\ \psi_{\text{e}}(\hat{\theta}_{\text{e,n}}^{(1)}) - \psi_{\text{e}}(\theta_{\text{e,n}}^{(1)})\end{bmatrix} + o_p(1). \tag{116}$$

Substituting (106) into (116) and multiplying the both sides by $\mathbf{I}_{\text{Q}}^{-1}/2$, we have

$$\sqrt{\pi n}\begin{bmatrix}\hat{\theta}_{\text{e,n}}^{(1)} - \theta_{\text{e,n}}^{(1)} \\ \psi_{\text{e}}(\hat{\theta}_{\text{e,n}}^{(1)}) - \psi_{\text{e}}(\theta_{\text{e,n}}^{(1)})\end{bmatrix} = -\mathbf{I}_{\text{Q}}^{-1}\Gamma_{\theta,1}^T\mathbf{I}_{\text{F}}\left\{\Gamma_{\theta,1}\mathbf{I}_{\text{Q}}^{-1}D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\leq\pi n} m_t(\theta_0)\right] - \mathbf{I}_{\text{B}}^{-1}D_{11}^T\Gamma_{m,1}\nu_e\right\} + o_p(1). \tag{117}$$

Proof of Proposition 8 of the Online Appendix. We first provide an approximation for $\mathcal{L}(\tilde{\theta}_{\text{e,n}}^{(1)}; \mathbf{y}_{\text{e}}^n)$. According to the second-order Taylor expansion around $(\theta_{\text{e,n}}^{(1)}, \psi_{\text{e}}(\theta_{\text{e,n}}^{(1)}))$, it follows that

$$\begin{aligned}\mathcal{L}(\tilde{\theta}_{\text{e,n}}^{(1)}; \mathbf{y}_{\text{e}}^n) &= \left[\frac{1}{\sqrt{\pi n}}\nabla J(\theta_{\text{e,n}}^{(1)}, \psi_{\text{e}}(\theta_{\text{e,n}}^{(1)}); \mathbf{y}_{\text{e}}^n)\right]^T\sqrt{\pi n}\begin{bmatrix}\tilde{\theta}_{\text{e,n}}^{(1)} - \theta_{\text{e,n}}^{(1)} \\ \psi_{\text{e}}(\tilde{\theta}_{\text{e,n}}^{(1)}) - \psi_{\text{e}}(\theta_{\text{e,n}}^{(1)})\end{bmatrix} \qquad (118)\\ &\quad + \sqrt{\pi n}\begin{bmatrix}\tilde{\theta}_{\text{e,n}}^{(1)} - \theta_{\text{e,n}}^{(1)} \\ \psi_{\text{e}}(\tilde{\theta}_{\text{e,n}}^{(1)}) - \psi_{\text{e}}(\theta_{\text{e,n}}^{(1)})\end{bmatrix}^T\mathbf{I}_{\text{Q}}\sqrt{\pi n}\begin{bmatrix}\tilde{\theta}_{\text{e,n}}^{(1)} - \theta_{\text{e,n}}^{(1)} \\ \psi_{\text{e}}(\tilde{\theta}_{\text{e,n}}^{(1)}) - \psi_{\text{e}}(\theta_{\text{e,n}}^{(1)})\end{bmatrix} + o_p(1).\end{aligned}$$

Thus, following (108) and (115),

$$\mathcal{L}(\tilde{\theta}_{\text{e,n}}^{(1)}; \mathbf{y}_{\text{e}}^n) = -2\left[L_{\text{F}}\zeta_{\text{e,n}} + L_{\Delta}\nu_e\right]^T\mathbf{I}_{\text{F}}L_{\text{B}}\zeta_{\text{e,n}} + \zeta_{\text{e,n}}^T L_{\text{B}}^T\mathbf{I}_{\text{F}}L_{\text{B}}\zeta_{\text{e,n}} + o_p(1). \tag{119}$$

We now provide an approximation for $\mathcal{L}(\tilde{\theta}_{\text{e,n}}^{(1)}; \mathbf{y}_{\text{o}}^n)$. According to the second-order Taylor expansion around

$(\theta^{(1)}_{\text{o,n}}, \psi_{\text{o}}(\theta^{(1)}_{\text{o,n}}))$, it follows that

$$\begin{aligned}\mathcal{L}(\tilde{\theta}^{(1)}_{\text{e,n}}; \mathbf{y}^n_{\text{o}}) &= \left[\frac{1}{\sqrt{\pi n}} \nabla J(\theta^{(1)}_{\text{o,n}}, \psi_{\text{o}}(\theta^{(1)}_{\text{o,n}}); \mathbf{y}^n_{\text{o}})\right]^T \sqrt{\pi n} \begin{bmatrix} \tilde{\theta}^{(1)}_{\text{e,n}} - \theta^{(1)}_{\text{o,n}} \\ \psi_{\text{o}}(\tilde{\theta}^{(1)}_{\text{e,n}}) - \psi_{\text{o}}(\theta^{(1)}_{\text{o,n}}) \end{bmatrix} \\ &\quad + \sqrt{\pi n} \begin{bmatrix} \tilde{\theta}^{(1)}_{\text{e,n}} - \theta^{(1)}_{\text{o,n}} \\ \psi_{\text{o}}(\tilde{\theta}^{(1)}_{\text{e,n}}) - \psi_{\text{o}}(\theta^{(1)}_{\text{o,n}}) \end{bmatrix}^T \mathbf{I}_{\mathcal{Q}} \sqrt{\pi n} \begin{bmatrix} \tilde{\theta}^{(1)}_{\text{e,n}} - \theta^{(1)}_{\text{o,n}} \\ \psi_{\text{o}}(\tilde{\theta}^{(1)}_{\text{e,n}}) - \psi_{\text{o}}(\theta^{(1)}_{\text{o,n}}) \end{bmatrix} + o_p(1).\end{aligned} \tag{120}$$

Thus, similar to the derivation of (119), we can show that

$$\mathcal{L}(\tilde{\theta}^{(1)}_{\text{e,n}}; \mathbf{y}^n_{\text{o}}) = -2 \left[L_{\text{F}} \zeta_{\text{o,n}} + L_{\Delta} \nu_o\right]^T \mathbf{I}_{\text{F}} L_{\text{B}} \zeta_{\text{e,n}} + \zeta^T_{\text{e,n}} L^T_{\text{B}} \mathbf{I}_{\text{F}} L_{\text{B}} \zeta_{\text{e,n}} + o_p(1). \tag{121}$$

We now provide an approximation for $\mathcal{L}(\hat{\theta}^{(1)}_{\text{e,n}}; \mathbf{y}^n_{\text{e}})$. According to the second-order Taylor expansion around $(\theta^{(1)}_{\text{e,n}}, \psi_{\text{o}}(\theta^{(1)}_{\text{e,n}}))$, it follows that

$$\begin{aligned}\mathcal{L}(\hat{\theta}^{(1)}_{\text{e,n}}; \mathbf{y}^n_{\text{e}}) &= \left[\frac{1}{\sqrt{\pi n}} \nabla J(\theta^{(1)}_{\text{e,n}}, \psi_{\text{e}}(\theta^{(1)}_{\text{e,n}}); \mathbf{y}^n_{\text{e}})\right]^T \sqrt{\pi n} \begin{bmatrix} \hat{\theta}^{(1)}_{\text{e,n}} - \theta^{(1)}_{\text{e,n}} \\ \psi_{\text{e}}(\hat{\theta}^{(1)}_{\text{e,n}}) - \psi_{\text{e}}(\theta^{(1)}_{\text{e,n}}) \end{bmatrix} \\ &\quad + \sqrt{\pi n} \begin{bmatrix} \hat{\theta}^{(1)}_{\text{e,n}} - \theta^{(1)}_{\text{e,n}} \\ \psi_{\text{e}}(\hat{\theta}^{(1)}_{\text{e,n}}) - \psi_{\text{e}}(\theta^{(1)}_{\text{e,n}}) \end{bmatrix}^T \mathbf{I}_{\mathcal{Q}} \sqrt{\pi n} \begin{bmatrix} \hat{\theta}^{(1)}_{\text{e,n}} - \theta^{(1)}_{\text{e,n}} \\ \psi_{\text{e}}(\hat{\theta}^{(1)}_{\text{e,n}}) - \psi_{\text{e}}(\theta^{(1)}_{\text{e,n}}) \end{bmatrix} + o_p(1).\end{aligned} \tag{122}$$

Thus, similar to the derivation of (119), we can show that

$$\mathcal{L}(\hat{\theta}^{(1)}_{\text{e,n}}; \mathbf{y}^n_{\text{e}}) = -\left[L_{\text{F}} \zeta_{\text{e,n}} + L_{\Delta} \nu_e\right]^T \mathbf{I}_{\text{F}} \left[L_{\text{F}} \zeta_{\text{e,n}} + L_{\Delta} \nu_e\right] + o_p(1). \tag{123}$$

We now provide an approximation for $\mathcal{L}(\hat{\theta}^{(1)}_{\text{e,n}}; \mathbf{y}^n_{\text{o}})$. According to the second-order Taylor expansion around $(\theta^{(1)}_{\text{o,n}}, \psi_{\text{o}}(\theta^{(1)}_{\text{o,n}}))$, it follows that

$$\begin{aligned}\mathcal{L}(\hat{\theta}^{(1)}_{\text{e,n}}; \mathbf{y}^n_{\text{o}}) &= \left[\frac{1}{\sqrt{\pi n}} \nabla J(\theta^{(1)}_{\text{o,n}}, \psi_{\text{o}}(\theta^{(1)}_{\text{o,n}}); \mathbf{y}^n_{\text{o}})\right]^T \sqrt{\pi n} \begin{bmatrix} \hat{\theta}^{(1)}_{\text{e,n}} - \theta^{(1)}_{\text{o,n}} \\ \psi_{\text{o}}(\hat{\theta}^{(1)}_{\text{e,n}}) - \psi_{\text{o}}(\theta^{(1)}_{\text{o,n}}) \end{bmatrix} \\ &\quad + \sqrt{\pi n} \begin{bmatrix} \hat{\theta}^{(1)}_{\text{e,n}} - \theta^{(1)}_{\text{o,n}} \\ \psi_{\text{o}}(\hat{\theta}^{(1)}_{\text{e,n}}) - \psi_{\text{o}}(\theta^{(1)}_{\text{o,n}}) \end{bmatrix}^T \mathbf{I}_{\mathcal{Q}} \sqrt{\pi n} \begin{bmatrix} \hat{\theta}^{(1)}_{\text{e,n}} - \theta^{(1)}_{\text{o,n}} \\ \psi_{\text{o}}(\hat{\theta}^{(1)}_{\text{e,n}}) - \psi_{\text{o}}(\theta^{(1)}_{\text{o,n}}) \end{bmatrix} + o_p(1).\end{aligned} \tag{124}$$

Thus, similar to the derivation of (119), we can show that

$$\mathcal{L}(\tilde{\theta}^{(1)}_{\text{e,n}}; \mathbf{y}^n_{\text{o}}) = -2 \left[L_{\text{F}} \zeta_{\text{o,n}} + L_{\Delta} \nu_o\right]^T \mathbf{I}_{\text{F}} \left[L_{\text{F}} \zeta_{\text{e,n}} + L_{\Delta} \nu_e\right] + \left[L_{\text{F}} \zeta_{\text{e,n}} + L_{\Delta} \nu_e\right]^T \mathbf{I}_{\text{F}} \left[L_{\text{F}} \zeta_{\text{e,n}} + L_{\Delta} \nu_e\right] + o_p(1). \tag{125}$$

4.3 Proofs of Corollaries

Proof of Corollary 1 of Chen, Dou, and Kogan (2021). We can derive the result following the same derivations for (85) under the baseline GMM model $\mathcal{Q}^{(1)}$.

Proof of Corollary 1 of the Online Appendix. The proof is similar to that of Lemma 1 of Li and Müller (2009), which is based on Le Cam's first lemma (e.g., van der Vaart, 1998, Page 88).

5 Miscellaneous Proofs and Derivations

5.1 Asymptotic Covariance of Moments

The derivations here are for the simulated studies in the example of time-varying disaster risk models. Let $g_t = x_t\left[x_{t+2} - \rho x_{t+1} - (1-\rho)\overline{p}\right]$. Then, it is obvious that $\mathrm{E}_t\left[g_t\right] = 0$ and $\mathrm{E}\left[g_{t+k}g_t\right] = 0$ when $|k| \geq 2$.

Expectation $\mathrm{E}\left[g_{t+1}g_t\right]$ can be computed as follows:

$$\begin{aligned}
\mathrm{E}\left[g_{t+1}g_t\right] &= -\rho\mathrm{E}\left[p_{t-1}p_tp_{t+1}(1-p_{t+1})\right] \\
&= \left[\rho(1-\rho)^2\overline{p}^2 - \rho(1-\rho)\overline{p}\right]\mathrm{E}\left[p_{t-1}p_t\right] \\
&\quad + \left[2\rho^2(1-\rho)\overline{p} - \rho^2 + \rho\sigma_p^2\right]\mathrm{E}\left[p_{t-1}p_t^2\right] + \rho^3\mathrm{E}\left[p_{t-1}p_t^3\right].
\end{aligned}$$

Expectations $\mathrm{E}\left[p_{t-1}p_t\right]$, $\mathrm{E}\left[p_{t-1}p_t^2\right]$, and $\mathrm{E}\left[p_{t-1}p_t^3\right]$ are computed as follows:

$$\begin{aligned}
\mathrm{E}\left[p_{t-1}p_t\right] &= \mathrm{E}\left[(\rho p_{t-1} + (1-\rho)\overline{p})p_{t-1}\right] \\
&= \rho\mathrm{E}\left[p_{t-1}^2\right] + (1-\rho)\overline{p}\mathrm{E}\left[p_{t-1}\right] \\
\mathrm{E}\left[p_{t-1}p_t^2\right] &= \mathrm{E}\left[(\rho p_{t-1} + (1-\rho)\overline{p})^2 p_{t-1}\right] + \sigma_p^2\mathrm{E}\left[p_{t-1}^2\right] \\
&= \rho^2\mathrm{E}\left[p_{t-1}^3\right] + 2\rho(1-\rho)\overline{p}\mathrm{E}\left[p_{t-1}^2\right] + (1-\rho)^2\overline{p}\mathrm{E}\left[p_{t-1}\right] \\
\mathrm{E}\left[p_{t-1}p_t^3\right] &= \mathrm{E}\left[(\rho p_{t-1} + (1-\rho)\overline{p})^3 p_{t-1}\right] + 3\sigma_p^2\mathrm{E}\left[p_{t-1}^2(\rho p_{t-1} + (1-\rho)\overline{p})\right] \\
&= \rho^3\mathrm{E}\left[p_{t-1}^4\right] + 3\left[\rho^2(1-\rho)\overline{p} + \rho\sigma_p^2\right]\mathrm{E}\left[p_{t-1}^3\right] + 3\left[\rho(1-\rho)^2\overline{p}^2 + (1-\rho)\overline{p}\sigma_p^2\right]\mathrm{E}\left[p_{t-1}^2\right] \\
&\quad + (1-\rho)^3\overline{p}^3\mathrm{E}\left[p_{t-1}\right].
\end{aligned}$$

Expectations $\mathrm{E}\left[p_{t-1}\right]$, $\mathrm{E}\left[p_{t-1}^2\right]$, $\mathrm{E}\left[p_{t-1}^3\right]$, and $\mathrm{E}\left[p_{t-1}^4\right]$ are computed as follows:

$$\begin{aligned}
\mathrm{E}\left[p_{t-1}\right] &= \overline{p} \\
\mathrm{E}\left[p_{t-1}^2\right] &= \overline{p}^2 + \frac{\sigma_p^2\overline{p}}{1-\rho^2} \\
\mathrm{E}\left[p_{t-1}^3\right] &= \overline{p}^3 + 3\overline{p}\frac{\sigma_p^2\overline{p}}{1-\rho^2} \\
\mathrm{E}\left[p_{t-1}^4\right] &= \overline{p}^4 + 6\overline{p}^2\frac{\sigma_p^2\overline{p}}{1-\rho^2} + 3\frac{\sigma_p^4\overline{p}^2}{(1-\rho^2)^2}.
\end{aligned}$$

Moreover, expectation $\mathrm{E}\left[g_tg_t\right]$ can be computed as follows:

$$\begin{aligned}
\mathrm{E}\left[g_tg_t\right] &= \mathrm{E}\left[x_t\left(u_{t+2} - \rho u_{t+1}\right)^2\right] \\
&= \mathrm{E}\left[x_tp_{t+1}(1-p_{t+1})\right] + \rho^2\mathrm{E}\left[x_tp_t(1-p_t)\right] \\
&= (1-\rho)\overline{p}\left[1-(1-\rho)\overline{p}\right]\mathrm{E}\left[p_{t-1}\right] - 2\rho^2\mathrm{E}\left[x_tp_t^2\right] + \left[\rho + \rho^2 - 2\rho(1-\rho)\overline{p} - \sigma_p^2\right]\mathrm{E}\left[x_tp_t\right].
\end{aligned}$$

Expectations $\mathrm{E}[x_t p_t]$ and $\mathrm{E}\left[x_t p_t^2\right]$ are

$$\mathrm{E}[x_t p_t] = (1-\rho)\overline{p}\mathrm{E}[p_{t-1}] + \rho\mathrm{E}\left[p_{t-1}^2\right] \tag{126}$$
$$\mathrm{E}\left[x_t p_t^2\right] = (1-\rho)^2\overline{p}^2\mathrm{E}[p_{t-1}] + 2\rho(1-\rho)\overline{p}\mathrm{E}\left[p_{t-1}^2\right] + \rho^2\mathrm{E}\left[p_{t-1}^3\right] + \sigma_p^2\mathrm{E}\left[p_{t-1}^2\right]. \tag{127}$$

Therefore, the asymptotic covariance is

$$\lim_{n\to+\infty} \mathrm{var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_t\right) = \mathrm{E}[g_t g_t] + 2\mathrm{E}[g_{t-1}g_t]. \tag{128}$$

5.2 Moment Rotations

Construct a lower block triangular matrix $L = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix}$ such that

$$\Omega^{-1} = L^T L. \tag{129}$$

It is most straightforward to analyze a rotated system of moment restrictions. Let

$$\tilde{m}_t(\theta) = Lm_t(\theta) = \begin{bmatrix} L_{11}m_t^{(1)}(\theta^{(1)}) \\ L_{21}m_t^{(1)}(\theta^{(1)}) + L_{22}m_t^{(2)}(\theta) \end{bmatrix} = \begin{bmatrix} \tilde{m}_t^{(1)}(\theta^{(1)}) \\ \tilde{m}_t^{(2)}(\theta) \end{bmatrix}. \tag{130}$$

Further, we let

$$\tilde{D} = LD = \begin{bmatrix} L_{11}D_{11} & 0 \\ L_{21}D_{11} + L_{22}D_{21} & L_{22}D_{22} \end{bmatrix} = \begin{bmatrix} \tilde{D}_{11} & 0 \\ \tilde{D}_{21} & \tilde{D}_{22} \end{bmatrix}. \tag{131}$$

For notational simplicity, we drop the ˜ but use the transformed system.

5.3 Hellinger-Differentiability Condition

The condition (72) is equivalent to the condition

$$\left(\frac{\mathrm{dQ}_{s,g}}{\mathrm{dQ}}\right)^{1/2} = 1 + \frac{1}{2}sg + s\varepsilon(s), \tag{132}$$

where $\varepsilon(s)$ converges to zero in $L^2(\mathrm{Q})$ as $s \to 0$. Equation (132) is equivalent to

$$\lim_{s\to 0}\int\left[\frac{1}{s}\left(\left(\frac{\mathrm{dQ}_{s,g}}{\mathrm{dQ}}\right)^{1/2} - 1\right) - \frac{1}{2}g\right]^2 \mathrm{dQ} = \lim_{s\to 0}\int \varepsilon(s)^2\mathrm{dQ} = 0. \tag{133}$$

5.4 The Expression of Λ

Let $D = [D_1, D_2]$ where $D_1^T = [D_{11}^T, D_{21}^T]$ and $D_2^T = [0, D_{22}^T]$. Thus, we have

$$P_2 = I - D_2\left(D_2^T D_2\right)^{-1} D_2^T = \begin{bmatrix} I & 0 \\ 0 & \Lambda_2 \end{bmatrix}, \tag{134}$$

where the matrix inversion is the generalized inversion.

Using the rules for the inversion of partitioned matrices, the matrix $\left(D^T D\right)^{-1}$ has the following expression:

$$\left[\begin{array}{cc} \left(D_1^T P_2 D_1\right)^{-1} & -\left(D_1^T P_2 D_1\right)^{-1} D_1^T D_2 \left(D_2^T D_2\right)^{-1} \\ -\left(D_2^T D_2\right)^{-1} D_2^T D_1 \left(D_1^T P_2 D_1\right)^{-1} & \left(D_2^T D_2\right)^{-1} + \left(D_2^T D_2\right)^{-1} D_2^T D_1 \left(D_1^T P_2 D_1\right)^{-1} D_1^T D_2 \left(D_2^T D_2\right)^{-1} \end{array}\right].$$

We can then show that

$$\begin{aligned} D\left(D^T D\right)^{-1} D^T &= D_1 \left(D_1^T P_2 D_1\right)^{-1} D_1^T - D_1 \left(D_1^T P_2 D_1\right)^{-1} D_1^T (I - P_2) \\ &\quad - (I - P_2) D_1 \left(D_1^T P_2 D_1\right)^{-1} D_1^T \\ &\quad + (I - P_2) + (I - M_2) D_1 \left(D_1^T P_2 D_1\right)^{-1} D_1^T (I - P_2) \\ &= I - P_2 + P_2 D_1 \left(D_1^T P_2 D_1\right)^{-1} D_1^T P_2. \end{aligned} \tag{135}$$

We conclude that

$$\Lambda = I - D\left(D^T D\right)^{-1} D^T = P_2 - P_2 D_1 \left(D_1^T P_2 D_1\right)^{-1} D_1^T P_2. \tag{136}$$

Recall that $\mathbf{I}_{\mathrm{F}} = D_1^T P_2 D_1$ (from Equation (27)). The matrix Λ can be rewritten as

$$\Lambda = \left[\begin{array}{cc} I - D_{11} \mathbf{I}_{\mathrm{F}}^{-1} D_{11}^T & D_{11} \mathbf{I}_{\mathrm{F}}^{-1} D_{11}^T \Lambda_2 \\ \Lambda_2 D_{11} \mathbf{I}_{\mathrm{F}}^{-1} D_{11}^T & \Lambda_2 - \Lambda_2 D_{21} \mathbf{I}_{\mathrm{F}}^{-1} D_{21}^T \Lambda_2 \end{array}\right]. \tag{137}$$

6 Disaster Risk Model: Solutions and Moments

6.1 Model Solution

We first show how to derive the Euler equation, and then how to obtain the dark matter measure $\varrho(p, \xi)$. The total return of market equity from t to $t+1$ is $e^{r_{M,t+1}}$, which is unknown at t, and the total return of the risk-free bond from t to $t+1$ is $e^{r_{f,t}}$, which is known at t. Thus, the excess log return of equity is $r_{t+1} = r_{M,t+1} - r_{f,t}$. The inter-temporal marginal rate of substitution is $\mathcal{M}_{t,t+1} = \delta e^{-\gamma g_{t+1}}$. The Euler equations for the market equity return and the risk-free rate are

$$1 = \mathbb{E}_t\left[\mathcal{M}_{t,t+1} e^{r_{M,t+1}}\right] \quad \text{and} \quad e^{-r_{f,t}} = \mathbb{E}_t\left[\mathcal{M}_{t,t+1}\right], \text{ respectively.} \tag{138}$$

Thus, we obtain the following Euler equation for the excess log return:

$$\mathbb{E}_t\left[\mathcal{M}_{t,t+1}\right] = \mathbb{E}_t\left[\mathcal{M}_{t,t+1} e^{r_{t+1}}\right]. \tag{139}$$

The left-hand side of (139) is equal to

$$\mathbb{E}_t\left[\mathcal{M}_{t,t+1}\right] = \mathbb{E}_t\left[e^{-\gamma g_{t+1}}\right] = (1-p) e^{-\gamma\mu + \frac{1}{2}\gamma^2\sigma^2} + p\xi \frac{e^{\gamma \underline{v}}}{\xi - \gamma},$$

and the right-hand side of (139) is equal to

$$\mathbb{E}_t\left[\mathcal{M}_{t,t+1}e^{r_{t+1}}\right] = \mathbb{E}_t\left[e^{-\gamma g_{t+1}+r_{t+1}}\right] = (1-p)e^{-\gamma\mu+\eta+\frac{1}{2}(\gamma^2\xi^2+\tau^2-2\gamma\rho\sigma\tau)} + p\xi\frac{e^{\frac{\varsigma^2}{2}+(\gamma-b)\underline{v}}}{\xi+b-\gamma}.$$

Thus, the Euler equation (139) can be rewritten as

$$(1-p)e^{-\gamma\mu+\frac{1}{2}\gamma^2\sigma^2}\left[e^{\eta+\frac{\tau^2}{2}-\gamma\rho\sigma\tau}-1\right] = p\Delta(\xi), \text{ where } \Delta(\xi) = \xi\left(\frac{e^{\gamma\underline{v}}}{\xi-\gamma}-\frac{e^{\frac{\varsigma^2}{2}+(\gamma-b)\underline{v}}}{\xi+b-\gamma}\right). \tag{140}$$

Using the Taylor expansion, we obtain the approximation

$$e^{\eta+\frac{\tau^2}{2}-\gamma\rho\sigma\tau}-1 \approx \eta+\frac{\tau^2}{2}-\gamma\rho\sigma\tau, \tag{141}$$

which, combined with (140), gives the following approximated Euler equation:

$$\overline{r}(p,\xi) = (1-p)\eta - p\ell\left(\underline{v}+1/\xi\right), \quad \text{where} \tag{142}$$

$$\eta = \gamma\rho\sigma\tau - \frac{\tau^2}{2} + e^{\gamma\mu-\frac{\gamma^2\sigma^2}{2}}\Delta(\xi)\frac{p}{1-p}, \text{ with } \Delta(\xi) = \xi\left[\frac{e^{\gamma\underline{v}}}{\xi-\gamma}-\frac{e^{\frac{\varsigma^2}{2}+(\gamma-\ell)\underline{v}}}{\xi+\ell-\gamma}\right]. \tag{143}$$

The term η in (142) is the log equity premium in the normal regime. The first two terms of η in (143) describe the market risk premia due to Gaussian consumption shocks; the third term is the disaster risk premium, which explodes as ξ approaches γ from above. In other words, there is an upper bound on the average disaster size for the equity premium to remain finite, which also limits how heavy the tail of the disaster size distribution can be.

6.2 Dark Matter Measure

Now, we show how to derive the dark matter measure. The Jacobian matrix of the moment restrictions and the asymptotic variance-covariance matrix are

$$D_{11} = \begin{bmatrix} -1 & 0 \\ 0 & -\frac{p}{\xi^2} \end{bmatrix} \text{ and } \Omega_{11} = \begin{bmatrix} p(1-p) & 0 \\ 0 & (1-p)\sigma^2+\frac{p}{\xi^2} \end{bmatrix} \approx \begin{bmatrix} p(1-p) & 0 \\ 0 & \frac{p}{\xi^2} \end{bmatrix}, \text{ respectively.} \tag{144}$$

The approximation above is simply due to the tiny magnitude of $\sigma^2 \approx 0$. The information matrix for the baseline model is

$$\Sigma_1 = D_{11}^T\Omega_{11}^{-1}D_{11} \approx \begin{bmatrix} \frac{1}{p(1-p)} & 0 \\ 0 & \frac{p}{\xi^2} \end{bmatrix}. \tag{145}$$

Next, the Jacobian matrix of moments restrictions and the asymptotic variance-covariance matrix for the full model are

$$D = \begin{bmatrix} -1 & 0 \\ 0 & -\frac{p}{\xi^2} \\ -(1-p)\frac{\partial \eta(p,\xi)}{\partial p} & -(1-p)\frac{\partial \eta(p,\xi)}{\partial \xi} - \frac{pb}{\xi^2} \end{bmatrix}, \tag{146}$$

and

$$\Omega = \begin{bmatrix} p(1-p) & 0 & 0 \\ 0 & (1-p)\sigma^2 + \frac{p}{\xi^2} & (1-p)\rho\sigma\tau + bp/\xi^2 \\ 0 & (1-p)\rho\sigma\tau + bp/\xi^2 & (1-p)\tau^2 + pb^2/\xi^2 \end{bmatrix}, \tag{147}$$

where

$$\eta(p,\xi) \equiv \gamma\rho\sigma\tau - \frac{\tau^2}{2} + \ln\left[1 + e^{\gamma\mu - \frac{\gamma^2\sigma^2}{2}}\xi\left(\frac{e^{\gamma\underline{v}}}{\xi-\gamma} - e^{\frac{1}{2}\varsigma^2}\frac{e^{(\gamma-b)\underline{v}}}{\xi+b-\gamma}\right)\frac{p}{1-p}\right]. \tag{148}$$

We can also derive the closed-form solution for the dark matter measure if we use the approximate Euler equation in (143). In this case, using the notation introduced in (143), we can express the information matrix for (p,ξ) under the full GMM model as

$$\Sigma \approx \begin{bmatrix} \frac{1}{p(1-p)} + \frac{\Delta(\xi)^2}{(1-\rho^2)\tau^2}\frac{e^{2\gamma_D\mu - \gamma_D^2\sigma^2}}{(1-p)^3} & \frac{p}{(1-\rho^2)\tau^2}\frac{e^{2\gamma_D\mu-\gamma_D^2\sigma^2}}{(1-p)^2}\Delta(\xi)\dot{\Delta}(\xi) \\ \frac{p}{(1-\rho^2)\tau^2}\frac{e^{2\gamma_D\mu-\gamma_D^2\sigma^2}}{(1-p)^2}\Delta(\xi)\dot{\Delta}(\xi) & \frac{p}{\xi^2} + \frac{\dot{\Delta}(\xi)^2}{(1-\rho^2)\tau^2}e^{2\gamma_D\mu-\gamma_D^2\sigma^2}\frac{p^2}{1-p} \end{bmatrix}, \tag{149}$$

where $\dot{\Delta}(\xi)$ is the first derivative of $\Delta(\xi)$, and

$$\dot{\Delta}(\xi) = -\frac{e^{\gamma\underline{v}}\gamma}{(\xi-\gamma)^2} + \frac{e^{(\gamma-\ell)\underline{v}}(\gamma-\ell)}{(\xi-\gamma+\ell)^2}e^{\varsigma^2/2}. \tag{150}$$

The largest eigenvalue of the matrix $\Sigma^{1/2}\Sigma_1^{-1}\Sigma^{1/2}$ is also the largest eigenvalue of $\Sigma_1^{-1/2}\Sigma\Sigma_1^{-1/2}$. In this case, the eigenvalues and eigenvectors are available in closed form. This gives us the formula for $\varrho(\theta)$ as follows:

$$\varrho(\theta) = 1 + \frac{p\Delta(\xi)^2 + p(1-p)\xi^2\dot{\Delta}(\xi)^2}{(1-\rho^2)\tau^2(1-p)^2}e^{2\gamma\mu-\gamma^2\sigma^2}. \tag{151}$$

7 Time-Varying Disaster Risk Model: Solutions and Moments

7.1 Model Solution

The model can be viewed as a discrete-time version of Wachter (2013). The representative agent has recursive preferences with unit elasticity of intertemporal substitution (EIS), and maximizes her lifetime utility V_t as follows:

$$\ln V_t = (1-\delta)\ln C_t + \delta(1-\gamma)^{-1}\ln \mathbb{E}_t\left[V_{t+1}^{1-\gamma}\right], \tag{152}$$

where C_t is consumption at time t, δ is the rate of time preference, and γ is the coefficient of risk aversion for timeless gambles. The log growth rate of consumption per capita, $\Delta c_{t+1} \equiv \ln(C_{t+1}/C_t)$, evolves as follows:

$$\Delta c_{t+1} = \mu + \sigma_c \varepsilon_{c,t+1} - \zeta_{t+1}, \tag{153}$$

where the consumption shock $\varepsilon_{c,t+1}$ follows a standard normal distribution, and ζ_{t+1} is a disaster variable characterized by

$$\zeta_{t+1} = z_{t+1} v_{t+1}, \tag{154}$$

where the variable u_{t+1} is a disaster shock following a truncated exponential distribution with lower bound $\underline{v}$:

$$v_{t+1} \sim \mathbf{1}\{v_{t+1} > \underline{v}\} \xi e^{-\xi(v_{t+1} - \underline{v})}, \tag{155}$$

and the Bernoulli variable z_{t+1} captures the occurrence of disasters with diaster probability $p_t = \max(\underline{p}, \widetilde{p}_t)$ and $\widetilde{p}_t$ evolving according to an AR(1) process:

$$\widetilde{p}_{t+1} = (1 - \rho)\overline{p} + \rho \widetilde{p}_t + \sigma_p \sqrt{p_t} \varepsilon_{p,t+1}. \tag{156}$$

We impose a small positive lower bound $\underline{p}$ (= 1 bps) on disaster probability p_t in solutions and simulations. Negative values of disaster probability can also be avoided by changing the specification. For example, the process of $\ln(p_t)$ can be specified as an AR(1) process as in Gourio (2012), and the disaster probability can be specified as $\max(p_t, 0)$ with boom jump to be $\max(-p_t, 0)$ as in Cheng, Dou, and Liao (2021).

We model dividends D_t as levered consumption with log dividend growth $\Delta d_{t+1} \equiv \ln(D_{t+1}/D_t)$:

$$\Delta d_{t+1} = \mu - \frac{1}{2}\varphi^2 \sigma_c^2 + \phi \sigma_c \varepsilon_{c,t+1} - \phi \zeta_{t+1} + \varphi \sigma_c \varepsilon_{d,t+1}, \tag{157}$$

similar in spirit to Abel (1999).

The shocks $(\varepsilon_{c,t+1}, \varepsilon_{d,t+1}, \varepsilon_{p,t+1}, J_{t+1})$ are mutually independent and i.i.d. over t. The Bernoulli variables z_{t+1} are independent of the contemporaneous jump probability shock $\varepsilon_{p,t+1}$ and its leads in the time series, but z_{t+1} and the lags of $\varepsilon_{p,t+1}$ are dependent through the jump probability p_t. The two processes z_{t+1} and $(\varepsilon_{c,t+1}, \varepsilon_{d,t+1}, J_{t+1})$ are mutually independent.

Because the EIS coefficient is one, the first-order condition of optimal consumption results in $C_t = (1-\delta)W_t$. Due to the homotheticity of the preference, it is natural to conjecture that

$$V_t = \mathcal{I}(p_t) C_t, \tag{158}$$

where $\mathcal{I}(p_t)$ is a deterministic function of p_t, capturing the marginal value of net worth. The specification of the dynamics is consistent with the exponential-affine models, and thus, we further conjecture that

$$\mathcal{I}(p_t) = e^{I_0 + I_1 p_t} \tag{159}$$

with constants I_0 and I_1 to be determined by equilibrium conditions.

The constants I_0 and I_1 can be solved by plugging (158) and (159) into (152). Specifically, it holds that

$$I_0 + I_1 p_t + \ln C_t = (1-\delta)\ln C_t + (1-\gamma)^{-1}\delta \ln \mathbb{E}_t\left[e^{(1-\gamma)(I_0+I_1 p_{t+1})} C_{t+1}^{1-\gamma}\right]. \tag{160}$$

By matching the constant term and p_t term, we obtain that

$$I_1 \approx \delta I_1 \rho + \frac{1}{2}(1-\gamma)\delta I_1^2 \sigma_p^2 + (1-\gamma)^{-1}\delta \Xi(\gamma-1) \tag{161}$$

$$I_0 \approx \delta I_0 + \delta I_1 (1-\rho)\overline{p} + \delta\mu + \frac{1}{2}(1-\gamma)\delta\sigma_c^2, \tag{162}$$

with $\Xi(x) \equiv e^{x\underline{v}}\dfrac{\xi}{\xi - x} - 1$. Equation (161) has two roots:

$$I_1 = \frac{1-\delta\rho \pm \sqrt{(1-\delta\rho)^2 - 2\delta^2\sigma_p^2 \Xi(\gamma-1)}}{(1-\gamma)\delta\sigma_p^2}. \tag{163}$$

Economic intuition can help select the reasonable root. When $\underline{v} \to 0$ and $\xi \to +\infty$, the disaster risk becomes negligible. In the limit, the value function should become independent of p_t, which rules out the root $\dfrac{1-\delta\rho + \sqrt{(1-\delta\rho)^2 - 2\delta^2\sigma_p^2\Xi(\gamma-1)}}{(1-\gamma)\delta\sigma_p^2}$ since it approaches to $\dfrac{2(1-\delta\rho)}{(1-\gamma)\delta\sigma_p^2} < 0$. Therefore, the relevant solution to (161) and (162) is

$$I_1 = \frac{1-\delta\rho - \sqrt{(1-\delta\rho)^2 - 2\delta^2\sigma_p^2 \Xi(\gamma-1)}}{(1-\gamma)\delta\sigma_p^2} \tag{164}$$

$$I_0 = \frac{\delta}{1-\delta}\left[I_1(1-\rho)\overline{p} + \mu + \frac{1}{2}(1-\gamma)\sigma_c^2\right]. \tag{165}$$

To ensure the existence of the equilibrium, the following restrictions on model parameters need to be satisfied:

$$(1-\delta\rho)^2 - 2\delta^2\sigma_p^2\Xi(\gamma-1) > 0. \tag{166}$$

The equilibrium stochastic discount factor (SDF) is

$$M_{t+1} = \delta\left(\frac{C_{t+1}}{C_t}\right)^{-1} \frac{V_{t+1}^{1-\gamma}}{\mathbb{E}_t\left[V_{t+1}^{1-\gamma}\right]}. \tag{167}$$

After plugging in the equilibrium value function and rearranging the terms, we get the log SDF, denoted by $m_{t+1} \equiv \ln M_{t+1}$ as follows:

$$m_{t+1} \approx \Gamma_0 + \Gamma_1 p_t - \lambda_c\sigma_c\varepsilon_{c,t+1} - \lambda_p\sigma_p\sqrt{p_t}\varepsilon_{p,t+1} + \lambda_\zeta\zeta_{t+1}, \tag{168}$$

where the predictive coefficients are

$$\Gamma_0 = \ln\delta - \mu - \frac{1}{2}(1-\gamma)^2\sigma_c^2 \tag{169}$$

$$\Gamma_1 = -\frac{1}{2}(1-\gamma)^2 I_1^2\sigma_p^2 - \Xi(\gamma-1), \tag{170}$$

and the loading coefficients are

$$\lambda_c = \gamma, \quad \lambda_p = (\gamma-1)I_1, \quad \text{and } \lambda_\zeta = \gamma. \tag{171}$$

The log risk-free rate, denoted by $r_{f,t} = -\ln\mathbb{E}_t[M_{t+1}]$, is

$$\begin{aligned} r_{f,t} &= -\ln\mathbb{E}_t\left[e^{m_{t+1}}\right] \\ &= -\Gamma_0 - \frac{1}{2}\lambda_c^2\sigma_c^2 - \left[\Gamma_1 + \frac{1}{2}\lambda_p^2\sigma_p^2 + \Xi(\lambda_\zeta)\right]p_t \quad (172) \\ &= -\ln\delta + \mu + \frac{1}{2}(1-\gamma)^2\sigma_c^2 - \frac{1}{2}\gamma^2\sigma_c^2 - \left[\Xi(\gamma) - \Xi(\gamma-1)\right]p_t \quad (173) \end{aligned}$$

Using the Campbell-Shiller decomposition and linearization, we can represent the return in terms of log price-dividend ratio and log dividend growth:

$$r_{m,t+1} = \kappa_{m,0} + \kappa_{m,1}z_{m,t+1} + \Delta d_{t+1} - z_{m,t}, \tag{174}$$

where

$$\kappa_{m,0} = \log(1+e^{\overline{z}_m}) - \kappa_{m,1}\overline{z}_m, \tag{175}$$

and

$$\kappa_{m,1} = \frac{e^{\overline{z}_m}}{1+e^{\overline{z}_m}} \tag{176}$$

and $\overline{z}_m$ is long-run mean of market log price-dividend ratio.

Using the log-linearization approximation, we search the equilibrium characterized by

$$z_{m,t} = A_{m,0} + A_{m,1}p_t, \tag{177}$$

where the constants $A_{m,0}$ and $A_{m,1}$ can be computed recursively as follows.

Define the period-t price of the dividend strip paid at the period $t+n$ as $H(D_t, p_t, n) = \mathbb{E}_t\left[M_{t,t+n}D_{t+n}\right]$ where $M_{t,t+n} \equiv e^{\sum_{i=1}^n m_{t+i}}$. The price function $H(D_t, p_t, n)$ satisfies the following recursive relations:

$$H(D_t, p_t, n) = \mathbb{E}_t\left[e^{m_{t+1}}H(D_{t+1}, p_{t+1}, n-1)\right] \tag{178}$$

$$H(D_t, p_t, 0) = D_t, \tag{179}$$

for arbitrary t and $n \geq 1$.

We conjecture that $H(D_t, p_t, n) = D_t e^{A_n + B_n p_t}$. Then, the recursive relations in (178) and (179) can be rewritten as follows:

$$\begin{aligned}
e^{A_n+B_np_t} &= \mathbb{E}_t\left[e^{\Delta d_{t+1}+m_{t+1}+A_{n-1}+B_{n-1}p_{t+1}}\right]\\
&= \mathbb{E}_t\left[e^{(\mu-\frac{1}{2}\varphi^2\sigma_c^2+\phi\sigma_c\varepsilon_{c,t+1}-\phi\zeta_{t+1}+\varphi\sigma_c\varepsilon_{d,t+1})+(\Gamma_0+\Gamma_1p_t-\lambda_c\sigma_c\varepsilon_{c,t+1}-\lambda_p\sigma_p\sqrt{p_t}\varepsilon_{p,t+1}+\lambda_\zeta\zeta_{t+1})+(A_{n-1}+B_{n-1}p_{t+1})}\right]\\
&= e^{\tilde{A}_n+\tilde{B}_np_t-\frac{1}{2}\varphi^2\sigma_c^2}\mathbb{E}_t\left[e^{(\phi-\lambda_c)\sigma_c\varepsilon_{c,t+1}+(B_{n-1}-\lambda_p)\sigma_p\sqrt{p_t}\varepsilon_{p,t+1}+(\lambda_\zeta-\phi)\zeta_{t+1}+\varphi\sigma_c\varepsilon_{d,t+1}}\right],
\end{aligned}$$

where $\tilde{A}_n = \mu + \Gamma_0 + A_{n-1} + B_{n-1}(1-\rho)\overline{p}$, and $\tilde{B}_n = \Gamma_1 + B_{n-1}\rho$.

The moment generating function of ζ_{t+1} is

$$\ln \mathbb{E}_t\left[e^{(\lambda_\zeta-\phi)\zeta_{t+1}}\right] \approx p_t\Xi(\gamma-\phi). \tag{180}$$

Thus, it holds that

$$\begin{aligned}
A_n &= \tilde{A}_n + \frac{1}{2}(\phi-\gamma)^2\sigma_c^2\\
&= \mu + \Gamma_0 + A_{n-1} + B_{n-1}(1-\rho)\overline{p} + \frac{1}{2}(\phi-\gamma)^2\sigma_c^2\\
&= \ln\delta - \frac{1}{2}(1-\gamma)^2\sigma_c^2 + A_{n-1} + B_{n-1}(1-\rho)\overline{p} + \frac{1}{2}(\phi-\gamma)^2\sigma_c^2,
\end{aligned}$$

and

$$\begin{aligned}
B_n &= \tilde{B}_n + \frac{1}{2}\left[B_{n-1}-(\gamma-1)I_1\right]^2\sigma_p^2 + \Xi(\gamma-\phi)\\
&= \rho B_{n-1} + \frac{1}{2}B_{n-1}^2\sigma_p^2 - (\gamma-1)I_1\sigma_p^2B_{n-1} + \Xi(\gamma-\phi) - \Xi(\gamma-1),
\end{aligned}$$

with the initial values $A_0 = B_0 = 0$.

Therefore, the log price-dividend ratio is

$$z_{m,t} = \ln\left[\sum_{n=1}^{+\infty} e^{A_n+B_np_t}\right]. \tag{181}$$

According to Taylor expansion in terms of p_t around $\overline{p}$, it follows that

$$A_{m,0} = \ln\left[\sum_{n=1}^{+\infty} e^{A_n+B_n\overline{p}}\right] - A_{m,1}\overline{p} \quad\text{and}\quad A_{m,1} = \frac{\sum_{n=1}^{+\infty} B_ne^{A_n+B_n\overline{p}}}{\sum_{n=1}^{+\infty} e^{A_n+B_n\overline{p}}}. \tag{182}$$

According to (174), the equilibrium log market return can be rewritten as

$$r_{m,t+1} = \mathbb{E}_t\left[r_{m,t+1}\right] + \beta_c\sigma_c\varepsilon_{c,t+1} + \beta_p\sigma_p\sqrt{p_t}\varepsilon_{p,t+1} - \beta_\zeta\left[\zeta_{t+1} - p_t(\underline{v}+1/\xi)\right] + \varphi\sigma_c\varepsilon_{d,t+1}, \tag{183}$$

where $\beta_c = \phi$, $\beta_p = \kappa_{m,1}A_{m,1}$, and $\beta_\zeta = \phi$.

The Euler equation for the log market return is

$$1 = \mathbb{E}_t\left[e^{r_{m,t+1}+m_{t+1}}\right], \tag{184}$$

which leads to the following equilibrium characterization of conditional equity premium:

$$
\begin{aligned}
&\mathbb{E}_t\left[r_{m,t+1}\right]-r_{f,t} \\
&\quad=\beta_c\lambda_c\sigma_c^2+\beta_p\lambda_p\sigma_p^2p_t+\left[\Xi(\gamma)-\Xi(\gamma-\phi)-\phi(\underline{v}+1/\xi)\right]p_t-\frac{1}{2}\left[(\beta_c^2+\varphi^2)\sigma_c^2+\beta_p^2\sigma_p^2p_t\right]. \qquad (185)
\end{aligned}
$$

The conditional variance of the log market return is

$$
\operatorname{var}_t(r_{m,t+1})=\beta_c^2\sigma_c^2+\varphi^2\sigma_c^2+\beta_p^2\sigma_p^2p_t+\beta_\zeta^2\left[(\underline{v}+1/\xi)^2(1-p_t)+1/\xi^2\right]p_t. \qquad (186)
$$

Now, we derive the yield of the defaultable government bond, denoted by $y_{b,t}$, and the expected return of the defaultable government bond, denoted by $\mu_{b,t}\equiv\mathbb{E}_t\left[r_{b,t+1}\right]$, where

$$
r_{b,t+1}=\begin{cases} y_{b,t}, & \text{if not default} \\ y_{b,t}-v_{t+1}, & \text{if default.}\end{cases} \qquad (187)
$$

A default on the government bond occurs with probability q conditional on the occurrence of a disaster. Thus, by definition, it holds that

$$
\mu_{b,t}=y_{b,t}-p_tq(\underline{v}+1/\xi). \qquad (188)
$$

According to the Euler equation of the defaultable government bond and the risk-free bond, it holds that

$$
\mathbb{E}_t\left[e^{m_{t+1}+r_{b,t+1}-r_{f,t}}\right]=\mathbb{E}_t\left[e^{m_{t+1}}\right]. \qquad (189)
$$

Some calculations show that the following relation approximately holds:

$$
\begin{aligned}
&\ln\mathbb{E}_t\left[e^{m_{t+1}+r_{b,t+1}-r_{f,t}}\right] \\
&\quad=\Gamma_0+\Gamma_1p_t+y_{b,t}-r_{f,t}+\frac{1}{2}\lambda_c^2\sigma_c^2+\frac{1}{2}\lambda_p^2\sigma_p^2p_t+\left[(1-q)\Xi(\gamma)-q\Xi(\gamma-1)\right]p_t. \qquad (190)
\end{aligned}
$$

Combining (172), (189), and (190), it follows that

$$
y_{b,t}-r_{f,t}=q\left[\Xi(\gamma)-\Xi(\gamma-1)\right]p_t. \qquad (191)
$$

Further, by putting together equations (188) and (191), we can obtain the following relation:

$$
\mu_{b,t}-r_{f,t}=q\left[\Xi(\gamma)-\Xi(\gamma-1)-(\underline{v}+1/\xi)\right]p_t. \qquad (192)
$$

Therefore, the conditional mean and variance of excess log returns of the market portfolio relative to the defaultable government bill are

$$
\begin{aligned}
\mathbb{E}_t\left[r_{m,t+1}-r_{b,t+1}\right]=&\beta_c\lambda_c\sigma_c^2+\beta_p\lambda_p\sigma_p^2p_t+\left[\Xi(\gamma)-\Xi(\gamma-\phi)-\phi(\underline{v}+1/\xi)\right]p_t \\
&-\frac{1}{2}\left[(\beta_c^2+\varphi^2)\sigma_c^2+\beta_p^2\sigma_p^2p_t\right]-q\left[\Xi(\gamma)-\Xi(\gamma-1)-(\underline{v}+1/\xi)\right]p_t, \qquad (193)
\end{aligned}
$$

and

$$\begin{aligned}\mathrm{var}_t\left[r_{m,t+1} - r_{b,t+1}\right] = \beta_c^2\sigma_c^2 + \beta_p^2\sigma_p^2 p_t + \varphi^2\sigma_c^2 + \left[(1-q)\phi^2 + q(\phi-1)^2\right]\left[(\underline{v}+1/\xi)^2 + 1/\xi^2\right]p_t \\ - (\phi - q)^2(\underline{v}+1/\xi)^2 p_t^2 \end{aligned} \tag{194}$$

Some More Derivations. From equation (164), it follows that

$$I_1 \approx \frac{\delta^{-1} - \rho}{(1-\gamma)\sigma_p^2}, \tag{195}$$

because $(\delta^{-1} - \rho)^2 \approx 2\sigma_p^2 \Xi(\gamma - 1)$.

The affine coefficient B_n has the following (approximate) recursive relation:

$$B_n = \left[\rho + \frac{2\Xi(\gamma-1)}{\delta^{-1}-\rho}\sigma_p^2\right] B_{n-1} + \Xi(\gamma - \phi) - \Xi(\gamma - 1). \tag{196}$$

Let $\hat{\rho} \equiv \rho + \frac{2\Xi(\gamma-1)}{\delta^{-1}-\rho}\sigma_p^2$. When $\hat{\rho} < 1$, it holds that

$$B_n = \frac{1-\hat{\rho}^n}{1-\hat{\rho}}\left[\Xi(\gamma-\phi) - \Xi(\gamma-1)\right]. \tag{197}$$

The coefficient A_n can be expressed as

$$\begin{aligned} A_n - A_{n-1} &= \ln\delta - \frac{1}{2}(1-\gamma)^2\sigma_c^2 + \frac{1-\hat{\rho}^n}{1-\hat{\rho}}\left[\Xi(\gamma-\phi) - \Xi(\gamma-1)\right](1-\rho)\overline{p} + \frac{1}{2}(\phi-\gamma)^2\sigma_c^2 && (198) \\ &\approx \ln\delta + \frac{1}{2}(\phi-1)(\phi+1-2\gamma)\sigma_c^2 + (1-\hat{\rho}^n)\left[\Xi(\gamma-\phi) - \Xi(\gamma-1)\right]\overline{p}. && (199) \end{aligned}$$

Thus, it holds that

$$A_n = n\left\{\ln\delta + \frac{1}{2}(\phi-1)(\phi+1-2\gamma)\sigma_c^2 + \left[\Xi(\gamma-\phi) - \Xi(\gamma-1)\right]\overline{p}\right\} - \hat{\rho}\frac{1-\hat{\rho}^n}{1-\hat{\rho}}\left[\Xi(\gamma-\phi) - \Xi(\gamma-1)\right]\overline{p}. \tag{200}$$

and hence, it holds that

$$A_n + B_n\overline{p} \approx -n\hat{\delta}, \quad \text{where} \quad \hat{\delta} \equiv -\left\{\ln\delta + \frac{1}{2}(\phi-1)(\phi+1-2\gamma)\sigma_c^2 + \left[\Xi(\gamma-\phi) - \Xi(\gamma-1)\right]\overline{p}\right\}. \tag{201}$$

Therefore, the coefficients in (177) satisfies the following relations in equilibrium:

$$\begin{aligned} A_{m,1} &= \left[\Xi(\gamma-\phi) - \Xi(\gamma-1)\right]\frac{\sum_{n=1}^{+\infty}\frac{1-\hat{\rho}^n}{1-\hat{\rho}}e^{-n\hat{\delta}}}{\sum_{n=1}^{+\infty}e^{-n\hat{\delta}}} && (202) \\ &\approx \Xi(\gamma-\phi) - \Xi(\gamma-1). && (203) \end{aligned}$$

and

$$A_{m,0} = \ln\left[\sum_{n=1}^{+\infty} e^{-n\hat{\delta}}\right] - A_{m,1}\overline{p} \tag{204}$$

$$= \ln\left[\frac{e^{-\hat{\delta}}}{1-e^{\hat{\delta}}}\right] - A_{m,1}\overline{p}. \tag{205}$$

And thus, the steady-state log price dividend ratio is

$$\overline{z}_m = A_{m,0} + A_{m,1}\overline{p} = \ln\left[\frac{e^{-\hat{\delta}}}{1-e^{\hat{\delta}}}\right]. \tag{206}$$

The log-linearization coefficient is

$$\kappa_{m,1} = e^{-\hat{\delta}} = \delta e^{\frac{1}{2}(\phi-1)(\phi+1-2\gamma)\sigma_c^2+[\Xi(\gamma-\phi)-\Xi(\gamma-1)]\overline{p}}. \tag{207}$$

The beta to the time-varying disaster risk is

$$\beta_p = \kappa_{m,1}A_{m,1} = \delta\left[\Xi(\gamma-\phi) - \Xi(\gamma-1)\right]. \tag{208}$$

The market price of risk is

$$\lambda_p = (\gamma-1)I_1 = -\frac{\delta^{-1}-\rho}{\sigma_p^2}. \tag{209}$$

The equilibrium conditional mean and variance of excess log returns of the market portfolio relative to the government bill are

$$\begin{aligned}\mathbb{E}_t\left[r_{m,t+1} - r_{b,t+1}\right] =& \phi\gamma\sigma_c^2 + \left[\Xi(\gamma-1) - \Xi(\gamma-\phi)\right](1-\delta\rho)p_t + \left[\Xi(\gamma) - \Xi(\gamma-\phi) - \phi(\underline{v} + 1/\xi)\right]p_t \\ &- \frac{1}{2}\phi^2\sigma_c^2 - q\left[\Xi(\gamma) - \Xi(\gamma-1) - (\underline{v}+1/\xi)\right]p_t,\end{aligned} \tag{210}$$

and

$$\begin{aligned}\mathrm{var}_t\left[r_{m,t+1} - r_{b,t+1}\right] =& \beta_c^2\sigma_c^2 + \beta_p^2\sigma_p^2 p_t + \varphi^2\sigma_c^2 + \left[(1-q)\phi^2 + q(\phi-1)^2\right]\left[(\underline{v}+1/\xi)^2 + 1/\xi^2\right]p_t \\ &- (\phi-q)^2(\underline{v}+1/\xi)^2 p_t^2.\end{aligned} \tag{211}$$

7.2 Generalized Methods of Moments

Denote the set of moment functions for the baseline model to be $m^{(1)}(\mathbf{y}_{t-1}, \mathbf{y}_t; \theta^{(1)})$ with the data $\mathbf{y}_t = (\Delta c_t, \Delta d_t, x_t, x_{t-1}, x_{b,t}, v_t, z_{m,t}, r_{e,t})$ and the baseline parameters $\theta^{(1)} = (\mu, \sigma_c^2, \overline{p}, \rho, \sigma_p^2, \xi, \phi, \varphi, q)^T$. More precisely, there are eight moment conditions specified as follows:

$$\mathbb{E}\left[m^{(1)}(\mathbf{y}_{t-1}, \mathbf{y}_t; \theta^{(1)})\right] = 0, \tag{212}$$

with

$$m^{(1)}(\mathbf{y}_{t-1}, \mathbf{y}_t; \theta^{(1)}) = \begin{bmatrix} x_t - \overline{p} \\ x_{t-2}\left[x_t - \rho x_{t-1} - (1-\rho)\overline{p}\right] \\ x_{t-1}(x_t - \overline{p}) - \rho\sigma_p^2\overline{p}/(1-\rho^2) \\ x_{b,t-1} - q\overline{p} \\ \Delta c_{t+1} - \mu + \overline{p}(\underline{v} + 1/\xi) \\ (\Delta c_{t+1} - \mu)^2 - \sigma_c^2 - \overline{p}[(\underline{v} + 1/\xi)^2 + 1/\xi^2] \\ \Delta c_{t+1} - z_{t+1}v_{t+1} - \mu \\ \Delta d_{t+1} - \phi\Delta c_{t+1} - (1-\phi)\mu + \frac{1}{2}\varphi^2\sigma_c^2 \\ [\Delta d_{t+1} - \phi\Delta c_{t+1} - (1-\phi)\mu + \frac{1}{2}\varphi^2\sigma_c^2]^2 - \varphi^2\sigma_c^2 \end{bmatrix}. \tag{213}$$

The first row of Jacobian matrix $D_{11}(\theta)$ for the baseline moment restrictions is

$$[0, 0, -1, 0, 0, 0, 0, 0, 0]. \tag{214}$$

The second row of Jacobian matrix $D_{11}(\theta)$ for the baseline moment restrictions is

$$[0, 0, -\overline{p}(1-\rho), -\rho\sigma_p^2\overline{p}/(1-\rho^2), 0, 0, 0, 0, 0]. \tag{215}$$

The third row of Jacobian matrix $D_{11}(\theta)$ for the baseline moment restrictions is

$$[0, 0, -\overline{p} - \rho\sigma_p^2/(1-\rho^2), -\sigma_p^2\overline{p}(1+\rho^2)/(1-\rho^2)^2, -\rho\overline{p}/(1-\rho^2), 0, 0, 0, 0]. \tag{216}$$

The fourth row of Jacobian matrix $D_{11}(\theta)$ for the baseline moment restrictions is

$$[0, 0, -q, 0, 0, 0, 0, 0, -\overline{p}]. \tag{217}$$

The fifth row of Jacobian matrix $D_{11}(\theta)$ for the baseline moment restrictions is

$$[-1, 0, \underline{v} + 1/\xi, 0, 0, -\overline{p}/\xi^2, 0, 0, 0]. \tag{218}$$

The sixth row of Jacobian matrix $D_{11}(\theta)$ for the baseline moment restrictions is

$$[2\overline{p}(\underline{v} + 1/\xi), -1, -(\underline{v} + 1/\xi)^2 - 1/\xi^2, 0, 0, 2\overline{p}(\underline{v}/\xi^2 + 2/\xi^3), 0, 0, 0]. \tag{219}$$

The seventh row of Jacobian matrix $D_{11}(\theta)$ for the baseline moment restrictions is

$$[-1, 0, 0, 0, 0, 0, 0, 0, 0]. \tag{220}$$

The eighth row of Jacobian matrix $D_{11}(\theta)$ for the baseline moment restrictions is

$$[-(1-\phi), \frac{1}{2}\varphi^2, 0, 0, 0, 0, \overline{p}(\underline{v} + 1/\xi), \varphi\sigma_c^2, 0]. \tag{221}$$

The ninth row of Jacobian matrix $D_{11}(\theta)$ for the baseline moment restrictions is

$$[0, -\varphi^2, 0, 0, 0, 0, 0, -2\varphi\sigma_c^2, 0]. \tag{222}$$

In time-varying disaster risk models, the major focus is to understand the equity premium, the average volatility, the predictability of excess returns based on price-dividend ratios, and the comovement between excess returns and price-dividend ratios explained by the consumption process and dividend process specified in (153) – (157). These asset pricing cross-equation restrictions can be specified as follows:

$$\mathbb{E}\left[m^{(2)}(\mathbf{y}_t, \mathbf{y}_{t-1}; \theta)\right] = 0, \tag{223}$$

with

$$m^{(2)}(\mathbf{y}_t, \mathbf{y}_{t-1}; \theta) = \left[\begin{array}{l} r_{m,t} - r_{b,t} - \chi_1(\theta) \\ (r_{m,t} - r_{b,t})^2 - \chi_2(\theta) \\ r_t^e - \chi_3(\theta) z_{m,t} - \chi_4(\theta) z_{m,t-1} - \chi_5(\theta) \end{array}\right],$$

where $r_t^e \equiv r_{m,t} - r_{b,t} + (\phi - x_{b,t})\zeta_t$, and

$$\begin{aligned}
\chi_1(\theta) &= \phi\gamma\sigma_c^2 + \beta_p\lambda_p\sigma_p^2\overline{p} + \left[\Xi(\gamma) - \Xi(\gamma-\phi) - \phi(\underline{v} + 1/\xi)\right]\overline{p} - \frac{1}{2}\left(\phi^2\sigma_c^2 + \beta_p^2\sigma_p^2\overline{p} + \varphi^2\sigma_c^2\right), \\
&\quad -q\left[\Xi(\gamma) - \Xi(\gamma-1) - (\underline{v} + 1/\xi)\right]\overline{p}, \\
\chi_2(\theta) &= \mathbb{E}\left[r_{m,t} - r_{b,t}\right]^2 + \mathrm{var}\left[r_{m,t} - r_{b,t}\right] = \chi_1(\theta)^2 + \mathbb{E}\left[\mathrm{var}_{t-1}(r_{m,t} - r_{b,t})\right] + \mathrm{var}\left[\mathbb{E}_{t-1}(r_{m,t} - r_{b,t})\right], \\
\chi_3(\theta) &= A_{m,1}^{-1}\beta_p, \\
\chi_4(\theta) &= A_{m,1}^{-1}\left\{\left[\beta_p\lambda_p\sigma_p^2 + \Xi(\gamma) - \Xi(\gamma-\phi) - \frac{1}{2}\beta_p^2\sigma_p^2\right] - q\left[\Xi(\gamma) - \Xi(\gamma-1)\right] - \beta_p\rho\right\}, \\
\chi_5(\theta) &= \phi\gamma\sigma_c^2 - A_{m,0}\left[\chi_3(\theta) + \chi_4(\theta)\right] - \beta_p(1-\rho)\overline{p} - \frac{1}{2}\left(\beta_c^2\sigma_c^2 + \varphi^2\sigma_c^2\right).
\end{aligned}$$

In the definition of $\chi_2(\theta)$ above, the expectation of conditional variance is

$$\begin{aligned}
\mathbb{E}\left[\mathrm{var}_{t-1}(r_{m,t} - r_{b,t})\right] &= \phi^2\sigma_c^2 + \varphi^2\sigma_c^2 + \beta_p^2\sigma_p^2\overline{p} + \left[(1-q)\phi^2 + q(\phi-1)^2\right]\left[(\underline{v} + 1/\xi)^2 + 1/\xi^2\right]\overline{p} \\
&\quad - (\phi - q)^2(\underline{v} + 1/\xi)^2\left(\overline{p}^2 + \frac{\sigma_p^2}{1-\rho^2}\overline{p}\right),
\end{aligned} \tag{224}$$

and, the variance of conditional expectation is

$$\mathrm{var}\left[\mathbb{E}_{t-1}(r_{m,t} - r_{b,t})\right] = A_{m,1}^2\left[\chi_4(\theta) + \rho\chi_3(\theta)\right]^2\sigma_p^2\overline{p}/(1-\rho^2). \tag{225}$$

The parameter vector $\theta = \left[\begin{array}{c} \theta^{(1)} \\ \theta^{(2)} \end{array}\right]$ includes the baseline parameter $\theta^{(1)} = (\mu, \sigma_c^2, \overline{p}, \rho, \sigma_p^2, \xi, \phi, \varphi, q)^T$ and the nusance parameter $\theta^{(2)} = \gamma$. The auxiliary parameters are $\underline{v}$ and δ, treated as part of the functional-form specification of the model.

The analytical formulas for the Jacobian matrix of the over-identification moment conditions are quite complicated. We ignore the formulas here and, in fact, we calculate them numerically in obtaining the fragility measures. Moreover, we compute the Fisher Information matrices for the moments in $m^{(1)}(\cdot, \theta^{(1)})$ and $m(\cdot, \theta)$ based on simulated stationary time series using the Monte Carlo method.

8 Long-Run Risk Model: Solutions and Moments

8.1 Model Solution

We consider a long-run risk model similar to Bansal and Yaron (2004) and Bansal, Kiku, and Yaron (2012). The log growth rate of aggregate consumption Δc_t, the long-run risk component in consumption growth x_t, and stochastic volatility σ_t follow the joint processes

$$\Delta c_{t+1} = \mu_c + x_t + \sigma_t \epsilon_{c,t+1} \tag{226a}$$

$$x_{t+1} = \rho x_t + \varphi_x \sigma_t \epsilon_{x,t+1} \tag{226b}$$

$$\widetilde{\sigma}^2_{t+1} = \overline{\sigma}^2 + \nu(\widetilde{\sigma}^2_t - \overline{\sigma}^2) + \sigma_w \epsilon_{\sigma,t+1} \tag{226c}$$

$$\sigma^2_{t+1} = \max(\underline{\sigma}^2, \widetilde{\sigma}^2_{t+1}), \tag{226d}$$

where the shocks $\epsilon_{c,t}, \epsilon_{x,t}$, and $\epsilon_{\sigma,t}$ are i.i.d. standard normal variables and they are mutually independent. Similar to Bansal and Yaron (2004) and Bansal, Kiku, and Yaron (2012), we adopt the local approximation method to linearize the model and thus the solution. In the log-linearized approximation system, it is fair to assume that $\sigma^2_t = \widetilde{\sigma}^2_t$.

The representative agent has Epstein-Zin-Weil preferences:

$$V_t = \left[(1-\delta) C_t^{\frac{1-\gamma}{\vartheta}} + \delta \left(\mathbb{E}_t \left[V_{t+1}^{1-\gamma} \right] \right)^{\frac{1}{\vartheta}} \right]^{\frac{\vartheta}{1-\gamma}} \tag{227}$$

where $\vartheta = (1-\gamma)/(1-\psi^{-1})$. Define the wealth process and the gross return on consumption claims:

$$W_{t+1} = (W_t - C_t) R_{c,t+1}. \tag{228}$$

Therefore, the stochastic discount factor (SDF) can be expressed as follows:

$$M_{t+1} = \delta^{\vartheta} \left(\frac{C_{t+1}}{C_t} \right)^{-\vartheta/\psi} R_{c,t+1}^{\vartheta - 1}. \tag{229}$$

The log SDF can be written as

$$m_{t+1} = \vartheta \log \delta - \frac{\vartheta}{\psi} \Delta c_{t+1} + (\vartheta - 1) r_{c,t+1}. \tag{230}$$

The state variables in long-run risk models are x_t and σ^2_t. The dependence of $r_{c,t+1}$ on the state variables are endogenous. To turn the model into an affine system, we first exploit the Campbell-Shiller log-linearization approximation:

$$r_{c,t+1} = \kappa_0 + \kappa_1 z_{t+1} + \Delta c_{t+1} - z_t, \tag{231}$$

where $z_t = \log(W_t/C_t)$ is log wealth-consumption ratio and wealth is the price of consumption claims. The log-linearization constants are determined by the long-run steady state:

$$\kappa_0 = \log(1 + e^{\overline{z}}) - \kappa_1 \overline{z} \tag{232}$$

$$\kappa_1 = \frac{e^{\overline{z}}}{1 + e^{\overline{z}}}, \tag{233}$$

where $\overline{z}$ is the long-run mean of the log price-consumption ratio.

Given the log-linearization approximation (231) – (233), we can search the equilibrium log consumption-wealth ratio characterized by

$$z_t = A_0 + A_1 x_t + A_2 \sigma_t^2, \tag{234}$$

where the constants A_0, A_1 and A_2 are to be determined by the equilibrium conditions.

Thus, the log return on the consumption claim can be written as

$$r_{c,t+1} = \kappa_0 + \kappa_1 \left(A_0 + A_1 x_{t+1} + A_2 \sigma_{t+1}^2\right) + \Delta c_{t+1} - \left(A_0 + A_1 x_t + A_2 \sigma_t^2\right). \tag{235}$$

Therefore, the log SDF can be re-written in terms of state variables and exogenous shocks

$$m_{t+1} = \Gamma_0 + \Gamma_1 x_t + \Gamma_2 \sigma_t^2 - \lambda_c \sigma_t \epsilon_{c,t+1} - \lambda_x \sigma_t \varphi_x \epsilon_{x,t+1} - \lambda_\sigma \sigma_w \epsilon_{\sigma,t+1}, \tag{236}$$

where predictive coefficients are

$$\Gamma_0 = \log \delta - \psi^{-1} \mu_c - \frac{1}{2} \vartheta (\vartheta - 1) \left(\kappa_1 A_2 \sigma_w\right)^2, \tag{237}$$

$$\Gamma_1 = -\psi^{-1}, \tag{238}$$

$$\Gamma_2 = (\vartheta - 1)(\kappa_1 \nu - 1) A_2 = \frac{1}{2}(\gamma - 1)(\psi^{-1} - \gamma) \left[1 + \left(\frac{\kappa_1 \varphi_x}{1 - \kappa_1 \rho}\right)^2\right], \tag{239}$$

and the market price of risk coefficients are

$$\lambda_c = \gamma, \tag{240}$$

$$\lambda_x = \left(\gamma - \psi^{-1}\right) \frac{\kappa_1 \varphi_x}{1 - \kappa_1 \rho}, \tag{241}$$

$$\lambda_\sigma = -(\gamma - 1) \left(\gamma - \psi^{-1}\right) \frac{\kappa_1}{2(1 - \kappa_1 \nu)} \left[1 + \left(\frac{\kappa_1 \varphi_x}{1 - \kappa_1 \rho}\right)^2\right]. \tag{242}$$

The coefficients A_j's are determined by the equilibrium condition (i.e., the Euler equation for price of consumption claim) as follows:

$$1 = \mathbb{E}_t \left[M_{t+1} R_{c,t+1}\right] = \mathbb{E}_t \left[e^{m_{t+1} + r_{c,t+1}}\right]. \tag{243}$$

It leads to the equilibrium conditions:

$$A_0 = \frac{1}{1-\kappa_1}\left[\log\delta + \kappa_0 + \left(1-\psi^{-1}\right)\mu_c + \kappa_1 A_2(1-\nu)\overline{\sigma}^2 + \frac{\vartheta}{2}\left(\kappa_1 A_2 \sigma_w\right)^2\right], \tag{244}$$

$$A_1 = \frac{1-\psi^{-1}}{1-\kappa_1\rho}, \tag{245}$$

$$A_2 = -\frac{(\gamma-1)(1-\psi^{-1})}{2(1-\kappa_1\nu)}\left[1+\left(\frac{\kappa_1\varphi_x}{1-\kappa_1\rho}\right)^2\right]. \tag{246}$$

The long-run mean $\overline{z}$ is also determined endogenously in the equilibrium. More precisely, given all parameters fixed, we have $A_j = A_j(\overline{z})$ in Equations (244) – (246) because κ_0 and κ_1 are functions of $\overline{z}$. In the long-run steady state, we have

$$\overline{z} = A_0(\overline{z}) + A_2(\overline{z})\overline{\sigma}^2. \tag{247}$$

Thus, in the equilibrium, the long-run mean $\overline{z}$ is a function of all parameters in the model, according to (247) and the Implicit Function Theorem,

$$\overline{z} = \overline{z}\left(\mu_c, \rho, \varphi_x, \overline{\sigma}^2, \nu, \sigma_w, \cdots\right). \tag{248}$$

And hence, based on equation (248), we can also solve out $\kappa_0 = \kappa_0(\mu_c, \rho, \varphi_x, \overline{\sigma}^2, \nu, \sigma_w, \cdots)$ and $\kappa_1 = \kappa_1(\mu_c, \rho, \varphi_x, \overline{\sigma}^2, \nu, \sigma_w, \cdots)$, whose explicit forms are usually not available. The gradients κ_0 and κ_1 with respect to the parameters, such as ρ and ν, can be calculated using the Implicit Function Theorem in (247).

We specify the joint distribution of the exogenous state variables and the log dividend growth Δd_t, these joint distributional assumptions are part of the structural component of the model. More precisely, we assume that the log dividend growth process is

$$\Delta d_{t+1} = \mu_d + \phi_d x_t + \varphi_{d,c}\sigma_t\epsilon_{c,t+1} + \varphi_{d,d}\sigma_t\epsilon_{d,t+1}. \tag{249}$$

Market Equity Return. Using the Campbell-Shiller decomposition and linearization, we can represent the return in terms of log price-dividend ratio and log dividend growth:

$$r_{m,t+1} = \kappa_{m,0} + \kappa_{m,1} z_{m,t+1} + \Delta d_{t+1} - z_{m,t}, \tag{250}$$

where

$$\kappa_{m,0} = \log(1+e^{\overline{z}_m}) - \kappa_{m,1}\overline{z}_m, \tag{251}$$

and

$$\kappa_{m,1} = \frac{e^{\overline{z}_m}}{1+e^{\overline{z}_m}}, \tag{252}$$

and $\overline{z}_m$ is long-run mean of market log price-dividend ratio. We search for the equilibrium where the log market price-dividend ratio is a linear function of the states in the following form:

$$z_{m,t} = A_{m,0} + A_{m,1} x_t + A_{m,2}\sigma_t^2, \tag{253}$$

where the constants $A_{m,0}, A_{m,1}$ and $A_{m,2}$ are to be determined by equilibrium condition (i.e., Euler equation for market equity returns). Thus, we have

$$\begin{aligned} r_{m,t+1} - \mathbb{E}_t\left[r_{m,t+1}\right] = \varphi_{d,c}\sigma_t\epsilon_{c,t+1} + \kappa_{m,1}A_{m,1}\varphi_x\sigma_t\epsilon_{x,t+1} \\ + \kappa_{m,1}A_{m,2}\sigma_w\epsilon_{\sigma,t+1} + \varphi_{d,d}\sigma_t\epsilon_{d,t+1}, \end{aligned} \tag{254}$$

where

$$\mathbb{E}_t\left[r_{m,t+1}\right] = \mu_d + \kappa_{m,0} + (\kappa_{m,1}-1)A_{m,0} + \kappa_{m,1}A_{m,2}(1-\nu)\overline{\sigma}^2 \tag{255}$$

$$+ \left[\phi_d + (\kappa_{m,1}\rho - 1)A_{m,1}\right]x_t + (\kappa_{m,1}\nu - 1)A_{m,2}\sigma_t^2. \tag{256}$$

Plugging the equation above into the following Euler equation:

$$1 = \mathbb{E}_t\left[e^{m_{t+1}+r_{m,t+1}}\right], \tag{257}$$

we can derive the coefficients,

$$A_{m,0} = \frac{1}{1-\kappa_{m,1}}\left[\Gamma_0 + \kappa_{m,0} + \mu_d + \frac{1}{2}\sigma_{d,u}^2 + \kappa_{m,1}A_{m,2}(1-\nu)\overline{\sigma}^2 + \frac{1}{2}(\kappa_{m,1}A_{m,2}-\lambda_w)^2\sigma_w^2\right],$$

$$A_{m,1} = \frac{\phi_d - \psi_{\mathrm{L}}^{-1}}{1-\kappa_{m,1}\rho}, \tag{258}$$

and

$$A_{m,2} = \frac{1}{1-\kappa_{m,1}\nu}\left[\Gamma_2 + \frac{1}{2}\left(\varphi_{d,d}^2 + (\varphi_{d,c}-\lambda_c)^2 + (\kappa_{m,1}A_{m,1}\varphi_x - \lambda_x)^2\right)\right]. \tag{259}$$

Taken together, according to (254), the log return of the market portfolio can be re-written as the following beta representation for the priced aggregate shocks:

$$r_{m,t+1} - \mathbb{E}_t\left[r_{m,t+1}\right] = \beta_c\sigma_t\epsilon_{c,t+1} + \beta_x\sigma_t\epsilon_{x,t+1} + \beta_\sigma\sigma_w\epsilon_{\epsilon,t+1} + \varphi_{d,d}\sigma_t\epsilon_{d,t+1}, \tag{260}$$

where the equilibrium betas are

$$\beta_c = \varphi_{d,c}, \quad \beta_x = \kappa_{m,1}A_{m,1}\varphi_x, \quad \text{and} \quad \beta_\sigma = \kappa_{m,1}A_{m,2}. \tag{261}$$

Excess Market Return and Equity Premium. The Euler equations for the market equity return and risk-free rate can be written in one equation

$$\mathbb{E}_t\left[e^{m_{t+1}}\right] = \mathbb{E}_t\left[e^{m_{t+1}+r_{m,t+1}^e}\right]. \tag{262}$$

The risk premium is given by the beta pricing rule:

$$\mathbb{E}_t\left[r_{m,t+1}^e\right] = \lambda_c\sigma_t^2\beta_c + \lambda_x\sigma_t^2\beta_x + \lambda_\sigma\sigma_w^2\beta_\sigma - \frac{1}{2}\sigma_{r_m,t}^2, \tag{263}$$

$$\text{where} \quad \sigma_{r_m,t}^2 = \beta_c^2\sigma_t^2 + \beta_x^2\sigma_t^2 + \beta_\sigma^2\sigma_w^2 + \varphi_{d,d}^2\sigma_t^2. \tag{264}$$

Similarly, the long-run mean of log market price-dividend ratio is

$$\overline{z}_m = A_{m,0}(\overline{z}_m) + A_{m,2}(\overline{z}_m)\overline{\sigma}^2. \tag{265}$$

Based on equation (260), the excess log return of the market portfolio, defined by $r^e_{m,t+1} = r_{m,t+1} - r_{f,t}$, has the following expression:

$$r^e_{m,t+1} - \mathbb{E}_t\left[r^e_{m,t+1}\right] = \beta_c\sigma_t\epsilon_{c,t+1} + \beta_x\sigma_t\epsilon_{x,t+1} + \beta_\sigma\sigma_w\epsilon_{\epsilon,t+1} + \varphi_{d,d}\sigma_t\epsilon_{d,t+1}. \tag{266}$$

Therefore, the equilibrium excess log return of the market portfolio follows the dynamics below:

$$r^e_{m,t+1} = \mu^e_{r,t} + \beta_c\sigma_t\epsilon_{c,t+1} + \beta_x\sigma_t\epsilon_{x,t+1} + \beta_\sigma\sigma_w\epsilon_{\sigma,t+1} + \varphi_{d,d}\sigma_t\epsilon_{d,t+1}, \tag{267}$$

where $\mu^e_{r,t} = \lambda_c\beta_c\sigma^2_t + \lambda_x\beta_x\sigma^2_t + \lambda_\sigma\beta_\sigma\sigma^2_w - \frac{1}{2}\left(\beta^2_c\sigma^2_t + \beta^2_x\sigma^2_t + \beta^2_\sigma\sigma^2_w + \varphi^2_{d,d}\sigma^2_t\right)$. To avoid the stochastic singularity, we assume that the underlying marginal distribution of $(\Delta c_{t+1}, x_t, \sigma^2_t, \Delta d_{t+1})$ has some features not captured by the structural model $\mathcal{Q}$. More precisely, we assume that the excess log return's true distribution is characterized by

$$r^e_{m,t+1} = \mu^e_{r,t} + \beta_c\sigma_t\epsilon_{c,t+1} + \beta_x\sigma_t\epsilon_{x,t+1} + \beta_\sigma\sigma_w\epsilon_{\sigma,t+1} + \varphi_{d,d}\sigma_t\epsilon_{d,t+1} + \varphi_r\sigma_t\epsilon_{r,t+1}, \tag{268}$$

which augments the characterization in (267) by adding a normal shock $\varphi_r\sigma_t\epsilon_{r,t+1}$.

8.2 Generalized Methods of Moments

The likelihood function of the baseline statistical model can be seen clearly below when re-arranging the terms:

$$\frac{\Delta c_{t+1} - \mu_c - x_t}{\sigma_t} = \epsilon_{c,t+1}, \tag{269a}$$

$$\frac{x_{t+1} - \rho x_t}{\varphi_x\sigma_t} = \epsilon_{x,t+1}, \tag{269b}$$

$$\text{and} \quad \frac{(\sigma^2_{t+1} - \overline{\sigma}^2) - \nu(\sigma^2_t - \overline{\sigma}^2)}{\sigma_w} = \epsilon_{\sigma,t+1}, \tag{269c}$$

where $\epsilon_{c,t}$, $\epsilon_{x,t}$ and $\epsilon_{\sigma,t}$ are i.i.d. standard normal variables and they are mutually independent. The dividend growth process is

$$\Delta d_{t+1} = \mu_d + \phi_d x_t + \varphi_{d,c}\left(\Delta c_{t+1} - \mu_c - x_t\right) + \varphi_{d,d}\sigma_t\epsilon_{d,t+1}. \tag{270}$$

We consider the GMM setup where the baseline moments functions are identical to the score functions of the likelihood function. We denote the set of baseline moment functions by $m^{(1)}(\Delta c_{t+1}, x_{t+1}, x_t, \sigma^2_{t+1}, \sigma^2_t, \Delta d_{t+1}; \theta^{(1)})$, which includes ten moment conditions. The moment conditions that only involve Δc_{t+1}, x_t, and σ^2_t are the following six baseline moment conditions:

$$\mathbb{E}\left[\frac{\Delta c_{t+1}-\mu_c-x_t}{\sigma_t^2}\right]=0$$
$$\mathbb{E}\left[\frac{(x_{t+1}-\rho x_t)x_t}{\varphi_x^2\sigma_t^2}\right]=0$$
$$\mathbb{E}\left[\frac{(x_{t+1}-\rho x_t)^2}{\varphi_x^2\sigma_t^2}-1\right]=0$$
$$\mathbb{E}\left[\frac{\left[(\sigma_{t+1}^2-\overline{\sigma}^2)-\nu(\sigma_t^2-\overline{\sigma}^2)\right](\sigma_t^2-\overline{\sigma}^2)}{\sigma_w^2}\right]=0$$
$$\mathbb{E}\left[\frac{\left[(\sigma_{t+1}^2-\overline{\sigma}^2)-\nu(\sigma_t^2-\overline{\sigma}^2)\right]^2}{\sigma_w^2}-1\right]=0$$
$$\mathbb{E}\left[\frac{(\sigma_{t+1}^2-\overline{\sigma}^2)-\nu(\sigma_t^2-\overline{\sigma}^2)}{\sigma_w^2}\right]=0.$$

The six baseline moment conditions above captures the distribution characterized by (269a) – (269c). The joint distribution of fundamental variables $(\Delta c_{t+1}, x_t, \sigma_t^2)$ and dividend growth Δd_{t+1} is captured by the following four additional baseline moment conditions:

$$\mathbb{E}\left[\frac{\Delta d_{t+1}-\mu_d-\phi_d x_t-\varphi_{d,c}(\Delta c_{t+1}-\mu_c-x_t)}{\varphi_{d,d}^2\sigma_t^2}\right]=0$$
$$\mathbb{E}\left[\frac{x_t\left[\Delta d_{t+1}-\mu_d-\phi_d x_t-\varphi_{d,c}(\Delta c_{t+1}-\mu_c-x_t)\right]}{\varphi_{d,d}^2\sigma_t^2}\right]=0$$
$$\mathbb{E}\left[\frac{(\Delta c_{t+1}-\mu_c-x_t)\left[\Delta d_{t+1}-\mu_d-\phi_d x_t-\varphi_{d,c}(\Delta c_{t+1}-\mu_c-x_t)\right]}{\varphi_{d,d}^2\sigma_t^2}\right]=0$$
$$\mathbb{E}\left[\frac{\left[\Delta d_{t+1}-\mu_d-\phi_d x_t-\varphi_{d,c}(\Delta c_{t+1}-\mu_c-x_t)\right]^2}{\varphi_{d,d}^2\sigma_t^2}-1\right]=0.$$

In the long-run risk model, the primary goal is to understand how the excess log return of the market portfolio is affected by the consumption process and dividend process specified in (269a) – (269c) and (270). The joint distribution of the excess log return $r_{m,t+1}^e$, the consumption variables, and the dividend variables can be seen clearly from the following formula:

$$\begin{aligned}\varphi_r\sigma_t\epsilon_{r,t+1}=r_{m,t+1}^e-\mu_{r,t}^e-(\beta_c-\varphi_{d,c})(\Delta c_{t+1}-\mu_c-x_t)-\beta_x\frac{x_{t+1}-\rho x_t}{\varphi_x}\\ -\beta_\sigma\left[\widehat{\sigma}_{t+1}^2-\nu\widehat{\sigma}_t^2\right]-(\Delta d_{t+1}-\mu_d-\phi_d x_t)\,,\end{aligned}\tag{271}$$

where $\widehat{\sigma}_t^2\equiv\sigma_t^2-\overline{\sigma}^2$ and

$$\mu_{r,t}^e=\lambda_\eta\beta_\eta\sigma_t^2+\lambda_e\beta_e\sigma_t^2+\lambda_w\beta_w\sigma_w^2-\frac{1}{2}\left(\beta_\eta^2\sigma_t^2+\beta_e^2\sigma_t^2+\beta_w^2\sigma_w^2+\varphi_{d,d}^2\sigma_t^2\right).\tag{272}$$

Because $\beta_c = \varphi_{d,c}$, equation (271) can be rewritten as

$$\varphi_r \sigma_t \epsilon_{r,t+1} = r^e_{m,t+1} - \mu^e_{r,t} - \beta_x \frac{x_{t+1} - \rho x_t}{\varphi_x} - \beta_\sigma \left[\widehat{\sigma}^2_{t+1} - \nu \widehat{\sigma}^2_t\right] - (\Delta d_{t+1} - \mu_d - \phi_d x_t). \tag{273}$$

We choose the asset pricing cross-equation moments, denoted by $m^{(2)}(\Delta c_{t+1}, x_{t+1}, x_t, \sigma^2_{t+1}, \sigma^2_t, \Delta d_{t+1}, r^e_{m,t+1}; \theta)$, to include the score functions of the conditional likelihood of $r^e_{m,t+1}$ above. Thus, the moment conditions for the optimal GMM setup to assess the fragility of the benchmark version of long-run risk model are

$$m(\Delta c_{t+1}, x_{t+1}, x_t, \sigma^2_{t+1}, \sigma^2_t, r^e_{t+1}, \Delta d_{t+1}, \theta) \equiv \begin{bmatrix} m^{(1)}(\Delta c_{t+1}, x_{t+1}, x_t, \sigma^2_{t+1}, \sigma^2_t, \Delta d_{t+1}, \theta^{(1)}) \\ m^{(2)}(\Delta c_{t+1}, x_{t+1}, x_t, \sigma^2_{t+1}, \sigma^2_t, \Delta d_{t+1}, r^e_{m,t+1}, \theta) \end{bmatrix}.$$

Intuitively, the over-identification moment conditions imposed by the long-run risk model on the dynamic parameter θ is through the cross-equation restrictions on the beta coefficients $\beta_c, \beta_x, \beta_\sigma$ and the pricing coefficients $\lambda_c, \lambda_x, \lambda_\sigma$. Because the shocks $\epsilon_{c,t+1}$, $\epsilon_{x,t+1}$, $\epsilon_{\sigma,t+1}$, and $\epsilon_{r,t+1}$ are mutually independent, the GMM setup is actually first-order asymptotically equivalent to the MLE for the joint distribution of $(\Delta c_{t+1}, x_t, \sigma^2_t, \Delta d_{t+1}, r^e_{m,t+1})$. It should be noted that the whole joint distribution of the variables, including $(\Delta c_{t+1}, x_t, \sigma^2_t, \Delta d_{t+1}, r^e_{m,t+1})$ and many other variables such as price-dividend ratios, may have more stochastic singularities and many features that are not the targets of the long-run risk model to explain at the first place. Following the spirits of GMM-based estimation and hypothesis testing for structural models, we focus on the moments targeted by a particular long-run risk model.

The analytical formulas for the over-identification moment conditions are quite complicated, since how the beta coefficients and market price of risk coefficients depend on model parameters in equilibrium is very complicated for the long-run risk model. We ignore the formulas here and, in fact, we calculate them numerically in obtaining the fragility measures. Moreover, we compute the information matrices for the baseline and full moment conditions based on simulated stationary time series.

References

Abel, A. B., 1999, "Risk premia and term premia in general equilibrium," *Journal of Monetary Economics*, 43, 3 – 33.

Andrews, D. W. K., 1993, "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

Bansal, R., D. Kiku, and A. Yaron, 2012, "An Empirical Evaluation of the Long-Run Risks Model for Asset Prices," *Critical Finance Review*, 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, 1481–1509.

Bickel, P. J., 1981, "Quelques aspects de la statistique robuste," in *Ninth Saint Flour Probability Summer School—1979 (Saint Flour, 1979)*, vol. 876 of *Lecture Notes in Math.*, . pp. 1–72, Springer, Berlin-New York.

Bickel, P. J., C. A. J. Klaassen, Y. Ritov, and J. A. Wellner, 1993, *Efficient and adaptive estimation for semiparametric models* . Johns Hopkins Series in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD.

Bradley, R. C., 2005, "Basic properties of strong mixing conditions. A survey and some open questions," *Probab. Surv.*, 2, 107–144, Update of, and a supplement to, the 1986 original.

Chamberlain, G., 1987, "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 34, 305–334.

Chen, H., W. W. Dou, and L. Kogan, 2021, "Supplement to "Measuring the 'Dark Matter' in Asset Pricing Models"," Discussion paper.

Chen, X., and A. Santos, 2018, "Overidentification in Regular Models," *Econometrica*, 86, 1771–1817.

Cheng, X., W. W. Dou, and Z. Liao, 2021, "Macro-Finance Decoupling: Robust Evaluations of Macro Asset Pricing Models," Working papers, University of Pennsylvania.

Dou, W., D. Pollard, and H. H. Zhou, 2010, "Functional Regression for General Exponential Families," *Submitted to Annals of Statistics.*

Gordin, M. I., 1969, "The central limit theorem for stationary processes," *Dokl. Akad. Nauk SSSR*, 188, 739–741.

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

Greenwood, P. E., and W. Wefelmeyer, 1995, "Efficiency of Empirical Estimators for Markov Chains," *Ann. Statist.*, 23, 132–143.

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

Hansen, L. P., 1985, "A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators," *Journal of Econometrics*, 30, 203–238.

Hansen, L. P., 2007, "Generalized Method of Moments Estimation," in *The New Palgrave Dictionary of Economics* . edited by Steven N. Durlauf and Lawrence E. Blume, Palgrave Macmillan.

Hansen, L. P., 2012, "Proofs for large sample properties of generalized method of moments estimators," *Journal of Econometrics*, 170, 325 – 330, Thirtieth Anniversary of Generalized Method of Moments.

Jones, G. L., 2004, "On the Markov chain central limit theorem," *Probab. Surv.*, 1, 299–320.

Kitamura, Y., T. Otsu, and K. Evdokimov, 2013, "Robustness, Infinitesimal Neighborhoods, and Moment Restrictions," *Econometrica*, 81, 1185–1201.

Le Cam, L., and G. L. Yang, 2000, *Asymptotics in statistics* . Springer Series in Statistics, Springer-Verlag, New York, second edn., Some basic concepts.

Levit, B. Y., 1976, "On the Efficiency of a Class of Non-Parametric Estimates," *Theory of Probability & Its Applications*, 20, 723–740.

Li, H., and U. K. Müller, 2009, "Valid Inference in Partially Unstable Generalized Method of Moments Models," *Review of Economic Studies*, 76, 343–365.

McLeish, D. L., 1975a, "Invariance principles for dependent variables," *Wahrscheinlichkeitstheorie verw Gebiete*, 32, 165–178.

McLeish, D. L., 1975b, "A maximal inequality and dependent strong laws," *Ann. Probability*, 3, 829–839.

Meyn, S., and R. L. Tweedie, 2009, *Markov chains and stochastic stability*, Cambridge University Press, Cambridge, second edn., With a prologue by Peter W. Glynn.

Nevelson, M. B., 1977, "On One Informational Lower Bound," *Problemy Peredachi Informatsii*, 13, 26–31.

Newey, W. K., 1985, "Generalized method of moments specification testing," *Journal of Econometrics*, 29, 229–256.

Phillips, P. C. B., and S. N. Durlauf, 1986, "Multiple Time Series Regression with Integrated Processes," *The Review of Economic Studies*, 53, 473–495.

Severini, T. A., and G. Tripathi, 2013, "Semiparametric Efficiency Bounds for Microeconometric Models: A Survey," *Foundations and Trends in Econometrics*, 6, 163–397.

Sowell, F., 1996, "Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework," *Econometrica*, 64, 1085–1107.

van der Vaart, A. W., 1988, *Statistical estimation in large parameter spaces* . , vol. 44 of *CWI Tract*, Stichting Mathematisch Centrum, Centrum voor Wiskunde en Informatica, Amsterdam.

van der Vaart, A. W., 1998, *Asymptotic Statistics* . , vol. 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, Cambridge.

van der Vaart, A. W., and J. A. Wellner, 1996, *Weak convergence and empirical processes* . Springer Series in Statistics, Springer-Verlag, New York, With applications to statistics.

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

White, H., and I. Domowitz, 1984, "Nonlinear Regression with Dependent Observations," *Econometrica*, 52, 143–61.