NBER WORKING PAPER SERIES

MEASURING "DARK MATTER" IN ASSET PRICING MODELS

Hui Chen
Winston Wei Dou
Leonid Kogan

Working Paper 26418
http://www.nber.org/papers/w26418

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2019

We thank Alon Brav, Bradyn Breon-Drish, Murray Carlson, Xiaohong Chen, Xu Cheng, John Cochrane, George Constantinides, Ian Dew-Becker, Frank Diebold, Stefano Giglio, Dan Greenwald, Lars Peter Hansen, Yan Ji, Karen Lewis, Ye Luo, Robert Novy-Marx, Christian Opp, Giorgio Primiceri, Tom Sargent, Frank Schorfheide, Martin Schneider, Bill Schwert, Chris Sims, Rob Stambaugh, Luke Taylor, Di Tian, Laura Veldkamp, Pietro Veronesi, S. Viswanathan, Jessica Wachter, Tan Wang, Ivo Welch, Yu Xu, Amir Yaron, Fernando Zapatero, Stan Zin, participants at the 2019 MIT Capital Markets Research Workshop, and seminar participants at University of Chicago, Chicago CITE, CKGSB, INSEAD, ITAM Finance Conference, Macrofinance Workshop, MFM Summer Session, MFM Winter Conference, MIT Sloan, Northwestern, Harvard, NYU, NBER Asset Pricing Meeting, NBER Capital Markets Meeting, Red Rock Finance Conference, UBC Winter Finance Conference, University of Pennsylvania (Economics), Utah Winter Finance Conference, WFA, Washington University (Olin), Wharton, and Yale SOM for comments. We thank Xiaoliang Wang and Kan Xu for their research assistance. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Measuring “Dark Matter” in Asset Pricing Models
Hui Chen, Winston Wei Dou, and Leonid Kogan
NBER Working Paper No. 26418
November 2019
JEL No. C52,D81,E32,G12

ABSTRACT

We introduce an information-based fragility measure for GMM models that are potentially misspecified and unstable. A large fragility measure signifies a GMM model's lack of internal refutability (weak power of specification tests) and external validity (poor out-of-sample fit). The fragility of a set of model-implied moment restrictions is tightly linked to the quantity of additional information the econometrician can obtain about the model parameters by imposing these restrictions. Our fragility measure can be computed at little cost even for complex dynamic structural models. We illustrate its applications via two models: a rare-disaster risk model and a long-run risk model.

Hui Chen
MIT Sloan School of Management
100 Main Street, E62-637
Cambridge, MA 02142
and NBER
huichen@mit.edu

Leonid Kogan
MIT Sloan School of Management
100 Main Street, E62-636
Cambridge, MA 02142
and NBER
lkogan@mit.edu

Winston Wei Dou
2318 Steinberg Hall - Dietrich Hall
The Wharton School at
University of Pennsylvania
3620 Locust Walk
Philadelphia, PA 19104
wdou@wharton.upenn.edu

1. INTRODUCTION

In cosmology, dark matter is a form of matter that is not directly observable, yet its presence is required for Einstein's theory of general relativity to be consistent with the observable motions of stars and galaxies. Certain economic models rely on an analogous form of "dark matter," namely model components or parameters that are difficult to verify and measure directly in the data, despite having significant effects on the models' performance.[1]

Should we be worried about models with dark matter features? Two common defenses for economic models relying on dark matter are: (i) these features are not inconsistent with the data, i.e., they cannot be rejected by the data; and (ii) the fact that they improve the model's ability to fit the data provides indirect evidence supporting their very existence, analogous to dark matter as posited in cosmology. In this paper, we propose a measure for economic dark matter and show that the measure helps quantify the degree of model fragility. Specifically, models with more dark matter tend to lack refutability and have higher overfitting tendency.

We define a measure of economic dark matter for a general class of GMM models (Hansen, 1982) that summarize a structural model with a set of unconditional moment restrictions. Our measure is based on the (relative) informativeness of the cross-equation restrictions imposed by the structural model. Essentially, the measure compares the asymptotic variances of two efficient GMM estimators for the model parameters θ, one based on the full set of moment restrictions which we refer to as the full GMM model, and the other based on a subset of the moment restrictions which we refer to as the baseline GMM model. The dark matter measure is obtained by searching for the largest discrepancy between the two asymptotic variances in all linear directions.

Our dark matter measure above is inspired by rational expectations econometrics, where the key assumption is that the agents in an economic model know more about model parameters than conveyed by the primitive sources inside the model. Intuitively, our dark matter measure will have a large value when the primary source of information about model parameters is the cross-equation restrictions, rather than the primitive sources in the model. In such cases, it may be challenging to argue that, as an approximation, economic agents have inferred parameters from rich histories of primitive data. Hence, the information obtained through the cross-equation restrictions becomes unreliable because it heavily relies on the validity of cross-

[1] For example, John Campbell referred to rare disasters as "dark matter for economists" in his 2008 Princeton Lectures in Finance (Campbell, 2018).

equation restrictions. In this paper, we formalize this intuition by showing what happens when highly informative restrictions are potentially misspecified. Our setting applies, but is not limited to the rational expectations models.

We show that the dark matter measure is linked to two key model properties: refutability and overfitting tendency. Specifically, we show that the power of the optimal specification tests vanishes as the dark matter measure for a GMM model approaches infinity. We also show that those GMM models with more dark matter tend to overfit the data; namely, the out-of-sample fit of the moment restrictions can deteriorate significantly relative to its in-sample fit once we take into account the possibility that the GMM model is potentially misspecified and the true data-generating process (DGP) is subject to local instability (e.g., Li and Müller, 2009).

The dark matter measure has an intuitive sample-size interpretation, which can be thought of as the amount of additional data needed for the baseline efficient GMM estimator to match or exceed the precision that the full efficient GMM estimator can achieve with the aid of the cross-equation restrictions. Technically, we extend the results on semiparametric local minimax efficiency bounds by Levit (1976), Nevelson (1977), and Chamberlain (1987, Theorem 2) for unconditional moment restrictions to Markov processes with local instability.[2] This result formalizes the information interpretation of our measure.

Model fragility is usually examined through the lens of sensitivity analysis in practice. The fact that quantifying the informativeness of cross-equation restrictions is linked to sensitivity analysis is intuitive. Cross-equation restrictions are informative under two conditions: (i) the additional moments from the full model are sensitive to small changes in the parameter values; and (ii) these additional moments can be measured precisely in the data relative to the baseline moments. Thus, computing the dark matter measure resembles the dual of conducting a form of sensitivity analysis. Intuitively, a GMM model is considered fragile if its key implications are excessively sensitive to small perturbations of the data-generating process. Formally, one needs to specify the relevant magnitude of "local perturbations" and define "excessive sensitivity." In multivariate settings, there is the additional challenge that the simultaneous responses of model parameters to the perturbations in the data-generating process must be considered in

[2] Hansen (1985) and Chamberlain (1987, Thereom 3) study semiparametric local minimax efficiency bounds for conditional moment restrictions. Hansen (1985) derives the efficiency bounds from the perspective of characterizing the optimal instrument in the estimation of generalized instrumental variables in a non-i.i.d. context. Chamberlain (1987, Theorem 3) focuses on moment restrictions parameterized in terms of a finite-dimensional vector in an i.i.d. context. Newey (1990, 1993) proposes an estimator that attains Chamberlain's bounds. Ai and Chen (2003) propose an estimation method, as well as its efficiency, for models of conditional moment restrictions, which contain finite dimensional unknown parameters and infinite dimensional unknown functions.

order to assess the full scope of model fragility. However, little work has been done to formalize the sensitivity analysis for model fragility quantification.

Our dark matter measure formalizes the sensitivity analysis by (i) benchmarking the local perturbation in the data-generating process against the identification derived from the baseline model, and (ii) defining the excessive sensitivity of efficient GMM estimators to the local perturbations in the data-generating process based on the sampling variability of the efficient GMM estimators. Naturally, we require the baseline model to be a correct benchmark similar to Eichenbaum, Hansen, and Singleton (1988), because we need the identification provided by the baseline model to define the reasonable perturbations in the data-generating process. In addition, our measure identifies the worst direction of perturbation for the multivariate setting by searching for the direction in the model parameter space in which the cross-equation restrictions are the most informative.

An intuitive justification for the concern regarding model sensitivity is the high effective degrees of freedom in models with high sensitivity. Such models can be fitted to a wide range of empirical moments with minor changes to the parameter values. Models with high degrees of freedom are well-documented as being more prone to overfitting, which is why statistical model selection procedures impose penalties for model complexity, as measured by AIC, BIC, LASSO, etc.

Formally, the concern regarding model sensitivity originates in potential misspecification and local instability of the data-generating process. Under the assumption that a model is correctly specified, high sensitivity of the moments to parameter perturbations is a beneficial feature. Imposing these model restrictions will facilitate estimating the model parameters significantly more precisely, which is the foundation for structural estimation. However, if the model restrictions are potentially misspecified, the information obtained from imposing such restrictions may no longer be valid. Kocherlakota (2007) similarly emphasizes the "fallacy of fit." The nontestable assumption of identification strength is related to the informativeness of the cross-equation restrictions. They indicate the implicit degrees of freedom postulated by the modeler, as the modeler effectively passes the statistical challenge of learning about model parameters from the data onto economic agents.

We analyze the consequences of misspecification and local instability by generalizing the local instability framework of Li and Müller (2009) to the semiparametric setting.[3] We show

[3]We need a semiparametric framework for at least three reasons: (i) it provides a formal general econometric framework for local perturbations in the space of local data-generating processes; (ii) it is needed for justifying

that models with large dark matter measures tend to generate an excessively high quality of in-sample fit. Due to this finding, these models are difficult to reject even when they are misspecified, hence the lack of refutability. Moreover, we show that under local instability, models with larger dark matter measures have a higher worst-case asymptotic expected degree of overfitting, as measured by the gap between the in-sample model fit and the out-of-sample model fit based on the Sargan-Hansen J statistic. Thus, we generalize the notion of sensitivity analysis from the perturbations of model parameters to the perturbations of the underlying data-generating processes.

The recursive (two-stage) GMM estimation (e.g., Christiano and Eichenbaum, 1992; Ogaki, 1993; Newey and McFadden, 1994; Hansen and Heckman, 1996; Hansen, 2007b; Lee, 2007; Hansen, 2012) has necessarily worse in-sample fit than does the efficient GMM estimation (Hansen, 1982); however, the former can in fact deliver better out-of-sample fit when the dark matter measure (i.e., the model fragility) is excessively high. Although the original impetus of the recursive (two-stage) GMM estimation was primarily computational, we advocate it as a robust estimation procedure against high model fragility. Thorough analyses on optimal robust estimation, and even on optimal model selection, however, are beyond the scope of this paper.

We evaluate the fragility of two models from the asset pricing literature. The first example is a rare-disaster model. In this model, parameters describing the likelihood and magnitude of economic disasters are difficult to estimate from the data unless information in asset prices is used. We derive the dark matter measure in this example analytically. We also illustrate how to incorporate uncertainty about the structural parameters (in this context, preference parameters) when computing model fragility. The second example is a long-run risk model with a nine-dimensional parameter space. We use this example to show that two calibrations of the model with similar in-sample fit can differ vastly in fragility properties. We conduct Monte Carlo simulation experiments for both examples and show that the calibrated models with large dark matter measures lack in-sample refutability and have poor out-of-sample fit, consistent with the theory.

Related Literature

The idea that a model's fragility is connected to its degrees of freedom (i.e., its complexity) dates back at least to Fisher (1922). Traditionally, effective degrees of freedom of a model are

the information matrix interpretation of our dark matter measure based on semiparametric efficiency bounds; and (iii) it is a natural way to connect the GMM model to the structural economic model.

measured by the number of parameters, simply because the two coincide in Gaussian-linear models (e.g. Ye, 1998; Efron, 2004). Numerous statistical model selection procedures are based on this idea.[4]

However, the limitations of using the number of parameters to measure model's degrees of freedom have been well documented. Hence, new methods have been developed to measure the sensitivity of model implications to parameter perturbations in the statistics literature.[5] A common feature of these proposals is that they rely on the same model being evaluated to determine the parameter perturbations; this is potentially problematic when evaluating economic models that are themselves fragile, partly due to a lack of internal refutability. In contrast, we propose using a baseline model to assign weights to potential alternative underlying data-generating processes or to determine the possible "reasonable" perturbations of data-generating processes.

To be more specific, our fragility measure is different from the extant measures in four aspects. First, we use a semiparametric framework to allow for general local perturbations of data-generating processes similar to Hansen and Sargent (2001), but not only the local perturbations of model parameters that fit the model. Second, the reasonable local perturbations of data-generating processes are generated by the baseline model that is less likely to be misspecified, and we use the baseline model as a benchmark to respect the primary purpose of the economic structural model in the fragility assessment. Third, we directly connect the model fragility measure to a model's internal refutability, i.e., the optimal power of specification tests. Fourth, we also directly connect the model fragility measure to a model's out-of-sample fit, emphasized by, for example, Schorfheide and Wolpin (2012) and Athey and Imbens (2017, 2019), for assessing economic models.[6]

Further, we have built our model fragility measure based on a multivariate sensitivity analysis. Müller (2012) studies multivariate sensitivity analysis in Bayesian inference and the worst-case direction. He focuses on the sensitivity of the posterior distribution with respect to the prior

[4]Examples include the Akaike information criterion (AIC) (Akaike, 1973), the Bayesian information criterion (BIC) (Schwarz, 1978), the risk inflation criterion (RIC) (Foster and George, 1994), and the covariance inflation criterion (CIC) (Tibshirani and Knight, 1999).

[5]Extant statistics literature covers several alternative approaches to measuring the "implicit degrees of freedom" or "generalized degrees of freedom" (e.g., Ye, 1998; Shen and Ye, 2002; Efron, 2004; Spiegelhalter, Best, Carlin, and van der Linde, 2002; Ando, 2007; Gelman, Hwang, and Vehtari, 2013).

[6]The terms "in-sample fit" and "out-of-sample fit", as well as other similar terms, are used commonly in statistics, econometrics, and empirical asset pricing (e.g., Hastie, Tibshirani, and Friedman, 2001; Ferson, Nallareddy, and Xie, 2013; Varian, 2014; Athey and Imbens, 2015; Mullainathan and Spiess, 2017; Müller and Watson, 2016). The terms "out-of-sample fit" and "external validity" have been used interchangeably in finance and economics literature (e.g., Bossaerts and Hillion, 1999; Stock and Watson, 2002; Schorfheide and Wolpin, 2012).

distribution by asking how much the posterior distribution changes in response to a perturbation of the prior distribution. Differently, we perturb the underlying data-generating process, and ask how much the moment restrictions and the baseline model parameters change. Similar to Müller (2012), we consider a multivariate sensitivity problem using asymptotic methods and eigenvalue decomposition to identify the worst-case direction. In another related paper, Andrews, Gentzkow, and Shapiro (2017) propose a local measure of the relationship between parameter estimates and moments. In Section 5.2, we establish the connection between our information-based dark matter measure and their sensitivity matrix. Their focus is to add transparency in structural estimation, and they do not link the magnitude of the sensitivity matrix to model properties. We formally connect the dark matter measure to model fragility, and link its magnitude to the model's lack of refutability and overfitting tendency. Moreover, our measure differs from their sensitivity matrix in two aspects: (i) it is a relative sensitivity measure that uses the baseline GMM model as a benchmark; and (ii) it normalizes the expected change of the estimator by its asymptotic covariance in the full model.

Our work contributes to the literature on local instability in time series analysis. Evidence abounds on structural changes and nonstationarity in asset pricing (e.g., Pesaran and Timmermann, 1995; Bossaerts and Hillion, 1999; Pastor and Stambaugh, 2001; Lettau and Van Nieuwerburgh, 2008; Lettau, Ludvigson, and Wachter, 2008; Welch and Goyal, 2008; Koijen and Van Nieuwerburgh, 2011; Dangl and Halling, 2012). Econometric theory has largely focused on testing whether or not the model is stable (see, e.g., Nyblom, 1989; Andrews, 1993; Andrews and Ploberger, 1994; Sowell, 1996; Bai and Perron, 1998; Hansen, 2000; Andrews, 2003; Elliott and Müller, 2006, for recent contributions). However, little research has explored the next step: what implications arise once instabilities are suspected? One exception is Li and Müller (2009) who show that the standard GMM inference (Hansen, 1982), despite ignoring the partial instability of a subset of model parameters, remains asymptotically valid for the subset of stable parameters. We show that the GMM models tend to have poor out-of-sample fit if their dark matter measure, i.e., model fragility measure, is excessively large.

Our work connects to the literature on structural estimation, including rational expectations econometrics, in which economic assumptions (the cross-equation restrictions) have been used extensively to increase efficiency in estimating the structural parameters. Classic examples include Saracoglu and Sargent (1978), Hansen and Sargent (1980), Campbell and Shiller (1988), among others, and textbook treatments by Lucas and Sargent (1981), Hansen and Sargent (1991). In a fragile model, cross-equation restrictions may imply excessively tight confidence regions for the parameters, with low coverage probability under reasonable parameter perturbations. An important potential source of fragility in this context is that the structural model relies heavily on the agents possessing accurate knowledge of hard-to-estimate parameters.

Hansen (2007a) offers an extensive discussion of the informational burden that rational expectations models place on the agents, which is one of the key motivations for research in Bayesian learning, model ambiguity, and robustness (e.g., Gilboa and Schmeidler, 1989; Hansen and Sargent, 2001; Epstein and Schneider, 2003; Klibanoff, Marinacci, and Mukerji, 2005). This literature recognizes that the traditional assumption that agents possess precise knowledge of the relevant probability distributions is not justifiable in certain contexts, and explicitly incorporates robustness considerations into agents' decision problems. Our approach complements this line of research, in that our measure of fragility helps diagnose situations in which incorporating parameter uncertainty and agents' robustness considerations within an economic model could be particularly important.

Our analysis of the disaster-risk model relates to studies that have highlighted the challenges in testing such models. One implication of the low probability of disasters is the so-called "peso problem" (see Lewis, 2008, for an overview): if observations of disasters in a particular sample under-represent their population distribution, standard inference procedures may lead to distorted conclusions. Thus, the peso problem is a particular case of the weak identification problem. Our analysis highlights that in applications subject to the peso problem, it is important to guard against model fragility. On this front, Zin (2002) shows that certain specifications of higher-order moments of the endowment growth distribution may help the model fit the asset pricing moments while being difficult to reject in the endowment data. Our analysis of model fragility encapsulates such considerations in a general quantitative measure.

2. AN INTUITIVE EXAMPLE

In this section, we use a version of the Gordon growth model to illustrate how the dark matter measure connects to model fragility. Suppose the dividend process for a stock is

$$Y_{t+1}/Y_t = 1 + \theta + \sigma_Y \epsilon_{Y,t+1}, \qquad \epsilon_{Y,t} \text{ is i.i.d., with } \mathrm{E}[\epsilon_{Y,t}] = 0,\ \mathrm{E}[\epsilon_{Y,t}^2] = 1. \tag{1}$$

The parameters θ and σ_Y are the mean and volatility of dividend growth. According to the Gordon growth model, the price of the stock is the present value of expected future dividends.

Assuming the risk-adjusted discount rate is r, then

$$P_t = \sum_{s=1}^{\infty} \mathrm{E}_t\left[Y_{t+s}\right]/(1+r)^s, \tag{2}$$

which implies a constant price-dividend ratio,

$$P_t/Y_t = F(\theta), \quad \text{with } F(\theta) \equiv (1+\theta)/(r-\theta). \tag{3}$$

The econometrician evaluates a GMM version of this model in a sample of size n. To avoid the stochastic singularity, we add i.i.d. shocks to the price-dividend ratio such that (3) only holds on average,

$$P_{t+1}/Y_{t+1} = F(\theta) + \sigma_P \epsilon_{P,t+1}, \qquad \epsilon_{P,t} \text{ is i.i.d., with } \mathrm{E}[\epsilon_{P,t}] = 0,\ \mathrm{E}[\epsilon_{P,t}^2] = 1. \tag{4}$$

Moreover, $\epsilon_{P,t}$ and $\epsilon_{Y,t}$ are mutually independent. For simplicity, we assume that the econometrician knows all the parameters except for average dividend growth θ and focuses on the following moment conditions:

$$\mathbb{E}\left[m(\mathbf{y}_t,\theta)\right] = 0, \quad \text{with } m(\mathbf{y}_t,\theta) \equiv \begin{bmatrix} Y_{t+1}/Y_t - 1 - \theta \\ P_t/Y_t - F(\theta) \end{bmatrix}, \tag{5}$$

where $\mathbf{y}_t \equiv (Y_t, P_t)^T$. We denote the first element of $m(\mathbf{y}_t,\theta)$ by $m^{(1)}(\mathbf{y}_t,\theta)$ and refer to it as the baseline moment.

Next, We assume that the Gordon growth model can be misspecified and that the true local data-generating processes is

$$\begin{bmatrix} Y_{t+1}/Y_t \\ P_{t+1}/Y_{t+1} \end{bmatrix} = \frac{f_{n,t}}{\sqrt{n}} + \begin{bmatrix} 1+\theta_0 \\ F(\theta_0) \end{bmatrix} + \begin{bmatrix} \sigma_Y \epsilon_{Y,t+1} \\ \sigma_P \epsilon_{P,t+1} \end{bmatrix}. \tag{6}$$

The term $f_{n,t}$ captures the potential local bias and instability:

$$f_{n,t} = \lambda_1 + \lambda_2 b(t/n), \quad \text{with } \lambda_i \equiv \left[\lambda_i^{(1)},\ \lambda_i^{(2)}\right]^T \text{ for } i \in \{1,2\}, \tag{7}$$

where $b(\cdot)$ is an unknown deterministic function on $[0,1]$ whose path has a finite number of discontinuities and one-sided limits everywhere. Without loss of generality, we assume that

$\sup_{u\in[0,1]}|b(u)|\leq 1$ and $\int_0^1 b(u)\mathrm{d}u = 0$.

Instability in the data-generating process, for example, structural breaks and nonstationarity, is an important consideration in asset pricing.[7] Our specification in (7) follows the the literature on local instability (e.g., Andrews, 1993; Sowell, 1996; Li and Müller, 2009). Intuitively, λ_1 captures the stable local biases in the moments, while λ_2 captures their local instability.

For calibration, we set $\theta_0 = 0$, $r = 3\%$, $\sigma_Y = 4\%$, and $\sigma_P = 5$. Under the calibrated distribution Q_0 of $\mathbf{y}_t$, which corresponds to the parameter value $\theta_0 = 0$, the Jacobian matrices of the baseline and the full moment restrictions are

$$D_{11} = \mathrm{E}\left[\frac{\partial m^{(1)}(\mathbf{y}_t,\theta)}{\partial\theta}\right] = -1 \quad\text{and}\quad D = \mathrm{E}\left[\frac{\partial m(\mathbf{y}_t,\theta)}{\partial\theta}\right] = \begin{bmatrix} -1 \\ (1+r)/r^2 \end{bmatrix}$$

The spectral density matrices (at zero frequency) for the baseline and full models are

$$\Omega_{11} = \mathrm{E}\left[m^{(1)}(\mathbf{y}_t,\theta_0)^2\right] = \sigma_Y^2 \quad\text{and}\quad \Omega = \mathrm{E}\left[m(\mathbf{y}_t,\theta_0)m(\mathbf{y}_t,\theta_0)^T\right] = \begin{bmatrix} \sigma_Y^2 & 0 \\ 0 & \sigma_P^2 \end{bmatrix}.$$

The growth rate θ can be estimated using either the baseline moment restrictions alone or using the full moment restrictions. We denote these two efficient GMM estimators for θ by $\tilde{\theta}_n$ and $\hat{\theta}_n$, which minimize the objectives

$$J^{(1)}(\theta^{(1)},\mathbf{y}^n) = n\widehat{m}_n^{(1)}(\theta)^T\Omega_{11}^{-1}\widehat{m}_n^{(1)}(\theta) \text{ and } J(\theta,\mathbf{y}^n) = n\widehat{m}_n(\theta)^T\Omega^{-1}\widehat{m}_n(\theta), \text{ respectively,}$$

where $\widehat{m}_n^{(1)}(\theta)$ and $\widehat{m}_n(\theta)$ are the sample means for $m^{(1)}(\mathbf{y}_t,\theta)$ and $m(\mathbf{y}_t,\theta)$, respectively.

The dark matter measure

Panel A of Figure 1 displays the asymptotic distribution of the baseline estimator $\tilde{\theta}_n$ (the efficient GMM estimator for θ based on the baseline moment of (5)),

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) \xrightarrow{d} N(0, \mathbf{I}_{\mathrm{B}}^{-1}), \quad\text{where}\quad \mathbf{I}_{\mathrm{B}} \equiv D_{11}^T\Omega_{11}^{-1}D_{11}.$$

[7] See e.g., Pesaran and Timmermann (1995); Pastor and Stambaugh (2001); Lettau, Ludvigson, and Wachter (2008); Lettau and Van Nieuwerburgh (2008); Welch and Goyal (2008); Koijen and Van Nieuwerburgh (2011).

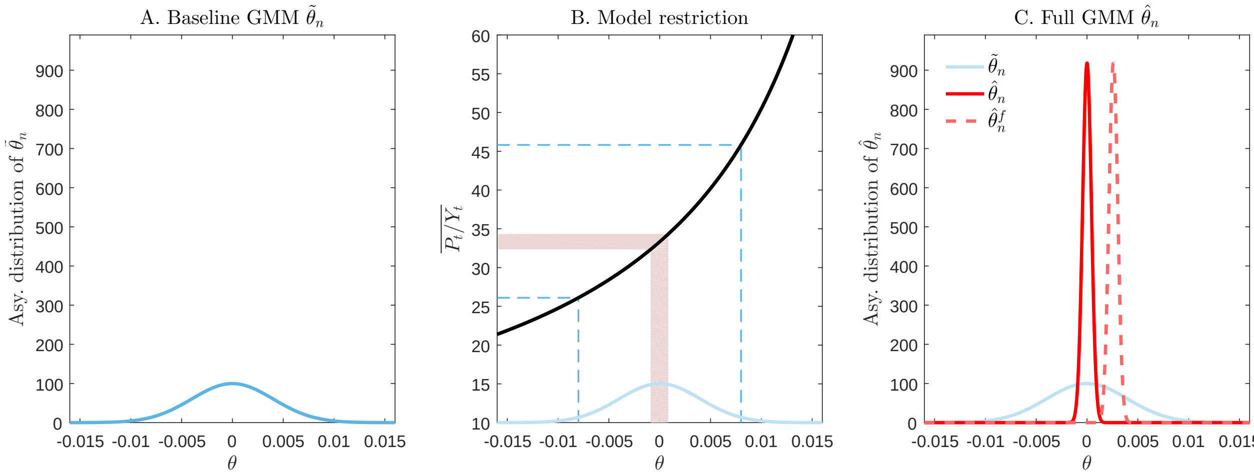

FIGURE 1.— An example of an informative asset pricing restriction on the parameters. $\tilde{\theta}_n$ is the efficient GMM estimator only based on the baseline GMM model characterized by the baseline moment $m^{(1)}(\mathbf{y}_t, \theta)$ and sample $\mathbf{y}^n$. Both $\hat{\theta}_n$ and $\hat{\theta}_n^f$ are the efficient GMM estimators based on the full GMM model and sample $\mathbf{y}^n$, except that $\hat{\theta}_n^f$ is the estimator when the model is misspecified ($f_{n,t} \equiv 0$). Panel A plots the asymptotic distribution of $\sqrt{n}(\tilde{\theta}_n - \theta_0)$ with $n = 100$. Panel B plots the average price-dividend ratio as a function of θ. The light blue dashed vertical lines represent the confidence band for $\tilde{\theta}_n$ with sample size $n = 100$, and the corresponding light blue dashed horizontal lines represent the model-implied average price-dividend ratio when perturbing θ within the confidence band. The red shaded horizontal area represents the confidence band of the average price-dividend ratio $\overline{P_t/Y_t}$ according to the data with $n = 100$, and the corresponding red shaded vertical area represents the model-implied parameter θ according the data with $n = 100$. Panel C plots the asymptotic distribution of $\hat{\theta}_n$ based on the assumed asset pricing restriction displayed in Panel B and the whole sample $\mathbf{y}^n$ with $n = 100$; the normal density indicated by the red solid line is the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ when the model is correctly specified ($f_{n,t} \equiv 0$). By contrast, the normal density indicated by the red dashed line is the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n^f - \theta_0)$ when the model is locally misspecified ($\lambda_1 = [0, 30]^T$, $\lambda_2 = [0, 0]^T$, and $b = 0$).

The graph illustrates the degree of uncertainty about the value of θ according to the baseline model and the dividend data. When $n = 100$, the 95% asymptotic confidence interval for the growth rate θ is approximately $[-0.8\%, 0.8\%]$.

Panel B of Figure 1 plots the model-implied average price-dividend ratio $F(\theta)$ as a function of the dividend growth rate θ. Because $F(\theta)$ rises quickly with θ, there is only a narrow range of θ for which the model-implied average price-dividend ratios would be consistent with the sample mean. To see this, we use the shaded horizontal region near 30 on the y-axis to denote the 95% confidence interval for the average price-dividend ratio from the data (assuming there is no local misspecification, i.e. $f_{n,t} \equiv 0$). The corresponding shaded vertical region on the x-axis shows the range of θ consistent with the model, which is significantly more concentrated than the asymptotic distribution of the baseline estimator.

The full estimator $\hat{\theta}_n$, which is the efficient GMM estimator based on the full set of moment restrictions in (5), is significantly more precise than the baseline estimator. With the moment restrictions correctly specified (i.e., $f_{n,t} \equiv 0$),

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \mathbf{I}_{\mathrm{F}}^{-1}), \quad \text{where} \quad \mathbf{I}_{\mathrm{F}} \equiv D^T \Omega^{-1} D.$$

When $n = 100$, the 95% asymptotic confidence interval for the growth rate θ is approximately $[-0.06\%, 0.06\%]$. A comparison between the asymptotic distributions of the baseline and full estimators in Panel C of Figure 1 reveals how informative the asset pricing restriction in (5) is about θ. This incremental informativeness of the asset pricing restrictions is the focus of our dark matter measure. More precisely, we define the dark matter measure as

$$(8) \qquad \varrho(\theta_0) \equiv \mathbf{I}_{\mathrm{F}}/\mathbf{I}_{\mathrm{B}} - 1 = \left[F'(\theta_0)\sigma_Y/\sigma_P\right]^2 = 83.82.$$

This simple example shows that the informativeness ($\varrho(\theta_0)$) increases in the sensitivity of the asset pricing moment to the parameter value ($F'(\theta_0)$) and decreases in the variability of the asset pricing moment (σ_P).

Dark matter and model fragility

While informative moment restrictions can be very helpful in identifying parameters in the absence of local misspecification, they become a symptom of model fragility with misspecification. It is conventional to define model fragility as the excessive sensitivity of the model's implications, specifically, how well it fits the data, to small perturbations of the data-generating process. To formalize this procedure, we need to be precise on (i) what constitutes small perturbations in data-generating processes, and (ii) how to define excessive sensitivity of the model fit. For (i), if trusting the baseline moment restrictions but lacking confidence in the additional moment restrictions from the structural model, it makes sense to benchmark the magnitude of the perturbation to the uncertainty about θ in the baseline moment restrictions.[8] For (ii), we can assess the extent to which the changes in the moments resulting from parameter perturbations are statistically distinguishable from zero according to the data.

Following the rules of the sensitivity analysis set above, Panel B of Figure 1 indicates that a perturbation of θ within the 95% confidence region of the baseline estimator $\tilde{\theta}_n$ (marked by the two vertical dashed lines) can move the average model-implied price-dividend ratio far away from the confidence region implied by the empirical moments (the narrow shaded region near

[8] We formally justify this benchmark in Section 4.3 by considering local instability and misspecification in the data-generating process.

30 on the y-axis), which means the perturbed GMM model will likely be rejected by the data. Similar to the informativeness of the asset pricing restriction, model sensitivity is higher when the asset pricing moment has a larger gradient with respect to the model parameter θ, and when the error of the asset pricing moment is smaller.

More generally, the dark matter measure can be viewed as a multivariate model sensitivity measure. The direction in which the two efficient asymptotic variances based on the baseline and full models differ the most is also the one in which a local perturbation of the parameter vector θ results in the largest changes in the model fit.

It is also worth noting that the connection between the informativeness of the structural model restrictions and model fragility only makes sense in light of the possibility of model misspecification. If a GMM model is correctly specified, the more sensitive the moments, then the more precise will be the estimate of the parameters by imposing valid restrictions; this can be seen from Panel C of Figure 1, where the asymptotic distribution of $\hat{\theta}_n$ with $f_{n,t} \equiv 0$ is tightly distributed around the true value $\theta_0 = 0$. However, if the asset pricing restrictions are potentially misspecified and unstable, then the extra information they provide may not be valid, thereby making inference problematic.

Lack of refutability

We are interested in testing the validity of the asset pricing moment restriction (i.e., whether $\lambda_1^{(2)} = 0$) given the prior information that the baseline moment restriction is correctly specified (i.e., $\lambda_1^{(1)} = 0$). Considering an upper bound for the maximin local power, we can focus on a subset of alternatives satisfying $b \equiv 0$ to obtain an upper bound characterized by the dark matter measure, albeit it is not necessarily the tightest bound. In this example, the C test (Eichenbaum, Hansen, and Singleton, 1988) is numerically equivalent to the J test (Hansen, 1982); and Newey (1985a) shows that it is asymptotically optimal for these particular alternatives. We consider the set of alternatives $\mathcal{A}_\kappa(\mathrm{Q}_0) \equiv \left\{\lambda_1 \in \mathbb{R}^2 : \lambda_1^{(1)} = 0 \text{ and } |\lambda_1^{(2)}| \geq \kappa\right\}$.

The C test statistic can be rewritten as:

$$C_n = \left[Z_n + \sqrt{\frac{\varrho(\theta_0)}{1+\varrho(\theta_0)}}\left(\frac{\lambda_1^{(1)}}{\sigma_Y}\right) + \sqrt{\frac{1}{1+\varrho(\theta_0)}}\left(\frac{\lambda_1^{(2)}}{\sigma_P}\right)\right]^2 + o_p(1), \tag{9}$$

where $Z_n = \sqrt{\varrho(\theta_0)/(1+\varrho(\theta_0))}\left(n^{-1/2}\sum_{t=1}^n \epsilon_{Y,t}\right) + \sqrt{1/(1+\varrho(\theta_0))}\left(n^{-1/2}\sum_{t=1}^n \epsilon_{P,t}\right)$ converges to a standard normal variable in distribution. According to Newey (1985a) and Chen and Santos

(2018), the maximin asymptotic power of the GMM specification tests of size α is bounded from above by

$$\inf_{\lambda_1 \in \mathcal{A}_\kappa(\mathrm{Q}_0)} \lim_{n\to\infty} \mathbb{P}\left\{C_n > c_{1-\alpha}\right\} \leq \lim_{n\to\infty} \mathbb{P}\left\{\left[Z_n + \sigma_P^{-1}\kappa/\sqrt{1+\varrho(\theta_0)}\right]^2 > c_{1-\alpha}\right\}, \tag{10}$$

where $c_{1-\alpha}$ is the $1-\alpha$ quantile of a chi-square distribution with degree of freedom one. The right-hand side of (10) is an upper bound on the maximin asymptotic power, constructed by choosing $\lambda_1^{(1)} = 0$ and $\lambda_1^{(2)} = \kappa$. Further, according to the continuous mapping theorem, the right-hand side (10) is equal to

$$\lim_{n\to\infty} \mathbb{P}\left\{\left[Z_n + \sigma_P^{-1}\kappa/\sqrt{1+\varrho(\theta_0)}\right]^2 > c_{1-\alpha}\right\} = \mathbb{P}\left\{\chi_1^2(\mu) > c_{1-\alpha}\right\}, \tag{11}$$

where $\chi_1^2(\mu)$ represents a noncentral chi-square distribution with degree of freedom one and the noncentrality parameter $\mu = \dfrac{(\kappa/\sigma_P)^2}{1+\varrho(\theta_0)}$. Combining (10) and (11) leads to

$$\inf_{\lambda_1 \in \mathcal{A}_\kappa(\mathrm{Q}_0)} \lim_{n\to\infty} \mathbb{P}\left\{C_n > c_{1-\alpha}\right\} \leq \mathbb{P}\left\{\chi_1^2(\mu) > c_{1-\alpha}\right\}. \tag{12}$$

Thus, when the GMM model has too much dark matter, the noncentrality parameter μ is very close to zero and thus the upper bound $\mathbb{P}\left\{\chi_1^2(\mu) > c_{1-\alpha}\right\}$ is very close to α, which means that the test power is close to zero. The local power function is visualized in Panel A of Figure 2. The power only starts to get close to one when the misspecification is 40 times of the standard deviation σ_P of the price-dividend ratio.

Intuitively, the baseline moment restrictions have limited ability to refute the cross-equation restrictions implied by the structural model when the dark matter measure $\varrho(\theta_0)$ is large. It is clear from Panel B of Figure 1 that, by tuning the parameter value of θ inside the "acceptable region" imposed by the baseline GMM model (i.e., within the 95% confidence region of the baseline estimator $\tilde{\theta}_n$, marked by the two vertical dashed lines), the econometrician can fit the average model-implied price-dividend ratio over an immensely wide range (i.e., the range between the two horizontal dashed lines), which means the model-implied cross-equation restriction can hardly be rejected by the data. This can also be seen in Panel C of Figure 1: even when the moment condition for the price-dividend ratio is severely misspecified (with $\kappa/\sigma_P = 6$), the baseline GMM model still cannot reject the point estimate $\hat{\theta}_n^f$ (i.e., $\hat{\theta}_n^f$ is still within the confidence interval of θ based on $\tilde{\theta}_n$).

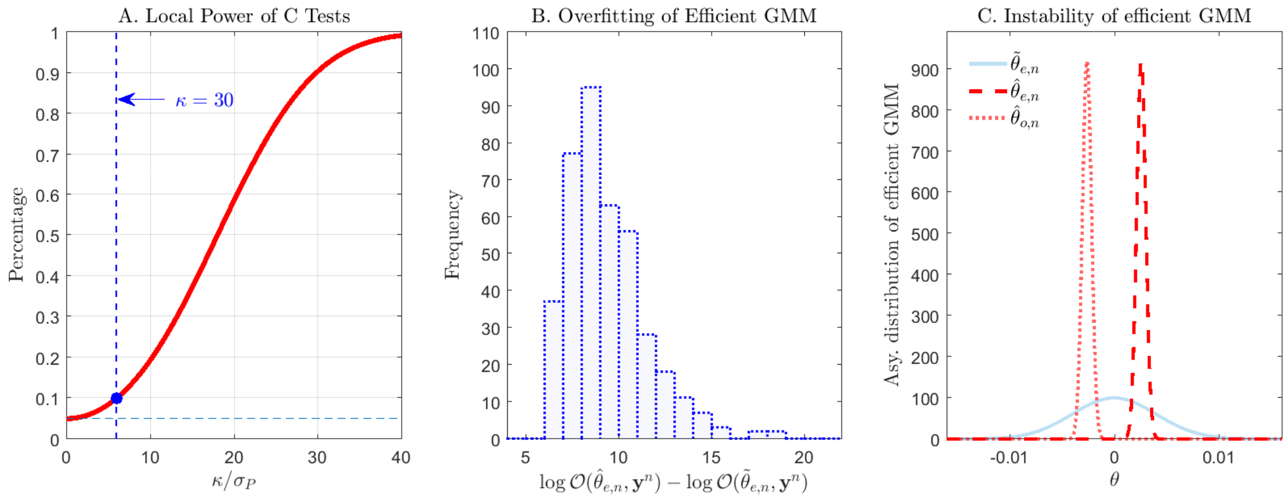

FIGURE 2.— Panel A plots the local power of C tests with $n = 100$. In panel B we simulate 400 independent time series with length $n = 200$. We set the break point $\pi = 1/2$, misspecification $\lambda_1 = [0, 0]^T$, and instability $\lambda_2 = [0, 30]^T$ and $b(\cdot)$ specified in (13) for panels B and C. Panel B displays the difference between the logged overfitting measure of the full efficient GMM estimator $\hat{\theta}_{e,n}$ and that of the baseline efficient GMM estimator $\tilde{\theta}_{e,n}$. Panel C plots the asymptotic distribution of $\hat{\theta}_{e,n}$ based on the assumed asset pricing restriction and the whole sample $\mathbf{y}^n$ with instability; the normal density indicated by the red dotted line is the asymptotic distribution of $\sqrt{\pi n}(\hat{\theta}_{e,n} - \theta_0)$. By contrast, the normal density indicated by the red dashed line is the asymptotic distribution of $\sqrt{\pi n}(\hat{\theta}_{\mathrm{o,n}} - \theta_0)$.

Poor out-of-sample fit

A common method for evaluating out-of-sample fit of GMM models is to hold out data from the model estimation (e.g., Schorfheide and Wolpin, 2012; Müller and Watson, 2016). We split the entire time series $\mathbf{y}^n \equiv \{\mathbf{y}_1, \cdots, \mathbf{y}_n\}$ into two non-overlapping subsamples $\mathbf{y}^n_{\mathrm{e}} \equiv \left\{\mathbf{y}_1, \cdots, \mathbf{y}_{\lfloor \pi n \rfloor}\right\}$ and $\mathbf{y}^n_{\mathrm{o}} \equiv \left\{\mathbf{y}_{\lfloor \pi n \rfloor + 1}, \cdots, \mathbf{y}_n\right\}$ with $\pi \in (0, 1/2]$. The first segment $\mathbf{y}^n_{\mathrm{e}}$ is used as the estimation sample, while the second segment $\mathbf{y}^n_{\mathrm{o}}$ is used as the holdout sample.[9] In particular, for the true local data-generating process in (6) and (7), we assume that $\lambda_1 = [0, 0]^T$, $\lambda_2 = [0, 30]^T$, $\pi = 1/2$, and

$$b(t/n) = \begin{cases} 1, & \text{when } 1 \le t \le \lfloor \pi n \rfloor \\ -1, & \text{when } \lfloor \pi n \rfloor < t \le n, \end{cases} \tag{13}$$

where the sequence $b(t/n)$ captures the structural breaks and π is the break point.

We consider the overfitting measure of the efficient GMM estimator $\hat{\theta}_{\mathrm{e,n}}$ based on the estimation sample $\mathbf{y}^n_{\mathrm{e}}$, defined as the extent to which the out-of-sample fitting error exceeds the

[9]The out-of-sample approach, treating $\mathbf{y}^n_{\mathrm{o}}$ as future hypothetical data, can be viewed as a standard cross-validation method based on observed data.

in-sample fitting error:[10]

$$\mathcal{O}(\hat{\theta}_{\mathrm{e,n}}, \mathbf{y}^n) \equiv \frac{1}{2}\left[\mathcal{L}(\hat{\theta}_{\mathrm{e,n}}, \mathbf{y}^n_{\mathrm{o}}) - \mathcal{L}(\hat{\theta}_{\mathrm{e,n}}, \mathbf{y}^n_{\mathrm{e}})\right], \tag{14}$$

with $\mathcal{L}(\theta, \mathbf{y}^n_s) \equiv J(\theta, \mathbf{y}^n_s) - J(\theta_0, \mathbf{y}^n_s)$ for $s \in \{e, o\}$ and the parameter value θ_0 ensures that the baseline moment restriction $\mathrm{E}\left[m^{(1)}(\mathbf{y}_t, \theta_0)\right] = 0$ is perfectly satisfied. It can be shown that (see Theorem 2) the expected overfitting measure can be approximated by

$$\mathrm{E}\left[\mathcal{O}(\hat{\theta}_{\mathrm{e,n}}, \mathbf{y}^n)\right] \approx 1 + (\lambda_2^{(2)}/\sigma_P)^2 \varrho(\theta_0), \quad \text{as } n \text{ approaches } +\infty. \tag{15}$$

Panel B of Figure 2 displays the histogram of difference between logged overfitting measures $\log \mathcal{O}(\hat{\theta}_{e,n}, \mathbf{y}^n)$ of the full efficient GMM estimator $\hat{\theta}_{e,n}$ and those of the baseline efficient GMM estimator $\tilde{\theta}_{e,n}$. For the instability $\kappa = 30$ (i.e. 6 times of σ_P), which is hard to reject using C tests (see Panel A), the degree of overfitting by the efficient GMM estimator $\hat{\theta}_{e,n}$ is substantial. This result suggests that robust estimation is particularly relevant for GMM models with large dark matter measure.

Define $\hat{\theta}_{\mathrm{o,n}}$ to be the efficient GMM estimator based on the holdout sample $\mathbf{y}^n_{\mathrm{o}}$. The out-of-sample fit should be poor if the efficient GMM estimators $\hat{\theta}_{\mathrm{e,n}}$ and $\hat{\theta}_{\mathrm{o,n}}$, based on $\mathbf{y}^n_{\mathrm{e}}$ and $\mathbf{y}^n_{\mathrm{o}}$ respectively, are statistically separate from each other. This can be seen in the case of the asymptotic distributions of $\hat{\theta}_{\mathrm{e,n}}$ and $\hat{\theta}_{\mathrm{o,n}}$ (Panel C of Figure 1), which are centered away from the correct value $\theta_0 = 0$ for the baseline GMM model and therefore distant from each other. The distance between the in-sample estimator $\hat{\theta}_{\mathrm{e,n}}$ and the out-of-sample estimator $\hat{\theta}_{\mathrm{o,n}}$ is significant according to the asymptotic distribution, which implies that the estimator $\hat{\theta}_{\mathrm{e,n}}$ has a poor out-of-sample fit.

General theory and empirical examples

In the remainder of the paper, we formally develop the set of results illustrated by the simple example above. We consider the setting of weakly dependent time series data, which are prevalent in financial and macroeconomic studies, and allow for local perturbations (e.g., Hansen and Sargent, 2001) and instability (e.g., Li and Müller, 2009) of data-generating processes in a semiparametric framework. We then define the dark matter measure for a general class of GMM models, and formally establish the connection between the dark matter measure, the model's

[10]Overfitting measure is also studied by Mullainathan and Spiess (2017) and Hansen and Dumitrescu (2018).

refutability, and its out-of-sample fit. Further, We use the dark matter measure to analyze real data examples including the rare disaster risk and the long-run risk models. We also provide discussion on what to do with fragile models (i.e. models with large dark matter measure) in Section 6.

3. THE DARK MATTER MEASURE

In this section, we set up the model and introduce the dark matter measure. We then discuss the connections between the dark matter measure, sensitivity analysis, and model testability.

3.1. *Model Setup*

Let $\mathcal{Y} = \mathbb{R}^{d_y}$, the d_y-dimensional Euclidean space with Borel σ-field $\mathcal{F}$. Let $\mathcal{P}$ denote the collection of all probability measures on the measurable space $(\mathcal{Y} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{F})$ with the product sample space $\mathcal{Y} \times \mathcal{Y}$ and the product σ-field $\mathcal{F} \otimes \mathcal{F}$.

Markov processes

We consider a subspace of $\mathcal{P}$, denoted by $\mathcal{H}$, in which each probability measure is the bivariate marginal distribution Q for a time-homogeneous Harris ergodic and stationary Markov process $\{\mathbf{y}_t : t = 0, 1, \cdots\}$ satisfying the Doeblin condition.[11] Following Bickel and Kwon (2001), we parameterize time-homogeneous Markov processes by the bivariate marginal distributions Q of $(\mathbf{y}_{t-1}, \mathbf{y}_t)$ for any $t \geq 1$. We denote the $(n+1)$-variate joint distribution of $\mathbf{y}^n \equiv \{\mathbf{y}_0, \cdots, \mathbf{y}_n\}$ corresponding to Q by $\mathbb{P}_n$. 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).[12]

[11] The set of Markov processes satisfying the Doeblin condition includes a broad class of time series commonly used in finance and macroeconomics; see, e.g., Stokey and Lucas (1989) and Ljungqvist and Sargent (2004).

[12] First-order Markov models are widely adopted for approximating financial and economic time series. Many prominent structural asset pricing models feature state dynamics as first-order Markovian processes (e.g., Campbell and Cochrane, 1999; Bansal and Yaron, 2004; Gabaix, 2012; Wachter, 2013).

Structural models and moment restrictions.

Consider a stable structural model denoted by $\mathcal{Q}$, which aims to capture certain statistical features of the observed data $\mathbf{y}^n$. The parameters of such a "stable" model, denoted by θ, are constant over time (e.g., Li and Müller, 2009).[13] We assume that the restrictions imposed by the model on the data can be summarized by a set of moment restrictions, and that the model's performance in a given data sample can be measured by the degree to which these moment restrictions are violated (i.e., the fit of moment restrictions). As we will explain in Section 3.4 below, our notion of model fragility is also based on the moment restrictions, specifically their sensitivity to local perturbations of the underlying data-generating process.[14] We follow the literature (e.g., Li and Müller, 2009; Chen and Santos, 2018), and refer to these models as GMM models. As reflected in the original applications of GMM in asset pricing (Hansen and Singleton, 1982, 1983) and recently emphasized by Hansen (2014), structural asset pricing models are typically partially specified. Further, GMM has proven particularly valuable for analyzing structural models via focusing on key cross-equation restrictions such as Euler equations, without being overly influenced by the details and potential singularities of the remainder of the structural model. Accordingly, we adopt the semiparametric framework of GMM models.

Specifically, we assume the moment function corresponding to the full structural model is $m(\cdot,\theta) \in \mathbb{R}^{d_m}$, defined on a compact parameter set $\Theta \in \mathbb{R}^{d_\theta}$ with nonempty interior, and denote the full GMM model by $\mathcal{Q}$,

$$\mathcal{Q} = \left\{ \mathrm{Q} \in \mathcal{H} : \mathrm{E}^{\mathrm{Q}}\left[m(\cdot,\theta)\right] = 0, \text{for some } \theta \in \Theta \right\}, \tag{16}$$

which is a collection of probability measures under which the moment restrictions hold for some parameter vector θ. The system of moment restrictions is over-identified; that is, the number of model parameters is less than that of moment restrictions (i.e., $d_\theta < d_m$).

We assume that the moment function $m(\cdot,\theta)$ has a recursive structure:

$$m(\mathbf{y}_{t-1}, \mathbf{y}_t, \theta) = \begin{bmatrix} m^{(1)}(\mathbf{y}_{t-1}, \mathbf{y}_t, \theta^{(1)}) \\ m^{(2)}(\mathbf{y}_{t-1}, \mathbf{y}_t, \theta^{(1)}, \theta^{(2)}) \end{bmatrix}, \quad \text{with} \quad \theta = \begin{bmatrix} \theta^{(1)} \\ \theta^{(2)} \end{bmatrix}. \tag{17}$$

[13]Technically, the model parameters may vary with the sample size n, though they do not depend on the time index $t \in \{1, \cdots, n\}$.

[14]Kocherlakota (2016) adopts a similar notion in studying the sensitivity of real macro models to the specification of the Phillips curve.

Here $\theta^{(1)}$ is a $d_{\theta,1}$-dimensional sub-vector of θ, with $d_{\theta,1} \leq d_\theta$, and $m^{(1)}(\cdot, \theta^{(1)})$ has dimension $d_{m,1} \geq d_{\theta,1}$. The baseline moments can be represented by a selection matrix $\Gamma_{m,1}$:

$$m^{(1)}(\mathbf{y}_{t-1}, \mathbf{y}_t, \theta^{(1)}) = \Gamma_{m,1} m(\mathbf{y}_{t-1}, \mathbf{y}_t, \theta), \text{ with } \Gamma_{m,1} \equiv \left[I, 0_{d_{m,1} \times (d_m - d_{m,1})}\right]. \tag{18}$$

The assumption of the recursive structure for the moment function enables us to examine the fragility of a subset of moment restrictions, namely those in $m^{(2)}(\cdot, \theta)$. Such recursive structures are common in asset pricing. For example, the moments $m^{(1)}(\cdot, \theta^{(1)})$ could be derived from a statistical model of the real quantities (such as consumption), while the additional moments $m^{(2)}(\cdot, \theta)$ may apply to the joint dynamics of the real quantities and asset prices. Since the first coordinate block of the moment function $m(\cdot, \theta)$ only depends on $\theta^{(1)}$, the Jacobian matrix $D(\theta)$ is block lower triangular:

$$D(\theta) = \begin{bmatrix} D_{11}(\theta) & 0 \\ D_{12}(\theta) & D_{22}(\theta) \end{bmatrix}, \text{ where } D_{ij}(\theta) \equiv \mathrm{E}^{\mathrm{Q}}\left[\nabla_{\theta^{(j)}} m^{(i)}(\cdot, \theta)\right] \text{ and } i, j = 1, 2. \tag{19}$$

Corresponding to the first coordinate block $m^{(1)}(\cdot, \theta^{(1)})$ of the moment function $m(\cdot, \theta)$ in (17), we define the baseline GMM model $\mathcal{Q}^{(1)}$,

$$\mathcal{Q}^{(1)} = \left\{Q \in \mathcal{H} : \mathrm{E}^{\mathrm{Q}}\left[m^{(1)}(\cdot, \theta^{(1)})\right] = 0 \text{ for some } \theta^{(1)} \in \Theta^{(1)}\right\}. \tag{20}$$

The baseline model is a collection of probability measures under which the first block of moment restrictions, hereafter referred to as the baseline moments, hold for some parameter vector $\theta^{(1)}$. This definition is analogous to the definition of the full model, and clearly $\mathcal{Q} \subset \mathcal{Q}^{(1)}$. The subvector $\theta^{(2)}$ can only be identified by the moment restrictions not contained in the baseline model. We refer to $\theta^{(2)}$ as the nuisance parameters.

Although economic models often feature conditional moment restrictions, for estimation and testing, it is common to focus on a finite number of implied unconditional moment restrictions by using nonlinear instrumental variables (e.g., Hansen and Singleton, 1982, 1983; Hansen, 1985; Nagel and Singleton, 2011). For simplicity, we take these unconditional moments as the starting point in our analysis.

Following the definition of the full structural model (16), we define a mapping from the probability measure of the bivariate marginal distribution $\mathrm{Q} \in \mathcal{Q}$ to model parameters θ, $\theta =$

$\vartheta(\mathrm{Q})$, such that

$$\mathrm{E}^{\mathrm{Q}}\left[m(\cdot,\vartheta(\mathrm{Q}))\right]=0. \tag{21}$$

Calibrated models

Consider a calibrated model parameter value $\theta_0 \in int(\Theta)$, the interior of Θ, such that the moment restrictions evaluated at θ_0 hold under some $\mathrm{Q}_0 \in \mathcal{Q}$:

$$\mathrm{E}^{\mathrm{Q}_0}\left[m(\cdot,\theta_0)\right]=0. \tag{22}$$

Note that Q_0 remains unknown to the econometrician, even though θ_0 may be known. The calibrated full and baseline GMM models are sets of probability measures satisfying

$$\mathcal{Q}(\theta_0) \equiv \left\{\mathrm{Q} \in \mathcal{H} : \mathrm{E}^{\mathrm{Q}}\left[m(\cdot,\theta_0)\right]=0\right\}, \text{ and} \tag{23}$$

$$\mathcal{Q}^{(1)}(\theta_0^{(1)}) \equiv \left\{\mathrm{Q} \in \mathcal{H} : \mathrm{E}^{\mathrm{Q}}\left[m^{(1)}(\cdot,\theta_0^{(1)})\right]=0\right\}, \text{ respectively.} \tag{24}$$

By definition, $\mathcal{Q}(\theta_0) \subset \mathcal{Q}$. Moreover, $\mathcal{Q}(\theta_0)$ is non-empty (due to the assumption for θ_0). We pick one distribution from $\mathcal{Q}(\theta_0)$ and denote it by Q_0, which is a distribution under which the moment restrictions of the full model hold under the calibrated parameters θ_0.

3.2. *The Efficient GMM Estimation*

Under the distribution Q_0, the spectral density matrices (at zero frequency) for the baseline and full models are

$$\Omega_{11}=\sum_{t=-\infty}^{\infty}\mathrm{E}^{\mathrm{Q}_0}\left[m^{(1)}(\mathbf{y}_0,\mathbf{y}_1,\theta_0)m^{(1)}(\mathbf{y}_{t-1},\mathbf{y}_t,\theta_0)^T\right], \text{ and} \tag{25}$$

$$\Omega=\sum_{t=-\infty}^{\infty}\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_0,\mathbf{y}_1,\theta_0)m(\mathbf{y}_{t-1},\mathbf{y}_t,\theta_0)^T\right], \text{ respectively,} \tag{26}$$

where Ω_{11} is the upper-left block of Ω. We assume that both Ω and the Jacobian matrix D are known. In general, computing the expectations requires knowledge of the distribution Q_0. When Q_0 is unknown in practice, expectations can be replaced by their consistent estimators. For example, several consistent estimators of the covariance matrices are provided by Newey and West (1987), Andrews (1991), and Andrews and Monahan (1992). These estimation methods usually require a two-step plug-in procedure introduced by Hansen (1982) when θ_0 is unknown. We further assume that $\Omega = I$, which is innocuous because we can always rotate the system without altering the structure (Hansen, 2007b). More details are provided in Appendix G.1.

For any given θ, we define

$$m_t(\theta) \equiv m(\mathbf{y}_{t-1}, \mathbf{y}_t, \theta) \text{ and } m_t^{(1)}(\theta^{(1)}) \equiv m^{(1)}(\mathbf{y}_{t-1}, \mathbf{y}_t, \theta^{(1)}),$$

and denote the empirical moment functions for the full and baseline GMM models by

$$\widehat{m}_n(\theta) \equiv \frac{1}{n}\sum_{t=1}^{n} m_t(\theta) \text{ and } \widehat{m}_n^{(1)}(\theta^{(1)}) \equiv \frac{1}{n}\sum_{t=1}^{n} m_t^{(1)}(\theta^{(1)}), \text{ respectively.}$$

Then, the efficient GMM estimator $\hat{\theta}_n$ of the full model and that of the baseline model $\tilde{\theta}_n^{(1)}$ minimize

$$J(\theta, \mathbf{y}^n) \equiv n\left|\widehat{m}_n(\theta)\right|^2 \text{ and } J^{(1)}(\theta^{(1)}, \mathbf{y}^n) \equiv n\left|\widehat{m}_n^{(1)}(\theta^{(1)})\right|^2, \text{ respectively.} \tag{27}$$

3.3. *Information Matrices Based on Unconditional Moment Restrictions*

We now introduce information matrices for the GMM models. In statistics and econometrics, information regarding model parameters is often quantified by the efficiency bound on parameter estimators. One example is the Fisher information matrix for a given parametric family of likelihood functions, which is justified by the Cramér-Rao efficiency bound under the minimax criterion. The same idea can be extended to semiparametric models (e.g., Bickel, Klaassen, Ritov, and Wellner, 1993).

In this paper, we extend the semiparametric efficiency bound result for unconditional moment restrictions of Chamberlain (1987, Theorem 2) from i.i.d. data-generating processes to Markov processes with local instability. In Appendix B, we show that the optimal GMM covariance matrix derived by Hansen (1982) achieves the semiparametric minimax efficiency bound for unconditional moment restrictions with Markov data-generating processes that are locally unstable. We denote

$$D \equiv D(\theta_0) \text{ and } D_{ij} \equiv D_{ij}(\theta_0) \text{ for } i, j = 1, 2. \tag{28}$$

Then, the information matrices for $\theta^{(1)}$ in the baseline model and for θ in the full model, evaluated at $\theta_0^{(1)}$ and θ_0, respectively, are

$$\mathbf{I}_{\mathrm{B}} = D_{11}^T D_{11} \quad \text{and} \quad \mathbf{I}_{\Omega} = D^T D = \begin{bmatrix} D_{11}^T D_{11} + D_{21}^T D_{21} & D_{21}^T D_{22} \\ D_{22}^T D_{21} & D_{22}^T D_{22} \end{bmatrix}. \tag{29}$$

We define the marginal information matrix for $\theta^{(1)}$ in the full model, evaluated at $\theta_0^{(1)}$, as

$$\mathbf{I}_{\mathrm{F}} = \left[\Gamma_{\theta,1} \mathbf{I}_{\Omega}^{-1} \Gamma_{\theta,1}^T\right]^{-1}, \quad \text{where } \Gamma_{\theta,1} \equiv \left[I, 0_{d_{\theta,1} \times (d_\theta - d_{\theta,1})}\right], \tag{30}$$

which accounts for the uncertainty concerning the nuisance parameters $\theta^{(2)}$ when gauging the information about $\theta^{(1)}$ provided by the moment restrictions. Based on the inversion rule of partitioned matrices, the marginal information matrix $\mathbf{I}_{\mathrm{F}}$ can be rewritten as

$$\mathbf{I}_{\mathrm{F}} = D_{11}^T D_{11} + D_{21}^T \Lambda_2 D_{21}, \quad \text{with } \Lambda_2 \equiv I - D_{22}(D_{22}^T D_{22})^{-1} D_{22}^T. \tag{31}$$

Although our objective is to measure model fragility, in Section 3.4 we introduce a measure of the incremental informativeness of the moment restrictions about model parameters. We then argue that this informativeness measure is intuitively connected to the notion of model sensitivity described earlier in Section 2, and we shall formally establish the link between our informativeness measure and model fragility in Sections 4 and 5.

3.4. *The Dark Matter Measure*

In this section, we ask how informative the moment restrictions are regarding the model parameters $\theta^{(1)}$. Since $\theta^{(1)}$ appears in both the baseline moment restrictions and the additional moment restrictions in the full model, the cross-equation restrictions provide additional information about $\theta^{(1)}$ above and beyond the baseline model. The informativeness of the moment restrictions naturally depends on the sensitivity of the moments to changes in model parameters. If a small change in the parameter values can dramatically change the value of the moment function (i.e., high sensitivity), then imposing the moment restrictions empirically will tend to greatly restrict the parameter estimates (i.e., the moment restrictions are informative).

Before introducing our information measure, we discuss the relevant regularity conditions, including smoothness, rank, and identification.

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) *The moment restrictions are over-identified:* $d_\theta < d_m$*;*

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

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

Remark 1 *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) is the standard over-identification condition in GMM (Hansen, 1982). Condition (ii) is also a standard identification assumption to ensure that the sequence of GMM estimators has a unique limit (Hansen, 1982). Condition (iii) is the rank condition for moment restrictions, and is the sufficient condition for local identification enabling us to consistently estimate* θ_0*.*

We now introduce our dark matter measure.

Definition 1 *Let the incremental information matrix of the full model relative to the baseline models be*

$$\Pi \equiv \mathbf{I}_F^{1/2}\mathbf{I}_B^{-1}\mathbf{I}_F^{1/2} - I. \tag{32}$$

The dark matter measure is defined as the largest eigenvalue of Π*, denoted by*[15]

$$\varrho(\theta_0) \equiv \max_{|\mathbf{v}|=1} \mathbf{v}^T\Pi\mathbf{v}. \tag{33}$$

To better understand the dark matter measure, we rewrite it as

$$\varrho(\theta_0) = \max_{|\mathbf{v}|=1} \frac{\mathbf{v}^T\mathbf{I}_{\mathrm{B}}^{-1}\mathbf{v}}{\mathbf{v}^T\mathbf{I}_{\mathrm{F}}^{-1}\mathbf{v}} - 1. \tag{34}$$

As Equation (34) shows, our measure effectively compares the asymptotic covariance matrices of the two estimators of $\theta^{(1)}$: the matrix based on the baseline model, and the matrix based on the full model. It is the largest ratio, over all possible directions $\mathbf{v} \in \mathbb{R}^{d_{\theta,1}}$, of the two asymptotic variances of the efficient GMM estimator for a linear combination of model parameters $\mathbf{v}^T\theta^{(1)}$ under the baseline and full models.

[15] We focus on the one-dimensional worst-case fragility. There are straightforward extensions to the cases in which $\mathbf{v}$ is a matrix.

Equation (34) shows that the dark matter measure has a natural "effective-sample-size" interpretation. This equation gives the minimum sample size required for the estimator of the baseline model to match the asymptotic precision (the inverse of the variance) of the estimator of the full structural model in all directions of the parameter space. Because asymptotic variance scales inversely with the sample size, the effective sample size is $n\,[1+\varrho(\theta_0)]$.

Our dark matter measure isolates the information provided by the structural model above and beyond the baseline model. For the same structural model, alternative choices of the baseline model affect the magnitude of the dark matter measure. To this point, we have been silent on the question of how the baseline model should be chosen in relation to the full structural model. In general, there is no hard rule for this choice, beyond the technical requirement that the associated baseline parameters $\theta^{(1)}$ be identified by the baseline model. Desirable choices of the baseline model depend on which aspects of the model the fragility analysis aims to capture.

4. LOCAL MISSPECIFICATION AND INSTABILITY

A formal analysis of model fragility requires a framework for misspecification and instability. We adopt a statistical method similar to that of Hansen and Sargent (2001): the econometrician treats $\mathbb{P}_0$ as an approximation of the true data-generating process by taking into account a class of alternative data-generating processes that are statistically difficult to distinguish from Q_0 (i.e. a neighborhood of Q_0 in the space of probability measures) and assuming that the true process lies in such a collection of local alternatives. To model instability, we generalize the local instability framework of Li and Müller (2009) to the semiparametric setting, which provides a general econometric playground within which we analyze GMM model properties.

We first specify the true local data generating process in Subsection 4.1. We then introduce the concept of model misspecification in Subsection 4.2, and extend our framework by incorporating the concept of local instability in Subsection 4.3.

4.1. *Local Data-Generating Processes*

Our analysis is local in nature. We focus on a calibrated model with model parameter θ_0 as defined in (22), with the corresponding bivariate marginal distribution Q_0. To characterize the locally perturbed models, we define the collection of local perturbations of Q_0, denoted by $\mathcal{N}(\mathrm{Q}_0)$, as follows. Note that $\mathcal{N}(\mathrm{Q}_0)$ is a subset of $\mathrm{L}^2(\mathrm{Q}_0)$, the space of square-integrable random variables on the probability space $(\mathcal{Y}\times\mathcal{Y}, \mathcal{F}\otimes\mathcal{F}, \mathrm{Q}_0)$.

Definition 2 *The collection* $\mathcal{N}(\mathrm{Q}_0)$ *consists of the one-dimensional parametric family of bivariate distributions* $\mathrm{Q}_{s,f}$ *indexed by* $s \in (-\epsilon, \epsilon)$ *for some* $\epsilon > 0$ *and* $f \in L^2(\mathrm{Q}_0)$*, such that the path* $\mathrm{Q}_{s,f} \in \mathcal{H}$ *passes through the probability measure* $\mathrm{Q}_0 \in \mathcal{H}$ *at* $s = 0$*, and* $\mathrm{Q}_{s,f}$ *satisfies the smoothness condition (Hellinger-differentiability condition):*[16]

$$\frac{d\mathrm{Q}_{s,f}}{d\mathrm{Q}_0} = 1 + sf + s\Delta(s), \tag{35}$$

where $\Delta(s)$ *converges to* 0 *in* $L^2(\mathrm{Q}_0)$ *as* $s \to 0$*. Here, we refer to the scalar measurable function* $f \in L^2(\mathrm{Q}_0)$ *as the "score" of the parametric model* $s \mapsto \mathrm{Q}_{s,f}$*.*

Proposition 1 (Necessary Properties of Scores) *If* $f \in L^2(\mathrm{Q}_0)$ *satisfies* (35)*, then it follows that (i)* $\mathrm{E}^{\mathrm{Q}_0}[f] = \mathrm{E}^{\mathrm{Q}_0}[\Delta(s)] = 0$ *for all* s*, and (ii)* $\mathrm{E}^{\mathrm{Q}_0}[f(\mathbf{y}, \mathbf{y}')|\mathbf{y}] = \mathrm{E}^{\mathrm{Q}_0}[f(\mathbf{y}', \mathbf{y})|\mathbf{y}]$*.*

Now, appealing to the concept in Definition 2, we specify the true local data-generating process. We denote the joint distribution of $\mathbf{y}^n$ corresponding to the bivariate marginal distribution Q_0 by $\mathbb{P}_{0,n}$. Deviating from $\mathbb{P}_{0,n}$, the true local data-generating process for $\mathbf{y}^n$ has the joint distribution $\mathbb{P}_n^*$ with a sequence of bivariate marginal distributions for each consecutive pair $(\mathbf{y}_{t-1}, \mathbf{y}_t)$, $\mathrm{Q}_n^* \equiv \mathrm{Q}_{1/\sqrt{n}, f_{n,t}^*}$, which is characterized by

$$\frac{\mathrm{dQ}_{1/\sqrt{n}, f_{n,t}^*}}{\mathrm{dQ}_0} = 1 + \frac{f_{n,t}^*}{\sqrt{n}} + \Delta_n, \ \text{where} \tag{36}$$

$$f_{n,t}^* \equiv [1, b^*(t/n)] g^*(\mathbf{y}_{t-1}, \mathbf{y}_t) \ \text{ and } \ \sqrt{n}\Delta_n \to 0 \text{ in } L^2(\mathrm{Q}_0). \tag{37}$$

The vector g^* has two elements: $g^* = [g_1^*, g_2^*]^T$ with $g_1^*, g_2^* \in L^2(\mathrm{Q}_0)$. In other words, $f_{n,t}^* = g_1^*(\mathbf{y}_{t-1}, \mathbf{y}_t) + g_2^*(\mathbf{y}_{t-1}, \mathbf{y}_t) b^*(t/n)$ where g_1^* represents time-invariant perturbation, while g_2^* multiplied by $b(t/n)$ represents time-varying perturbation (i.e., local instability). The unknown function $b^*(\cdot)$ is a deterministic function on $[0, 1]$ that generates local instability.[17] When n is large, $1 + f_{n,t}^*/\sqrt{n}$ is approximately the Radon-Nikodym density of $\mathrm{Q}_{1/\sqrt{n}, f_{n,t}^*}$ with respect to Q_0.

Prior to imposing additional regularity conditions on the true score $f_{n,t}^*$, we define a set of square-integrable variables corresponding to Q_0.

[16] The smoothness condition (35) is equivalent to the Hellinger-differentiability, shown in Appendix G.2. It is a common regularity condition adopted for (semi)parametric inference (e.g., van der Vaart, 1988).

[17] Similar to, for example, Andrews (1993), Sowell (1996), and Li and Müller (2009), we assume instability to be non-stochastic in contrast to, for example, Stock and Watson (1996, 1998), Primiceri (2005), and Cogley and Sargent (2005). The assumption is for technical simplicity. We can extend from non-stochastic to stochastic instability following the arguments in Li and Müller (2009).

DEFINITION 3 (Set of Scores) *For* $\mathrm{Q}_0 \in \mathcal{Q}$, *define*

$$L_0^2(\mathrm{Q}_0) \equiv \left\{\varsigma \in L^2(\mathrm{Q}_0): \ \mathrm{E}^{\mathrm{Q}_0}\left[\varsigma(\mathbf{y}, \mathbf{y}')\right] = 0 \ \textit{and} \ \mathrm{E}^{\mathrm{Q}_0}\left[\varsigma(\mathbf{y}, \mathbf{y}')|\mathbf{y}\right] = \mathrm{E}^{\mathrm{Q}_0}\left[\varsigma(\mathbf{y}', \mathbf{y})|\mathbf{y}\right]\right\}. \tag{38}$$

Given the notation $L_0^2(\mathrm{Q}_0)$, the necessary conditions for scores derived in Proposition 1 can be restated as $f \in L_0^2(\mathrm{Q}_0)$. We then make the following assumption about the true $f_{n,t}^*$.

ASSUMPTION 2 (Local Data-Generating Process) *The true local data-generating process in (36) satisfies the following conditions:*

(i) $g^* \in \mathcal{G}(\mathrm{Q}_0)$, *which is defined as*

$$\mathcal{G}(\mathrm{Q}_0) \equiv \left\{g = [g_1, g_2]^T: \ \mathrm{E}^{\mathrm{Q}_0}\left[g_2(\mathbf{y}, \mathbf{y}')|\mathbf{y}\right] = 0 \ \textit{and} \ g_1, g_2 \in L_0^2(\mathrm{Q}_0)\right\};$$

(ii) $b^* \in \mathcal{B}$, *which is defined as*

$$\mathcal{B} \equiv \left\{ b: \begin{array}{l} |b(u)| \le 1 \ \textit{for all} \ u \in [0,1] \ \textit{and} \ \displaystyle\int_0^1 b(u)du = 0, \ \textit{whose path has a} \\ \textit{finite number of discontinuities and one-sided limits everywhere.} \end{array} \right\}.$$

REMARK 2 *The first part of Assumption 2 (i) implies that* $\mathrm{E}^{\mathrm{Q}_0}\left[f_{n,t}^*(\mathbf{y}, \mathbf{y}')|\mathbf{y}\right] = \mathrm{E}^{\mathrm{Q}_0}\left[g_1^*(\mathbf{y}, \mathbf{y}')|\mathbf{y}\right]$ *is invariant over time, which further ensures that the univariate marginal distribution of the true joint distribution* $\mathbb{P}_n^*$ *is invariant over time (Proposition 3). The second part of Assumption 2 (i) that* $g_1^*, g_2^* \in L_0^2(\mathrm{Q}_0)$ *is not restrictive since it is guaranteed by Proposition 1.*

Next, we impose additional assumptions about the heteroskedasticity of the locally unstable data-generating process 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) $n^{-1} \max_{1 \le t \le n} |g(\mathbf{y}_{t-1}, \mathbf{y}_t)|^2 = o_p(1)$;

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

REMARK 3 *Condition (i) 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 in our proofs. Condition (ii) 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 5 in Appendix A) and thus ensuring that the locally unstable data-generating process is contiguous to the stable data-generating process (see Corollary 2 in Appendix A). 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.

Finally, we extend the global identification condition in Assumption 1 (ii) from the reference distribution Q_0 to its perturbations.

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* $\mathbf{H}^2(\mathrm{Q}_{s,f}, \mathrm{Q}_0) < \epsilon$.

REMARK 4 *We define the collection of perturbed distributions* $\mathcal{N}(\mathrm{Q}_0)$ *in Definition 2, and the mapping* $\vartheta(\cdot)$ *in (21). This assumption ensures that the sequence of GMM estimators has a unique limit when the true distribution is a perturbation of* Q_0.

4.2. *Misspecification of GMM Models*

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

(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* $\{\mathbf{y}_{t-j}\}_{j=0}^{\infty}$.

REMARK 5 *Conditions (i) and (ii) 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, 1985a; 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 7 in Appendix A and Theorem 4 in Appendix B).*

Tangent space and misspecification

For a given $f \in L_0^2(\mathrm{Q}_0)$, we further require that the path of locally perturbed distributions satisfies $\mathrm{Q}_{s,f} \in \mathcal{Q}$ and that $\vartheta(\mathrm{Q}_{s,f})$, as a function of s, is differentiable with respect to s at $s=0$. The collection of such scores f is defined as follows:

$$\mathcal{T}(\mathrm{Q}_0) \equiv \left\{ f \in L_0^2(\mathrm{Q}_0) : \begin{array}{l} \exists \text{ a path } \mathrm{Q}_{s,f} \text{ such that } \mathrm{Q}_{s,f} \in \mathcal{Q} \cap \mathcal{N}(\mathrm{Q}_0) \text{ for all } s \in (-\epsilon, \epsilon) \\ \text{for some } \epsilon > 0 \text{ and } \vartheta(\mathrm{Q}_{s,f}) \text{ is differentiable at } s = 0 \end{array} \right\}.$$

We refer to the set $\mathcal{T}(\mathrm{Q}_0)$ above as the *tangent set* of $\mathcal{Q}$ at Q_0. We further characterize the tangent set $\mathcal{T}(\mathrm{Q}_0)$ as follows:

$$\mathcal{T}(\mathrm{Q}_0) = \left\{ f \in L_0^2(\mathrm{Q}_0) : \ \lambda(f) \in \mathsf{lin}(D) \right\}, \text{ with } \lambda(f) \equiv \mathrm{E}^{\mathrm{Q}_0}\left[m(\cdot, \theta_0) f \right], \tag{39}$$

where $\lambda(f)$ is a linear operator on $L_0^2(\mathrm{Q}_0)$ and linear space $\mathsf{lin}(D)$ is spanned by the columns of the Jacobian matrix D defined in (28). This characterization is standard in the literature (e.g., Severini and Tripathi, 2013; Chen and Santos, 2018) and can be proved using an implicit function theorem. Equation (39) implies that $\mathcal{T}(\mathrm{Q}_0)$ is a linear space. Whenever $f_{n,t}^* \in L_0^2(\mathrm{Q}_0) \setminus \mathcal{T}(\mathrm{Q}_0)$, the GMM model $\mathcal{Q}$ is *locally misspecified* with respect to the true local data-generating process $(\mathbf{y}_{t-1}, \mathbf{y}_t)$, $\mathrm{Q}_n^* \equiv \mathrm{Q}_{1/\sqrt{n}, f_{n,t}^*}$, defined in (36).

One direct implication of (39) is that, if $d_\theta < d_m$, then $\mathcal{T}(\mathrm{Q}_0) \neq L_0^2(\mathrm{Q}_0)$, and thus the distribution Q_0 is *locally overidentified* by $\mathcal{Q}$ (Chen and Santos, 2018); further, if $d_m = d_\theta$, then $\mathcal{T}(\mathrm{Q}_0) = L_0^2(\mathrm{Q}_0)$, and thus Q_0 is *locally just identified* by $\mathcal{Q}$.

Similar to (39), the tangent set of the baseline GMM model $\mathcal{Q}^{(1)}$ at Q_0 is characterized by

$$\mathcal{T}^{(1)}(\mathrm{Q}_0) \equiv \left\{ f \in L_0^2 : \lambda^{(1)}(f) \in \mathsf{lin}(D_{11}) \right\}, \text{ with } \lambda^{(1)}(f) \equiv \mathrm{E}^{\mathrm{Q}_0}\left[m^{(1)}(\cdot, \theta_0^{(1)}) f \right], \tag{40}$$

where operator $\lambda^{(1)}(f)$ is a linear operator on $L_0^2(\mathrm{Q}_0)$, linear space $\mathsf{lin}(D_{11})$ is spanned by the column vectors of D_{11}, and $m^{(1)}(\cdot, \theta_0^{(1)})$ contains the top $d_{m,1}$ elements of $m(\cdot, \theta_0)$.

4.3. *Local Instability of Data-Generating Processes*

To formalize the analysis on out-of-sample fit (i.e. external validity), we need to consider data-generating processes that allow for structural breaks in a non-stationary manner. More precisely, we consider the set $\mathcal{M}(\mathrm{Q}_0)$ consisting of all probability measures, each of which is a joint distribution for $\mathbf{y}^n$ following a Markov process with local instability around a Markov process characterized by Q_0. Now, we formalize the definition of $\mathcal{M}(\mathrm{Q}_0)$ as follows.

Definition 4 *The collection $\mathcal{M}(\mathrm{Q}_0)$ contains all join distributions $\mathbb{P}_{1/\sqrt{n},g,b}$, for local Markov data-generating processes, characterized by a sequence of bivariate marginal distributions $\mathrm{Q}_{s,f_{n,t}} \in \mathcal{N}(\mathrm{Q}_0)$ with $t = 1, 2, \cdots, n$ and index $s \in (-\epsilon, \epsilon)$ for some $\epsilon > 0$ such that*

$$f_{n,t} = [1, b(t/n)]g(\mathbf{y}_{t-1}, \mathbf{y}_t) \text{ with } g \in \mathcal{G}(\mathrm{Q}_0) \text{ and } b \in \mathcal{B}. \tag{41}$$

The unique corresponding model parameter value is also time-varying:

$$\theta_{n,t} \equiv \vartheta(\mathrm{Q}_{1/\sqrt{n},f_{n,t}}), \quad \text{for any } f_{n,t} \in \mathcal{T}(\mathrm{Q}_0) \text{ with } 1 \le t \le n \text{ and sufficiently large } n. \tag{42}$$

Definition 4 says that all the local data-generating processes in $\mathcal{M}(\mathrm{Q}_0)$ are characterized by the pair $(g, b) \in \mathcal{G}(\mathrm{Q}_0) \times \mathcal{B}$ and sample size n. The data-generating process is a time-homogeneous Markov process if $b(u) \equiv 0$ or $g_2(\mathbf{y}, \mathbf{y}') \equiv 0$. Assumption 4 ensures the uniqueness of (42).

Local moment biases

Under the local data-generating process $\mathbb{P}_{1/\sqrt{n},g,b}$ characterized by a sequence of bivariate marginal distributions $\mathrm{Q}_{1/\sqrt{n},f_{n,t}}$ for $t = 1, \cdots, n$, the moment restrictions evaluated at θ_0 are locally biased. We summarize the result in Proposition 2 with the proof in Appendix D.

Proposition 2 (Local Biases of Moment Restrictions) *Suppose Assumptions 1 – 5 hold. Under the bivariate marginal distribution $\mathrm{Q}_{1/\sqrt{n},f_{n,t}} \in \mathcal{M}(\mathrm{Q}_0)$ for the consecutive pair $(\mathbf{y}_{t-1}, \mathbf{y}_t)$ where $f_{n,t} = g_1(\mathbf{y}_{t-1}, \mathbf{y}_t) + g_2(\mathbf{y}_{t-1}, \mathbf{y}_t)b(t/n)$ and $(g, b) \in \mathcal{G}(\mathrm{Q}_0) \times \mathcal{B}$, the moment restrictions evaluated at θ_0 are locally biased:*

$$\mathrm{E}^{\mathrm{Q}_{1/\sqrt{n},f_{n,t}}}[m_t(\theta_0)] = [\lambda(g_1) + \lambda(g_2)b(t/n)]/\sqrt{n} + o\left(1/\sqrt{n}\right), \tag{43}$$

where the linear operator $\lambda(\cdot)$ is defined in (39).

5. WHY IS MODEL FRAGILITY A CONCERN?

We specify the notion of model fragility using high sensitivity of moment restrictions to local perturbations in the data-generating process. Using the semiparametric framework in Section 4, we now show that a fragile model lacks internal refutability and external validity. More specifically, we show that our dark matter measure is inversely linked to the power of the C test (i.e. internal refutability) and the out-of-sample fit (i.e. external validity) in Sections 5.1 and 5.2, respectively.

The baseline GMM model plays a special role in the test power and out-of-sample fit analyses as a “benchmark”, characterizing the correct baseline parameter values $\theta_{n,t}^{(1)}$ and discipline the asset pricing cross-equation restrictions.

Assumption 6 (Correct Baseline GMM Model) *We assume that the true local data-generating process with a joint distribution $\mathbb{P}_{1/\sqrt{n},g^*,b^*}$ satisfies*

$$\lambda^{(1)}(g_1^*) = 0 \text{ and } \lambda^{(1)}(g_2^*) \in \mathsf{lin}(D_{11}), \tag{44}$$

where the linear operator $\lambda^{(1)}(\cdot)$ is defined in (40).[18]

Remark 6 *Assumption 6 ensures that the baseline GMM model is correctly specified since $\lambda^{(1)}(f_{n,t}) = \lambda^{(1)}(g_2)b(t/n) \in \mathsf{lin}(D)$ for every $t \in \{1, \cdots, n\}$. If we define*

$$\mathcal{G}_B(\mathrm{Q}_0) \equiv \left\{g \in \mathcal{G}(\mathrm{Q}_0) : \lambda^{(1)}(g_1) = 0 \text{ and } \lambda^{(1)}(g_2) \in \mathsf{lin}(D_{11})\right\}, \tag{45}$$

Assumption 6 can be simply rewritten as $g^* \in \mathcal{G}_B(\mathrm{Q}_0)$.

The following corollary shows the correct baseline parameters are invariant under Assumption 6.

Corollary 1 (Correct Baseline Parameters) *Suppose Assumptions 1 – 6 hold. Then, the correct baseline parameters $\theta_{n,t}^{(1)} \equiv \vartheta^{(1)}(\mathrm{Q}_{1/\sqrt{n},f_{n,t}})$ exists for $f_{n,t} = g_1(\mathbf{y}_{t-1}, \mathbf{y}_t) + g_2(\mathbf{y}_{t-1}, \mathbf{y}_t)b(t/n)$ with $1 \le t \le n$, and they can be approximated by*

$$\theta_{n,t}^{(1)} - \theta_0^{(1)} = -(D_{11}^T D_{11})^{-1} D_{11}^T \lambda^{(1)}(f_{n,t})/\sqrt{n} + o\left(1/\sqrt{n}\right). \tag{46}$$

[18] We can replace (44) by a seemingly weaker assumption $\lambda^{(1)}(g_1^*), \lambda^{(1)}(g_2^*) \in \mathsf{lin}(D_{11})$. But, this does not add generality, because we can always replace θ_0 by a sequence of new reference points (reparametrization) to ensure that (44) is satisfied.

5.1. *Test Power and Dark Matter*

We now establish the connection between the dark matter measure and the local asymptotic maximin power. We focus on a subset of alternatives satisfying $b \equiv 0$ to obtain an upper bound on the maximin local power, which is characterized by the dark matter measure. A specification test for a GMM model $\mathcal{Q}$ against its baseline GMM model $\mathcal{Q}^{(1)}$ is a test of the null hypothesis that there exists some parameter for which all moment restrictions hold under the true data-generating process against the alternative that there exists some parameter for which only the baseline moment restrictions hold. That is,

$$H_0 : \mathrm{Q}_n^* \in \mathcal{Q} \quad vs. \quad H_{\mathcal{A}} : \mathrm{Q}_n^* \in \mathcal{Q}^{(1)} \setminus \mathcal{Q}. \tag{47}$$

Let $\check{\varphi}_n$ be an arbitrary GMM test statistic that maps $\mathbf{y}^n$ to $[0, 1]$ (e.g. Hansen, 1982; Newey, 1985a). We restrict our attention to GMM specification tests $\check{\varphi}_n$ that have local asymptotic level α and possess an asymptotic local power function.[19] More precisely, we consider the local data-generating process $\mathbb{P}_{1/\sqrt{n},g,0}$ for $\mathbf{y}^n$ with a bivariate marginal distribution $\mathrm{Q}_{1/\sqrt{n},g_1}$ that converges to $\mathrm{Q}_0 \in \mathcal{Q}(\theta_0)$ as $n \to \infty$.

The test $\check{\varphi}_n$ has a local asymptotic level α if

$$\limsup_{n\to\infty} \int \check{\varphi}_n \mathrm{d}\mathbb{P}_{1/\sqrt{n},g,0} \leq \alpha, \quad \forall\, g \in \mathcal{G}(\mathrm{Q}_0) \text{ such that } g_1 \in \mathcal{T}(\mathrm{Q}_0), \tag{48}$$

and the test $\check{\varphi}_n$ has a local asymptotic power function $q(g, \check{\varphi})$ if

$$q(g, \check{\varphi}) \equiv \lim_{n\to\infty} \int \check{\varphi}_n \mathrm{d}\mathbb{P}_{1/\sqrt{n},g,0}, \quad \forall\, g \in \mathcal{G}(\mathrm{Q}_0) \text{ such that } g_1 \in \mathcal{T}^{(1)}(\mathrm{Q}_0), \tag{49}$$

where $\check{\varphi} \equiv \{\check{\varphi}_n\}_{n\geq 1}$ is the sequence of test statistics.

Finally, a test $\check{\varphi}_n$ for (47) with a local asymptotic power function $q(\cdot, \check{\varphi}) : \mathcal{G}(\mathrm{Q}_0) \to [0, 1]$ is said to be *locally unbiased* if $q(g, \check{\varphi}) \leq \alpha$ for all g such that $g_1 \in \mathcal{T}(\mathrm{Q}_0)$, and $q(g, \check{\varphi}) \geq \alpha$ for all g such that $g_1 \in L_0^2(\mathrm{Q}_0) \setminus \mathcal{T}(\mathrm{Q}_0)$. We denote the set of locally unbiased GMM specification tests with level α as $\Phi_\alpha(\mathrm{Q}_0)$.

---

[19] As the sample size n approaches infinity, the distance between the null hypothesis and the data-generating process necessarily diminishes according to $n^{-1/2}$. If this distance were held fixed, then the power of all consistent tests would tend to unity as n increases to infinity. Local power analysis, the evaluation of the behavior of the power function of a hypothesis test in a neighborhood of the null hypothesis invented by Neyman (1937), has become an important and commonly utilized technique in econometrics (e.g., Newey, 1985b; Davidson and MacKinnon, 1987; Saikkonen, 1989; McManus, 1991).

The guaranteed local asymptotic power of tests, over all feasible local data-generating processes, can be characterized by the power of maximin tests (e.g., Lehmann and Romano, 1996, Chapter 8). Studies have demonstrated that the C test or incremental J test (e.g., Eichenbaum, Hansen, and Singleton, 1988) has the asymptotic optimality property in the maximin sense (e.g., Newey, 1985a; Chen and Santos, 2018).[20] Based on this observation, we establish Theorem 1 below, which formally connects the maximin optimal power of tests to the dark matter measure. We present the proof in Appendix C.

We consider the set of alternatives:

$$\mathcal{A}_\kappa(\mathrm{Q}_0) \equiv \left\{g \in \mathcal{G}_{\mathrm{B}}(\mathrm{Q}_0) : |\lambda^{(2)}(g_1)| \geq \kappa \text{ and } \lambda^{(2)}(g_1) \perp \mathsf{lin}(D_{22})\right\} \tag{50}$$

where $\lambda^{(2)}(g_1) \equiv \mathrm{E}^{\mathrm{Q}_0}\left[m^{(2)}(\cdot,\theta_0)g_1\right]$ is the bottom $d_m - d_{m,1}$ elements of $\lambda(g_1)$ defined in (39).

THEOREM 1 *Suppose Assumptions 1 – 6 hold. The local asymptotic power of maximin tests is bounded above by*

$$\sup_{\check{\varphi}\in\Phi_\alpha(\mathrm{Q}_0)} \inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} q(g,\check{\varphi}) \leq M_{\frac{d_{m,2}-d_{\theta,2}}{2}}\left(\sqrt{\frac{\kappa^2}{1+\varrho(\theta_0)}}, \sqrt{c_{1-\alpha}}\right), \tag{51}$$

where $c_{1-\alpha}$ is the $1-\alpha$ quantile of a chi-square distribution with degrees of freedom $d_{m,2} - d_{\theta,2}$, and $M_\gamma(x_1, x_2)$ is the generalized Marcum Q-function. By definition of $c_{1-\alpha}$, it holds that

$$M_{\frac{d_{m,2}-d_{\theta,2}}{2}}(0, \sqrt{c_{1-\alpha}}) = \alpha. \tag{52}$$

Therefore, the local asymptotic power of maximin tests vanishes as the dark matter measure rises:

$$\sup_{\check{\varphi}\in\Phi_\alpha(\mathrm{Q}_0)} \inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} q(g,\check{\varphi}) \to \alpha, \text{ as } \varrho(\theta_0) \to \infty. \tag{53}$$

The generalized Marcum Q-function $M_\gamma(x_1, x_2)$ is strictly increasing in γ and x_1, and it is strictly decreasing in x_2 (e.g., Sun, Baricz, and Zhou, 2010, Theorem 1). Intuitive interpretation of (51) is that there exists a difficult specific alternative characterized by the score $g \in \mathcal{A}_\kappa(\mathrm{Q}_0)$ such that the power of the optimal locally unbiased GMM specification test with level α cannot

[20] Alternative asymptotically equivalent approaches can be found in the literature (e.g., Newey, 1985a; Chen and Santos, 2018).

exceed $M_{\frac{d_{m,2}-d_{\theta,2}}{2}}\left(\sqrt{\frac{\kappa^2}{1+\varrho(\theta_0)}}, \sqrt{c_{1-\alpha}}\right)$, which is almost α when $\varrho(\theta_0)$ is extremely large. The coefficient κ captures the extent to which the alternatives under consideration are separate from the null, and thus, the upper bound for the test power (i.e., the right-hand side of (51)) naturally increases with κ.

5.2. *Overfitting Tendency and Dark Matter*

A common method adopted by economists and statisticians for assessing the external validity of models is to hold out data from the model estimation (e.g., Schorfheide and Wolpin, 2012). The assessment of external validity serves two important purposes: first, it mitigates the concern of in-sample overfitting (e.g., Foster, Smith, and Whaley, 1997; Kocherlakota, 2007; Lettau and Van Nieuwerburgh, 2008; Welch and Goyal, 2008; Koijen and Van Nieuwerburgh, 2011; Ferson, Nallareddy, and Xie, 2013; Athey and Imbens, 2017, 2019); and second, it serves as a primary criterion when the goal is long-run prediction (e.g., Valkanov, 2003; Müller and Watson, 2016). The literature has emphasized that out-of-sample fit evaluation captures model specification uncertainty, model instability, calibration uncertainty, and estimation uncertainty, in addition to the usual uncertainty of future events (Stock and Watson, 2008).

The holdout approach to selecting a model among competing structural models amounts to splitting the entire time series $\mathbf{y}^n \equiv \{\mathbf{y}_1, \cdots, \mathbf{y}_n\}$ into two non-overlapping subsamples $\mathbf{y}^n_{\mathrm{e}} \equiv \left\{\mathbf{y}_1, \cdots, \mathbf{y}_{\lfloor \pi n \rfloor}\right\}$ and $\mathbf{y}^n_{\mathrm{o}} \equiv \left\{\mathbf{y}_{\lfloor \pi n \rfloor + 1}, \cdots, \mathbf{y}_n\right\}$ with $\pi \in (0, 1/2]$.[21] Here, $\lfloor x \rfloor$ is the largest integer less than or equal to the real number x. The first segment $\mathbf{y}^n_{\mathrm{e}}$ is used as the estimation sample, while the second segment $\mathbf{y}^n_{\mathrm{o}}$ is used as the holdout sample (e.g., Schorfheide and Wolpin, 2012).[22] This approach has been commonly adopted in the literature on forecasting and model selection. Further, the holdout approach is also a natural way to investigate the long-run forecast problems in financial and macroeconomic time series, because the salient definition of a long-run forecast is that the prediction horizon is long relative to the sample length of the estimation sample (Müller and Watson, 2016, Section 5.2).

To consider the out-of-sample fit of estimated time-series models, we focus on model instability, because the constant misspecification over time would not affect the out-of-sample fit of estimated models based on in-sample data. Thus, in this subsection we focus on the case with

[21] We specify an upper bound for π to prevent the out-of-sample fit problem from becoming trivial. Without loss of generality, we choose the upper bound for π to be 1/2.

[22] The non-overlapping equal-length estimation and holdout subsamples are standard exercises in cross-validation for out-of-sample fit evaluation; in the statistics and machine learning literature, $\mathbf{y}^n_{\mathrm{e}}$ is also referred to as training sample, and $\mathbf{y}^n_{\mathrm{o}}$ as testing sample (e.g., Hastie, Tibshirani, and Friedman, 2001, Chapter 7).

$\lambda(g_1) \in \mathsf{lin}(D)$.

The idea is to quantify the overfitting tendency as a model property by focusing on the J statistic as the loss function. We define $\theta^{(1)}_{n,t} \equiv \vartheta^{(1)}(\mathrm{Q}_{1/\sqrt{n},f_{n,t}})$ for $t = 1, \cdots, n$, and

$$\theta^{(1)}_{\mathrm{e,n}} \equiv \frac{1}{\lfloor \pi n \rfloor} \sum_{t=1}^{\lfloor \pi n \rfloor} \theta^{(1)}_{n,t} \quad \text{and} \quad \theta^{(1)}_{\mathrm{o,n}} \equiv \frac{1}{\lfloor (1-\pi) n \rfloor} \sum_{t=\lfloor \pi n \rfloor + 1}^{n} \theta^{(1)}_{n,t}. \tag{54}$$

More precisely, we consider the goodness-of-fit of the full set of moments under any given baseline parameters $\theta^{(1)}$:

$$\mathcal{L}(\theta^{(1)}; \mathbf{y}^n_s) \equiv J(\theta^{(1)}, \psi_s(\theta^{(1)}), \mathbf{y}^n_s) - J(\theta^{(1)}_{\mathrm{s,n}}, \psi_s(\theta^{(1)}_{\mathrm{s,n}}), \mathbf{y}^n_s), \text{ with } s \in \{\mathrm{e, o}\} \tag{55}$$

where $\theta^{(1)}_{\mathrm{s,n}}$ is the average of those correct baseline parameter values that perfectly fit baseline moment restrictions (see (46)) and $\psi_s(\theta^{(1)})$ is chosen to minimize the J statistic while taking $\theta^{(1)}$ as given:[23]

$$\psi_s(\theta^{(1)}) \equiv \operatorname*{argmin}_{\theta^{(2)}} J(\theta^{(1)}, \theta^{(2)}, \mathbf{y}^n_s) \text{ for any fixed } \theta^{(1)} \text{ with } s \in \{\mathrm{e, o}\}. \tag{56}$$

In the definition of $\mathcal{L}(\theta^{(1)}; \mathbf{y}^n_s)$, we benchmark the goodness-of-fit measure against the J statistic evaluated at the average of correct baseline parameter values $\theta^{(1)}_{\mathrm{s,n}}$ for $s \in \{e, o\}$ to control the mechanical influence of instability on the J statistic. The lower the goodness-of-fit measure $\mathcal{L}(\theta^{(1)}; \mathbf{y}^n_s)$, the better the baseline parameter value $\theta^{(1)}$ fits the moments in the sample $\mathbf{y}^n_s$ with $s \in \{\mathrm{e, o}\}$. Importantly, by minimizing over all possible values of the nuisance parameters $\theta^{(2)}$, the measure $\mathcal{L}(\theta^{(1)}; \mathbf{y}^n_s)$ captures the best possible fit of the parameter value $\theta^{(1)}$ only.

We consider a GMM estimator of the baseline parameters $\theta^{(1)}$, denoted by $\check{\theta}^{(1)}_{\mathrm{e,n}}$, based on the estimation sample $\mathbf{y}^n_{\mathrm{e}}$ and all moment restrictions. We then assess the out-of-sample fit of $\check{\theta}^{(1)}_{\mathrm{e,n}}$ on the holdout sample by looking at the magnitude of the expected out-of-sample fitting error $\int \mathcal{L}(\check{\theta}^{(1)}_{\mathrm{e,n}}, \mathbf{y}^n_{\mathrm{o}}) \mathrm{d}\mathbb{P}_{1/\sqrt{n},g,b}$. The overfitting measure of the estimator $\check{\theta}^{(1)}_{\mathrm{e,n}}$ can be defined as the extent to which the out-of-sample fitting error is larger than the in-sample fitting error:

$$\mathcal{O}(\check{\theta}^{(1)}_{\mathrm{e,n}}, \mathbf{y}^n) \equiv \frac{1}{2} \left[\mathcal{L}(\check{\theta}^{(1)}_{\mathrm{e,n}}, \mathbf{y}^n_{\mathrm{o}}) - \mathcal{L}(\check{\theta}^{(1)}_{\mathrm{e,n}}, \mathbf{y}^n_{\mathrm{e}}) \right]. \tag{57}$$

[23]Mathematically, the formulation (55) and (56) follow the generic recursive GMM estimation procedure in Hansen (2007b) and Hansen (2012).

The asymptotic expected overfitting measure of the sequence of estimators $\check{\theta}^{(1)}_{\mathrm{e,n}}$ is [24]

$$\omega(g, b, \check{\theta}^{(1)}_{\mathrm{e}}) \equiv \lim_{l\to\infty} \lim_{n\to\infty} \int \mathcal{O}(\check{\theta}^{(1)}_{\mathrm{e,n}}, \mathbf{y}^n) \mathbf{1}_{\{|\mathcal{O}(\check{\theta}^{(1)}_{\mathrm{e,n}}, \mathbf{y}^n)| \le l\}} d\mathbb{P}_{1/\sqrt{n},g,b}, \tag{58}$$

where $\check{\theta}^{(1)}_{\mathrm{e}} \equiv \left\{\check{\theta}^{(1)}_{\mathrm{e,n}}\right\}_{n\ge 1}$ is a sequence of GMM estimators. $\omega(g, b, \check{\theta}^{(1)}_{\mathrm{e}})$ quantifies the extent to which the structural model over-fits the data when the true local data-generating process is $\mathbb{P}_{1/\sqrt{n},g,b}$. Similar in spirit to information criteria in model selection such as AIC and BIC, models whose expected overfitting measures are sizable compared with those of other models that fit the sample data equally well in sample should be penalized.

Two types of estimators

Here we focus on two particular estimation procedures – the efficient GMM estimation procedure and the recursive GMM estimation procedure.[25] The former is designed to use the identification strength provided by the additional asset pricing moment restrictions $\mathrm{E}^{\mathrm{Q}_0}\left[m_t^{(2)}(\theta)\right] = 0$ as much as possible, while the latter does not use any identification assumptions imposed by the additional asset pricing moment restrictions $\mathrm{E}^{\mathrm{Q}_0}\left[m_t^{(2)}(\theta)\right] = 0$. The identification strength is a nontestable assumption postulated by the structural model. The literature on recursive GMM estimation is substantial and dates back decades (e.g., Christiano and Eichenbaum, 1992; Ogaki, 1993; Newey and McFadden, 1994; Hansen and Heckman, 1996; Hansen, 2007b; Lee, 2007; Hansen, 2012). While the original impetus of the recursive GMM estimation was primarily computational, we advocate it as a more robust procedure against potential instability and misspecification since the procedure barely relies on the nontestable assumption of identification strength of the additional moment restrictions $m_t^{(2)}(\theta)$; the robustness of the recursive GMM estimation procedure is especially valuable when the dark matter measure is excessively large.

Characterized by selection matrices, the efficient GMM estimator and the recursive GMM estimator based on the estimation sample $\mathbf{y}^n_{\mathrm{e}}$, denoted by $\hat{\theta}_{\mathrm{e,n}}$ and $\tilde{\theta}_{\mathrm{e,n}}$ respectively, have the selection matrices $A = D$ and $A = \mathrm{diag}\{D_{11}, A_{22}\}$ with the (constrained) efficient selection matrix $A_{22} \equiv \left[D_{21}(D_{11}^T D_{11})^{-1} D_{21}^T + I\right]^{-1} D_{22}$ (Hansen, 2007b).

[24]The method of first calculating the truncated statistic, then letting the ceiling l increase to infinity, is commonly adopted in the literature for technical simplification (e.g. Bickel, 1981; Le Cam and Yang, 2000; Kitamura, Otsu, and Evdokimov, 2013).

[25]The recursive GMM estimation procedure is also referred to as the sequential (two-step) GMM estimation procedure in the literature.

Overfitting of the efficient GMM estimator $\hat{\theta}_{e,n}$

Recall that $\lambda(g_2)$ captures the magnitude of instability. We consider the set of possible magnitudes of instability $\mathcal{U}_\kappa(\mathrm{Q}_0) \equiv \{g \in \mathcal{G}_{\mathrm{B}}(\mathrm{Q}_0) : |\lambda(g_2)| \leq \kappa\}$. A larger κ allows for a higher degree of instability in the fragility analysis.

THEOREM 2 *Suppose Assumptions 1 – 6 hold. The overfitting of the efficient GMM estimator* $\hat{\theta}_{e,n}$ *based on the estimation sample* $\mathbf{y}_e^n$ *is defined as the worst-case asymptotic expected overfitting measure. It is characterized by the dark matter measure:*

$$\sup_{g\in\mathcal{U}_\kappa(\mathrm{Q}_0),b\in\mathcal{B}} \omega(g,b,\hat{\theta}_e^{(1)}) = d_{\theta,1} + c(\pi)\varrho(\theta_0)\kappa^2, \tag{59}$$

where $c(\pi) \equiv \pi\left(1+\sqrt{\frac{\pi}{1-\pi}}\right)$ *with* $0 < \pi \leq 1/2$. *Therefore, the asymptotic expected overfitting of the efficient GMM estimator sequence* $\hat{\theta}_e \equiv \left\{\hat{\theta}_{e,n}\right\}_{n\geq 1}$ *can be arbitrarily large, being linearly related to the dark matter measure.*

The overfitting of the efficient GMM estimator has two sources. The first term $d_{\theta,1}$ captures the traditional overfitting due to the sampling uncertainty in the estimation sample (see Theorem 3), while the second term $c(\pi)\varrho(\theta_0)\kappa^2$ captures the overfitting due to potential misspecification and instability. The latter component is the focus of our paper and directly depends on the dark matter measure. The second component, $c(\pi)\varrho(\theta_0)\kappa^2$, vanishes if there is no local instability (i.e., $\kappa = 0$).

Overfitting of the recursive GMM estimator $\tilde{\theta}_{e,n}$

THEOREM 3 *Suppose Assumptions 1 – 6 hold. The overfitting of the recursive GMM estimator* $\tilde{\theta}_{e,n}$ *based on the estimation sample* $\mathbf{y}_e^n$ *is defined as the worst-case asymptotic expected overfitting measure, which only depends on the number of baseline parameters:*

$$\sup_{g\in\mathcal{U}_\kappa(\mathrm{Q}_0),b\in\mathcal{B}} \omega(g,b,\tilde{\theta}_e^{(1)}) = d_{\theta,1}, \tag{60}$$

where $\mathcal{U}_\kappa(\mathrm{Q}_0) \equiv \{g \in \mathcal{G}_B(\mathrm{Q}_0) : |\lambda(g_2)| \leq \kappa\}$. *Therefore, the overfitting of the recursive GMM estimator sequence* $\tilde{\theta}_e \equiv \left\{\tilde{\theta}_{e,n}\right\}_{n\geq 1}$ *is determined by model parameter dimensionality, not affected by the dark matter of the model.*

The results above echo the traditional information criteria such as AIC and BIC, where the number of parameters captures the overfitting tendency due to sampling uncertainty. The recursive GMM estimator is not affected by the nontestable assumptions of identification strength imposed by $\mathrm{E}^{\mathrm{Q}_0}\left[m_t^{(2)}(\theta)\right] = 0$, thus its overfitting is not affected by the misspecified (local) instability. Importantly, Theorem 3 suggests that the recursive GMM estimator provides a robust estimator for models with large dark matter measures (i.e. large $\varrho(\theta_0)$) and thus subject to severe (local) instability concerns (i.e. large $\mathcal{U}_\kappa(\mathrm{Q}_0)$).

Instability of the efficient GMM estimator

Intuitively, the formal results about out-of-sample fit above can be appreciated through the sensitivity of efficient GMM estimators to local instability (see Panel C of Figure 1). We consider a local perturbation of the model from Q_0 in the direction of $g \in \mathcal{G}_{\mathrm{B}}(\mathrm{Q}_0)$ with instability $b \in \mathcal{B}$. According to Proposition 6 (in Appendix A),

$$\begin{bmatrix} \frac{1}{\sqrt{\pi n}} \sum_{t \le \pi n} m_t(\theta_0) \\ \frac{1}{\sqrt{(1-\pi)n}} \sum_{t > \pi n} m_t(\theta_0) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} m_e \\ m_o \end{bmatrix}, \text{ with } \mathrm{E}\begin{bmatrix} m_e \\ m_o \end{bmatrix} = \begin{bmatrix} \nu_e(g,b,\pi) \\ \nu_o(g,b,\pi) \end{bmatrix}, \tag{61}$$

where (m_e, m_o) are independent normals with the identity covariance matrix and means:

$$\nu_e(g,b,\pi) \equiv \frac{\lambda(g^T)}{\sqrt{\pi}} \begin{bmatrix} \pi \\ \int_0^\pi b(u)\mathrm{d}u \end{bmatrix} \text{ and } \nu_o(g,b,\pi) \equiv \frac{\lambda(g^T)}{\sqrt{1-\pi}} \begin{bmatrix} 1-\pi \\ \int_\pi^1 b(u)\mathrm{d}u \end{bmatrix}. \tag{62}$$

Further, Proposition 9 (in Appendix A) shows that the in- and out-of-sample estimators satisfy

$$\begin{bmatrix} \sqrt{\pi n}(\hat{\theta}_{\mathrm{e,n}}^{(1)} - \theta_{\mathrm{e,n}}^{(1)}) \\ \sqrt{(1-\pi)n}(\hat{\theta}_{\mathrm{o,n}}^{(1)} - \theta_{\mathrm{o,n}}^{(1)}) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \hat{\theta}_e^{(1)} \\ \hat{\theta}_o^{(1)} \end{bmatrix}, \text{ with } \mathrm{E}\begin{bmatrix} \hat{\theta}_e^{(1)} \\ \hat{\theta}_o^{(1)} \end{bmatrix} = -(L_{\mathrm{F}} - L_{\mathrm{B}}) \begin{bmatrix} \nu_e(g,b,\pi) \\ \nu_o(g,b,\pi) \end{bmatrix}, \tag{63}$$

where $L_{\mathrm{B}} \equiv \mathbf{I}_{\mathrm{B}}^{-1} D_{11}^T \Gamma_{m,1}$, $L_{\mathrm{F}} \equiv \Gamma_{\theta,1} \mathbf{I}_{\mathrm{Q}}^{-1} D^T$, and $(\hat{\theta}_e^{(1)}, \hat{\theta}_o^{(1)})$ are independent normals with covariance matrix $\mathbf{I}_{\mathrm{F}}^{-1}$. Therefore, the amount of estimator instability (normalized by covariance matrix $\mathbf{I}_{\mathrm{F}}^{-1}$) as a function of moment instability is

$$\mathbf{I}_{\mathrm{F}}^{1/2} \mathrm{E}\left[\hat{\theta}_{\mathrm{e,n}}^{(1)} - \hat{\theta}_{\mathrm{o,n}}^{(1)}\right] = \beta \mathrm{E}\left[m_o - m_e\right], \tag{64}$$

where $\beta = -\mathbf{I}_{\text{F}}^{1/2}(L_{\text{F}} - L_{\text{B}})$. The largest sensitivity can be captured by the spectral norm of the sensitivity matrix β; that is $||\beta||_{\mathbb{S}} = \sqrt{\varrho(\theta_0)}$. Thus, a large dark matter measure implies high sensitivity in the form of severe instability of the efficient GMM estimator out of sample versus in sample.

This result resembles that of Andrews, Gentzkow, and Shapiro (2017), but there are two key differences. When there is no nuisance parameter (i.e., $\theta^{(1)} = \theta$), L_{F} is the same as the sensitivity matrix in Andrews, Gentzkow, and Shapiro (2017). Relative to their measure, we add a baseline GMM model as a benchmark (replacing L_{F} by $L_{\text{F}} - L_{\text{B}}$), and we normalize the expected change in the efficient GMM estimator by its asymptotic covariance matrix in the full model (multiplying $\mathrm{E}[\hat{\theta}^{(1)}_{\text{e,n}} - \hat{\theta}^{(1)}_{\text{o,n}}]$ by $\mathbf{I}_{\text{F}}^{1/2}$).

6. EXAMPLES AND SIMULATION STUDIES

We now use the dark matter measure to analyze a rare disaster model, one of the leading consumption-based asset pricing models. We analyze a long-run risk model in Appendix H.

Rare economic disasters are a natural source of "dark matter" in asset pricing models. It is difficult to evaluate the likelihood and magnitude of rare disasters statistically. Yet, agents' aversion to large disasters can have large ex-ante effects on asset prices. In this subsection, we use our measure to analyze a disaster risk model similar to Barro (2006).

The model specifies the joint dynamics of the log growth rate of aggregate consumption (endowment) g_t and the excess log return on the market portfolio r_t. There is an observable state variable z_t, which follows an i.i.d. Bernoulli distribution and is equal to one with probability p. When $z_t = 1$, the economy is in a disaster regime, while the normal regime corresponds to $z_t = 0$. In the normal regime, the log consumption growth $g_t = u_t$, which is i.i.d. normal, $u_t \sim N(\mu, \sigma^2)$. In a disaster state, $g_t = -v_t$, where v_t follows a truncated exponential distribution with density

$$\nu_t \overset{\text{i.i.d.}}{\sim} \mathbf{1}\{v_t > \underline{v}\}\xi e^{-\xi(v_t - \underline{v})}. \tag{65}$$

Here the lower bound for disaster size is $\underline{v}$ and the average disaster size is $\underline{v} + 1/\xi$.

The joint distribution of log consumption growth g_t and excess log return r_t changes with the underlying state z_t. When the economy is in the normal regime ($z_t = 0$), g_t and r_t are jointly normal, and

$$r_t = \eta + \rho\frac{\tau}{\sigma}(g_t - \mu) + \sqrt{1-\rho^2}\tau\varepsilon_{0,t}, \tag{66}$$

where $\varepsilon_{0,t}$ is i.i.d. standard normal. The parameter τ is the return volatility in the normal regime, while ρ is the correlation between return and consumption growth in this regime. When the economy is in a disaster state ($z_t = 1$), $g_t = -v_t$, and

$$r_t = \ell g_t + \varsigma\varepsilon_{1,t}, \tag{67}$$

where $\varepsilon_{1,t}$ is i.i.d. standard normal.

Next, we assume that the representative agent has a constant relative risk aversion utility function $u_t(c_t) = \delta_{\mathrm{D}}^t c_t^{1-\gamma_{\mathrm{D}}}/(1-\gamma_{\mathrm{D}})$, where $\gamma_{\mathrm{D}} > 0$ is the coefficient of relative risk aversion and $\delta_{\mathrm{D}} < 1$ is the time preference parameter. The log equity premium, $\overline{r} \equiv \mathbb{E}[r_t]$, is available in closed form (see Appendix F for details) as follows:

$$\overline{r}(p,\xi) = (1-p)\eta - p\ell\left(\underline{v} + 1/\xi\right), \quad \text{where} \tag{68}$$

$$\eta \approx \gamma_{\mathrm{D}}\rho\sigma\tau - \frac{\tau^2}{2} + e^{\gamma_{\mathrm{D}}\mu - \frac{\gamma_{\mathrm{D}}^2\sigma^2}{2}}\Delta(\xi)\frac{p}{1-p}, \text{ with } \Delta(\xi) = \xi\left(\frac{e^{\gamma_{\mathrm{D}}\underline{v}}}{\xi-\gamma_{\mathrm{D}}} - \frac{e^{\frac{\varsigma^2}{2}+(\gamma_{\mathrm{D}}-\ell)\underline{v}}}{\xi+\ell-\gamma_{\mathrm{D}}}\right). \tag{69}$$

The term η in (68) is the log equity premium in the normal regime. The first two terms of η in (69) describe the market risk premium due to Gaussian consumption shocks; the third term is the disaster risk premium, which explodes as ξ approaches γ_{D} 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.

The fact that the equity premium explodes as ξ approaches γ_{D} is an important feature of our version of the disaster risk model. No matter how rare the disasters are (i.e., a very small p), an arbitrarily large equity premium can be generated as long as the average disaster size is sufficiently large (or equivalently, ξ is sufficiently small). Extremely rare but large disasters can be consistent with the data in the sense that they are difficult to rule out based on the observable data (and standard statistical tests). Below we illustrate how our dark matter measure can detect and quantify the fragility of these models.

To apply our framework to the disaster risk model, we first formulate the economic model above as a GMM model $\mathcal{Q}$ with the (transformed) moments:

$$m_t(\theta) = \Omega(\theta)^{-1/2} \begin{bmatrix} z_t - p \\ g_t - (1-z_t)\mu + z_t(\underline{v} + 1/\xi) \\ r_t - (1-z_t)\left[\eta + \rho\frac{\tau}{\sigma}(g_t - \mu)\right] - z_t \ell g_t \end{bmatrix}. \tag{70}$$

The first two moments in $m_t(\theta)$ are the baseline moments, the third is the asset pricing moment, and $\Omega(\theta)$ is the asymptotic covariance matrix of the untransformed moments. To simplify the example, we focus on the parameters $\theta = (p, \xi)'$ when constructing the dark matter measure, while treating the parameters $(\gamma_{\mathrm{D}}, \mu, \sigma, \underline{v}, \tau, \rho, \ell, \varsigma)$ as auxiliary parameters fixed at known values, making them a part of the functional-form specification. In other words, the nuisance parameter vector $\theta^{(2)}$ is empty in this example.

Based on the approximation (69), the dark matter measure is (see Appendix F for details):

$$\varrho(\theta) \approx 1 + \frac{p\Delta(\xi)^2 + p(1-p)\xi^2\dot{\Delta}(\xi)^2}{(1-\rho^2)\tau^2(1-p)^2} e^{2\gamma_{\mathrm{D}}\mu - \gamma_{\mathrm{D}}^2\sigma^2}, \tag{71}$$

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

$$\dot{\Delta}(\xi) = -\frac{e^{\gamma_{\mathrm{D}}\underline{v}}\gamma_{\mathrm{D}}}{(\xi - \gamma_{\mathrm{D}})^2} + \frac{e^{(\gamma_{\mathrm{D}} - \ell)\underline{v}}(\gamma_{\mathrm{D}} - \ell)}{(\xi - \gamma_{\mathrm{D}} + \ell)^2} e^{\varsigma^2/2}. \tag{72}$$

All else equal, when ξ approaches γ_{D}, both $\Delta(\xi)$ and $\dot{\Delta}(\xi)$ approach infinity, which suggests that disaster risk models featuring large but rare disasters (i.e., small ξ and small p) will be more fragile according to our measure.

Quantitative analysis

To take the model to the data, we use annual real per-capita consumption growth (nondurables and services) from the National Income and Product Accounts (NIPA) and returns on the CRSP value-weighted market portfolio from 1929 to 2011. We fix the following auxiliary parameters at the values of the corresponding moments of the empirical distribution of consumption growth and excess stock returns: $\mu = 1.87\%$, $\sigma = 1.95\%$, $\tau = 19.14\%$, $\varsigma = 34.89\%$ and $\rho = 0.59$. The lower bound for disaster size is set to $\underline{v} = 7\%$, and the leverage factor in the disaster regime is $\ell = 3$.

In Figure 3, we plot the 95% and 99% confidence regions for (p, ξ) based on the baseline

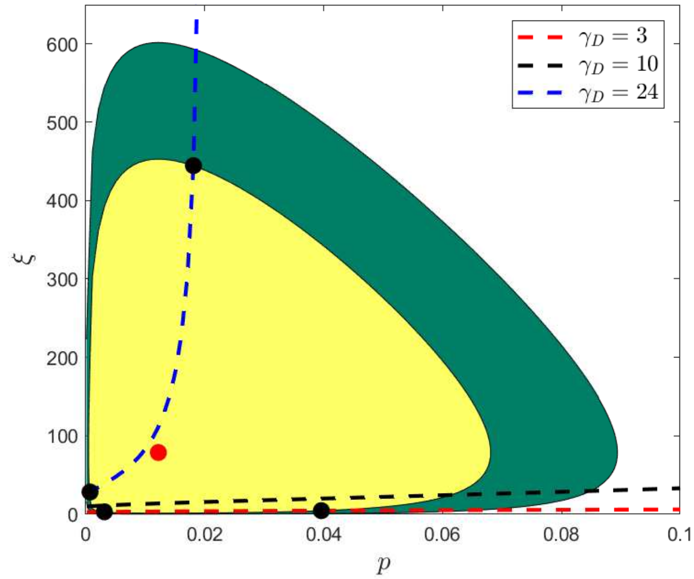

FIGURE 3.— The 95% and 99% confidence regions of (p, ξ) for the baseline model and the equity premium isoquants implied by the asset pricing moment restriction (68) for $\gamma_{\mathrm{D}} = 3, 10, 24$. p is disaster probability, and ξ characterizes the inverse of average disaster size. The efficient GMM estimates are $(\widehat{p}, \widehat{\xi}) = (0.012, 78.79)$, indicated by the red dot inside the confidence region. Four additional points mark the intersections of the equity premium isoquants for $\gamma_{\mathrm{D}} = 3$ and 24 and the boundary of the 95% confidence region. Only p and ξ are treated as unknown to the econometrician, and all other parameters are treated as auxiliary parameters with fixed known values; therefore, the dark matter measure is defined only based on $\theta = (p, \xi)$.

model. As expected, the confidence regions are large, which confirms that the baseline model provides limited information about p and ξ. We also plot the equity premium isoquants: for a given level of risk aversion γ_D, each dashed line in Figure 3 shows the different calibrations of the disaster risk model that match the unconditional equity premium of 5.09%. In particular, even for low risk aversion (e.g., $\gamma_{\mathrm{D}} = 3$), there exist models that not only match the observed equity premium, but are also consistent with the macro data in the sense that the model parameters (p, ξ) remain inside the 95% confidence region.[26]

While it is difficult to distinguish among a wide range of calibrations based on the fit with the macro data, these calibrated models can differ vastly based on our dark matter measure. For illustration, we focus on the following four calibrations, which are the four points where the equity premium isoquants for $\gamma_{\mathrm{D}} = 3$ and 24 intersect the boundary of the 95% confidence region in Figure 3. For $\gamma_{\mathrm{D}} = 3$, the two points are $(p = 3.96\%, \xi = 4.65)$ and $(p = 0.31\%, \xi = 3.179)$.

[26] Julliard and Ghosh (2012) estimate the consumption Euler equation using the empirical likelihood method and show that the model requires a high level of relative risk aversion to match the equity premium. Their empirical likelihood criterion rules out any large disasters that have not occurred in the historical sample, hence requiring the model to generate high equity premium using moderate disasters.

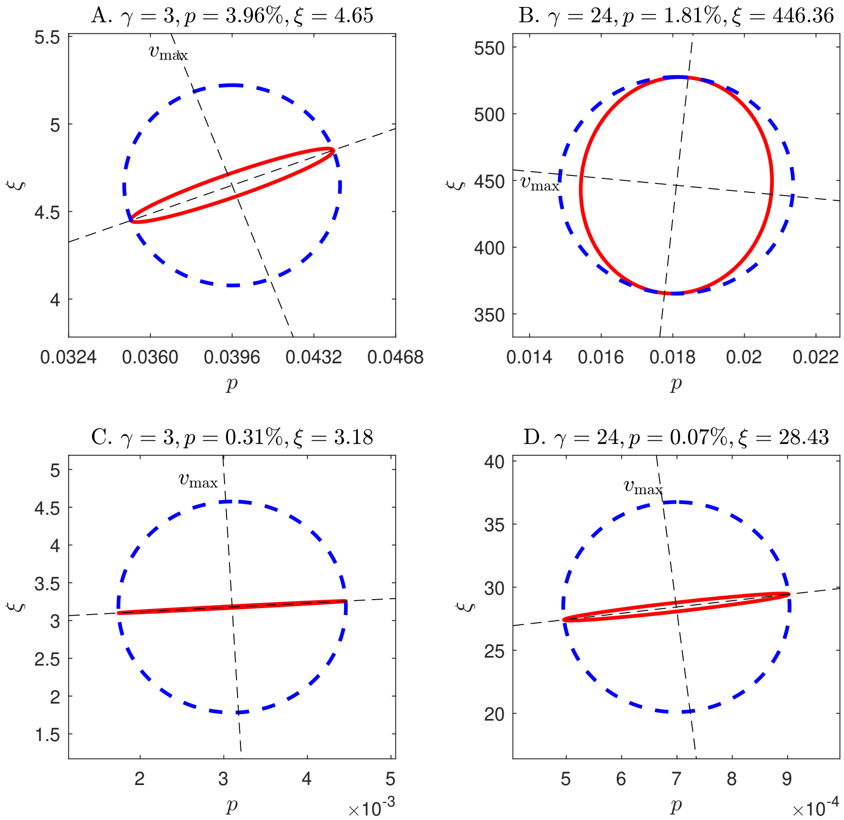

FIGURE 4.— Visualization of the dark matter measure through the 95% confidence regions for the asymptotic distribution of the efficient GMM estimators for four "acceptable" calibrations. In Panels A through D, the dark matter measures are $\varrho(\theta) = 74.03$, 1.49, $1.78 \cdot 10^4$, and $5.60 \cdot 10^2$, respectively, which are obtained in the direction marked by the vector $v_{\max}$. Parameter p is disaster probability, and ξ characterizes the inverse of average disaster size. Only p and ξ are treated as unknown to the econometrician, and all other parameters are treated as auxiliary parameters with fixed known values; therefore, the dark matter measure is defined only based on $\theta = (p, \xi)$.

For $\gamma_{\mathrm{D}} = 24$, the two points are $(p = 1.81\%, \xi = 446.36)$ and $(p = 0.07\%, \xi = 28.43)$.

With just two parameters in $\theta = (p, \xi)$, we can visualize the dark matter measure by plotting the asymptotic confidence regions for (p, ξ) in the baseline model and the full model, as determined by the respective information matrices $\mathbf{I}_{\mathrm{B}}$ and $\mathbf{I}_{\mathrm{F}}$. In each panel of Figure 4, the largest dashed-line circle marks the 95% asymptotic confidence region for (p, ξ) under the baseline model. The smaller solid-line ellipse indicates the 95% asymptotic confidence region for (p, ξ) under the full model. Intuitively, the direction in Figure 4, along which the asset pricing restriction does not provide additional information about the parameters $\theta = (p, \xi)$, is parallel to the tangent direction of the dashed lines (i.e., the equity premium isoquants) in Figure 3, evaluated at the black dots. This highlights the straightforward fact that the structural restriction does not increase informativeness in the direction along which the equity premium does not change.

In Panel A of Figure 4, the dark matter measure is $\varrho(\theta) = 74.07$. This means that under the baseline model, we need to increase the amount of consumption data by a factor of 74.07 to match or exceed the precision in estimation of any linear combination of p and ξ afforded by the equity premium constraint. Panels C and D of Figure 4 correspond to the calibrations with "extra rare and large disasters," and for $\gamma_{\text{D}} = 3$ and 24, the dark matter measure $\varrho(\theta)$ rises to 1.78×10^4 and 5.60×10^2, respectively. If, in Panel B of Figure 4, we raise γ_{D} to 24 while changing the annual disaster probability to 1.81% and lowering the average disaster size to 7.002% ($\xi = 446.36$), the dark matter measure $\varrho(\theta)$ declines to 1.49. The reason for the reduced fragility in this calibration is the combination of a higher disaster probability and a lower average disaster size.

Monte Carlo experiments

We use simulations to illustrate the connection between the dark matter measure and the model fragility (i.e., the internal refutability and external validity) of disaster risk models in finite samples. More precisely, we assume that the true local data-generating process has a time-varying relation between the expected log excess return and other dynamic parameters:

$$\overline{r}_n = \overline{r}(p_0, \xi_0) + \frac{\iota_t \delta_r}{\sqrt{n}}, \text{ with } \iota_t = \begin{cases} 1, & \text{when } 1 \leq t \leq \lfloor \pi n \rfloor \\ -1, & \text{when } \lfloor \pi n \rfloor < t \leq n, \end{cases} \tag{73}$$

where the time series ι_t captures the structural breaks and $\pi \in (0, 1/2]$ is the break point. Such a simple process ι_t characterizes one structural break in the middle of the time-series sample. The corresponding moment biases, evaluated at θ_0, are

$$\mathrm{E}^{\mathrm{Q}_0}\left[m_t(\theta_0)\right] = \left[0, 0, \lambda_t^{(2)}/\sqrt{n}\right]^T \text{ with } \lambda_t^{(2)} \equiv \frac{\iota_t \delta_r}{\sqrt{(1-p_0)(1-\rho^2)\tau^2 + p_0 \varsigma^2}}. \tag{74}$$

Therefore, data-generating processes A and C in Figure 4 have identical moment's local biases $\lambda_t^{(2)}$ after substituting the calibrated parameter values into (74), which guarantees that the comparisons across models in Panels A and B of Figure 5 are valid.

Figure 5 shows three different simulation experiments. Panel A displays the local power functions of C tests. The solid and dotted curves reflect the test powers when the data-generating processes are characterized by calibrations A and C in Figure 4, respectively. In this experiment, we vary the local misspecification δ_r in the risk premium moment restriction. The data-

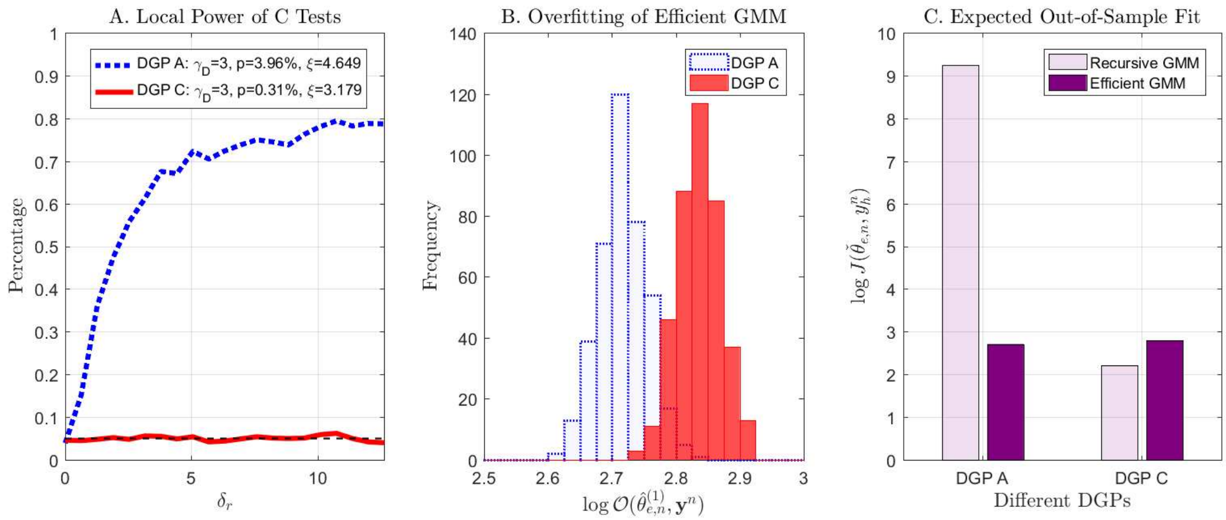

FIGURE 5.— Monte Carlo experiments for disaster risk models. In Panel A, we simulate 1000 independent yearly time series with length $n = 100$ (i.e. 100 years). In Panels B and C, we set $\delta_r = 0.4$, and simulate 400 independent yearly time series with length $n = 100$ (i.e. 100 years) and break point $\pi = 1/2$. Only p and ξ are treated as unknown to the econometrician, and all other parameters are treated as auxiliary parameters with fixed known values; therefore, the dark matter measure is defined only based on $\theta = (p, \xi)$.

generating process under calibration C features an excessively large amount of dark matter according to Panel C of Figure 4, and not surprisingly, it has little internal refutability (i.e. little test power) consistent with Theorem 1.

Panel B of Figure 5 displays the histograms of logged overfitting measures $\log \mathcal{O}(\hat{\theta}^{(1)}_{\mathrm{e,n}}, \mathbf{y}^n)$ of efficient GMM estimators for two data-generating processes under calibrations A and C in Figure 4. The estimator $\hat{\theta}_{\mathrm{e,n}}$ is based on the estimation sample $\mathbf{y}^n_{\mathrm{e}} = \{\mathbf{y}_1, \cdots, \mathbf{y}_{\lfloor n/2 \rfloor}\}$. In this experiment, we specify a structural break in the risk premium in the middle of the time-series sample with $\delta_r = 0.4$. Panel B shows that the efficient GMM estimator is likely to be overfitting the data in the calibrated structural model with a high dark matter measure, which is consistent with Theorem 2.

Panel C of Figure 5 compares the expected out-of-sample fit of the recursive GMM estimator $\tilde{\theta}_{\mathrm{e,n}}$ with that of the efficient GMM estimator $\hat{\theta}_{\mathrm{e,n}}$, based on the estimation sample $\mathbf{y}^n_{\mathrm{e}}$. We describe the two types of estimators in Section 5.2. Consistent with the conventional intuition, under the data-generating process A, the efficient GMM estimator yields a better expected out-of-sample fit than the recursive GMM estimator. This is because the additional identification information is reliable and meaningful when the amount of dark matter is not excessively large. In contrast, the recursive GMM estimator delivers a better expected out-of-sample fit under the data-generating process C, which exhibits a much higher dark matter measure. This finding indicates that the concern about misspecification and instability may offset – and even reverse – the efficiency gain from the additional moment restrictions. The result for the data-generating process C suggests that the econometrician should prioritize robustness over efficiency when estimating models that rely heavily relies on dark matter (i.e., with an excessively large $\varrho(\theta_0)$).

What to do with fragile models?

From an econometrician's perspective, robust estimation is particularly important for a GMM model with large dark matter measure, because the concern of misspecification and instability offsets the efficiency gain from imposing cross-equation restrictions. As discussed above, a combination of the recursive and efficient GMM estimators by deviating from the optimal weighting matrix is a potential way to construct estimators that balance robustness and efficiency. We leave a systematic econometric investigation on optimal robust estimation in the presence of dark matter for future research.

From a modeler's perspective, fragile models are unsatisfactory, because they lack refutability, and they are prone to over-fitting. Our analysis shows how to select model calibration based on robustness. Our analysis also highlights the parameter combinations in which the dark matter is embedded.

How to improve the robustness of a model? One approach is to bring in more data to identify the problematic parameter combinations better under the baseline model, (for example, see Barro and Ursúa, 2012; Nakamura, Steinsson, Barro, and Ursúa, 2013, who use international data to better estimate the distribution of consumption disasters). Another approach is to modify the preference specification or the belief formation mechanism so that the model parameters are better identified by the baseline moments and do not rely excessively on the restrictions implied by the asset pricing moments (e.g., Hansen and Sargent, 2010; Bidder and Dew-Becker, 2016; Collin-Dufresne, Johannes, and Lochstoer, 2016; Nagel and Xu, 2019). Yet another approach is to extend the model to connect the problematic parameter combinations of the baseline model to additional data – for example, Gârleanu, Panageas, and Yu (2012) and Kung and Schmid (2015) explicitly model production and innovation to endogenize the consumption dynamics, with a particular focus on low-frequency fluctuations. This ties the properties of R&D investment to those of the consumption process.

7. CONCLUSION

In this paper, we propose a new tractable measure of model fragility based on quantifying the informativeness of the cross-equation restrictions that a structural model imposes on the model parameters. We argue that our measure quantifies a useful model property related to the model's tendency to over-fit the data in sample.

Our fragility measure should be used as a model selection criterion. When faced with a set of candidate models consistent with available data, selecting the less fragile model can be an appealing criterion from the point of view of out of sample performance. We leave the formal development of model selection based on our measure of model fragility to future research.

Our model fragility measure is easy to implement, and the worst-case direction is particularly instructive. This direction provides guidance on which features of the model are most vulnerable to overfitting. Additional data or model elements would be needed to alleviate this tendency. Further, the worst-case direction is useful to consider when constructing robust estimators based on considerations of out-of-sample fit, e.g., following the idea of recursive GMM estimators.

Our methodology has a broad range of potential applications. In addition to the examples involving asset pricing, our measure can be used to assess the robustness of structural models in other areas of economics, such as industrial organization and corporate finance.

REFERENCES

Ai, C., and X. Chen, 2003, "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

Akaike, H., 1973, "Information theory and an extension of the maximum likelihood principle," in *Second International Symposium on Information Theory (Tsahkadsor, 1971)* . pp. 267–281, Akadémiai Kiadó, Budapest.

Ando, T., 2007, "Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models," *Biometrika*, 94, 443–458.

Andrews, D. W. K., 1991, "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–58.

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

Andrews, D. W. K., 2003, "End-of-Sample Instability Tests," *Econometrica*, 71, 1661–1694.

Andrews, D. W. K., and J. C. Monahan, 1992, "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953–66.

Andrews, D. W. K., and W. Ploberger, 1994, "Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative," *Econometrica*, 62, 1383–1414.

Andrews, I., M. Gentzkow, and J. M. Shapiro, 2017, "Measuring the Sensitivity of Parameter Estimates to

Estimation Moments," *The Quarterly Journal of Economics*, 132, 1553–1592.

Athey, S., and G. W. Imbens, 2015, "A Measure of Robustness to Misspecification," *American Economic Review*, 105, 476–480.

Athey, S., and G. W. Imbens, 2017, "The State of Applied Econometrics: Causality and Policy Evaluation," *Journal of Economic Perspectives*, 31, 3–32.

Athey, S., and G. W. Imbens, 2019, "Machine Learning Methods Economists Should Know About," Papers, Stanford University.

Bai, J., and P. Perron, 1998, "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47–78.

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., D. Kiku, and A. Yaron, 2016a, "Risks for the long run: Estimation with time aggregation," *Journal of Monetary Economics*, 82, 52 – 69.

Bansal, R., D. Kiku, and A. Yaron, 2016b, "Risks for the long run: Estimation with time aggregation," *Journal of Monetary Economics*, 82, 52 – 69.

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

Barro, R. J., 2006, "Rare disasters and asset markets in the twentieth century," *The Quarter Journal of Economics*, 121, 823–866.

Barro, R. J., and J. F. Ursúa, 2012, "Rare Macroeconomic Disasters," *Annual Review of Economics*, 4, 83–109.

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.

Bickel, P. J., and J. Kwon, 2001, "Inference for semiparametric models: some questions and an answer," *Statist. Sinica*, 11, 863–960, With comments and a rejoinder by the authors.

Bidder, R., and I. Dew-Becker, 2016, "Long-Run Risk Is the Worst-Case Scenario," *The American Economic Review*, 106, 2494–2527.

Bossaerts, P., and P. Hillion, 1999, "Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?," *The Review of Financial Studies*, 12, 405–428.

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.

Campbell, J. Y., 2018, *Financial Decisions and Markets: A Course in Asset Pricing*, Princeton University Press.

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

Campbell, J. Y., and R. J. Shiller, 1988, "Stock Prices, Earnings, and Expected Dividends," *Journal of Finance*, 43, 661–76.

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

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

Christiano, L. J., and M. Eichenbaum, 1992, "Current Real-Business-Cycle Theories and Aggregate Labor-Market Fluctuations," *The American Economic Review*, 82, 430–450.

Cogley, T., and T. J. Sargent, 2005, "Drifts and volatilities: monetary policies and outcomes in the post WWII US," *Review of Economic Dynamics*, 8, 262 – 302.

Collin-Dufresne, P., M. Johannes, and L. A. Lochstoer, 2016, "Parameter Learning in General Equilibrium: The Asset Pricing Implications," *American Economic Review*, 106, 664–698.

Constantinides, G. M., and A. Ghosh, 2011, "Asset Pricing Tests with Long-run Risks in Consumption Growth," *The Review of Asset Pricing Studies*, 1, 96–136.

Dangl, T., and M. Halling, 2012, "Predictive regressions with time-varying coefficients," *Journal of Financial Economics*, 106, 157 – 181.

Davidson, R., and J. G. MacKinnon, 1987, "Implicit Alternatives and the Local Power of Test Statistics," *Econometrica*, 55, 1305–1329.

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

Efron, B., 2004, "The estimation of prediction error: covariance penalties and cross-validation," *J. Amer. Statist. Assoc.*, 99, 619–642, With comments and a rejoinder by the author.

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

Elliott, G., and U. K. Müller, 2006, "Efficient Tests for General Persistent Time Variation in Regression Coefficients1," *The Review of Economic Studies*, 73, 907–940.

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

Epstein, L. G., and M. Schneider, 2003, "Recursive multiple-priors," *Journal of Economic Theory*, 113, 1–31.

Ferson, W., S. Nallareddy, and B. Xie, 2013, "The "out-of-sample" performance of long run risk models," *Journal of Financial Economics*, 107, 537 – 556.

Fisher, R. A., 1922, "On the mathematical foundations of theoretical statistics," *Phil. Trans. R. Soc. Lond. A*, 222, 309–68.

Foster, D. P., and E. I. George, 1994, "The risk inflation criterion for multiple regression," *Ann. Statist.*, 22, 1947–1975.

Foster, F. D., T. Smith, and R. E. Whaley, 1997, "Assessing Goodness-Of-Fit of Asset Pricing Models: The Distribution of the Maximal R^2," *The Journal of Finance*, 52, 591–607.

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

Gârleanu, N., S. Panageas, and J. Yu, 2012, "Technological Growth and Asset Pricing," *The Journal of Finance*, 67, 1265–1292.

Gelman, A., J. Hwang, and A. Vehtari, 2013, "Understanding predictive information criteria for Bayesian models," *Statsitics and Computing*.

Gilboa, I., and D. Schmeidler, 1989, "Maxmin expected utility with non-unique prior," *Journal of Mathematical*

Economics, 18, 141–153.

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

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

Hansen, B. E., 2000, "Testing for structural change in conditional models," *Journal of Econometrics*, 97, 93 – 115.

Hansen, L., and T. J. Sargent, 2010, "Fragile beliefs and the price of uncertainty," *Quantitative Economics*, 1, 129–162.

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., 2007a, "Beliefs, Doubts and Learning: Valuing Macroeconomic Risk," *American Economic Review*, 97, 1–30.

Hansen, L. P., 2007b, "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.

Hansen, L. P., 2014, "Nobel Lecture: Uncertainty Outside and Inside Economic Models," *Journal of Political Economy*, 122, 945 – 987.

Hansen, L. P., and J. J. Heckman, 1996, "The Empirical Foundations of Calibration," *Journal of Economic Perspectives*, 10, 87–104.

Hansen, L. P., and T. J. Sargent, 1980, "Formulating and estimating dynamic linear rational expectations models," *Journal of Economic Dynamics and Control*, 2, 7–46.

Hansen, L. P., and T. J. Sargent, 1991, *Rational Expectations Econometrics*, Westview Press, Boulder, Colorado.

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

Hansen, L. P., and K. J. Singleton, 1982, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–86.

Hansen, L. P., and K. J. Singleton, 1983, "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, 91, 249–65.

Hansen, P. R., and E.-I. Dumitrescu, 2018, "Parameter Estimation with Out-of-Sample Objective," Working paper.

Hastie, T., R. Tibshirani, and J. Friedman, 2001, *The elements of statistical learning* . Springer Series in Statistics, Springer-Verlag, New York, Data mining, inference, and prediction.

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

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

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

Klibanoff, P., M. Marinacci, and S. Mukerji, 2005, "A Smooth Model of Decision Making under Ambiguity," *Econometrica*, 73, 1849–1892.

Kocherlakota, N., 2016, "Fragility of Purely Real Macroeconomic Models," NBER Working Papers 21866, National Bureau of Economic Research, Inc.

Kocherlakota, N. R., 2007, "Model fit and model selection," *Federal Reserve Bank of St. Louis Review*, July/Augst, 349–360.

Koijen, R. S., and S. Van Nieuwerburgh, 2011, "Predictability of Returns and Cash Flows," *Annual Review of Financial Economics*, 3, 467–491.

Kung, H., and L. Schmid, 2015, "Innovation, Growth, and Asset Prices," *The Journal of Finance*, 70, 1001–1037.

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

Lee, L.-f., 2007, "The method of elimination and substitution in the GMM estimation of mixed regressive, spatial autoregressive models," *Journal of Econometrics*, 140, 155–189.

Lehmann, E., and J. P. Romano, 1996, *Weak convergence and empirical processes* . Springer Series in Statistics, Springer-Verlag, New York, With applications to statistics.

Lettau, M., S. Ludvigson, and J. Wachter, 2008, "The Declining Equity Premium: What Role Does Macroeconomic Risk Play?," *Review of Financial Studies*, 21, 1653–1687.

Lettau, M., and S. Van Nieuwerburgh, 2008, "Reconciling the Return Predictability Evidence," *Review of Financial Studies*, 21, 1607–1652.

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

Lewis, K. K., 2008, "Peso problem," in Steven N. Durlauf, and Lawrence E. Blume (ed.), *The New Palgrave Dictionary of Economics*, Palgrave Macmillan, Basingstoke.

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

Ljungqvist, L., and T. Sargent, 2004, *Recursive Macroeconomic Theory, 2nd Edition* . , vol. 1, The MIT Press, 2 edn.

Lucas, R. E., and T. J. Sargent, 1981, *Rational Expectations and Econometric Practice*, University of Minnesota Press, Minneapolis.

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.

McManus, D. A., 1991, "Who Invented Local Power Analysis?," *Econometric Theory*, 7, 265–268.

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.

Mullainathan, S., and J. Spiess, 2017, "Machine Learning: An Applied Econometric Approach," *Journal of Economic Perspectives*, 31, 87–106.

Müller, U. K., 2012, "Measuring prior sensitivity and prior informativeness in large Bayesian models," *Journal of Monetary Economics*, 59, 581 – 597.

Müller, U. K., and M. W. Watson, 2016, "Measuring Uncertainty about Long-Run Predictions," *The Review of*

Economic Studies, 83, 1711–1740.

Nagel, S., and K. J. Singleton, 2011, “Estimation and Evaluation of Conditional Asset Pricing Models,” *The Journal of Finance*, 66, 873–909.

Nagel, S., and Z. Xu, 2019, “Asset Pricing with Fading Memory,” Discussion paper.

Nakamura, E., J. Steinsson, R. Barro, and J. Ursúa, 2013, “Crises and Recoveries in an Empirical Model of Consumption Disasters,” *American Economic Journal: Macroeconomics*, 5, 35–74.

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

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

Newey, W. K., 1985b, “Maximum likelihood specification testing and conditional moment tests,” *Econometrica*, 53, 1047–1070.

Newey, W. K., 1990, “Efficient Instrumental Variables Estimation of Nonlinear Models,” *Econometrica*, 58, 809–837.

Newey, W. K., 1993, “Efficient Estimation of Models with Conditional Moment Restrictions,” in *Handbook of Statistics*, vol. 11, . chap. 16, Amsterdam: North-Holland, by g.s. maddala, c.r. rao, and h.d. vinod edn.

Newey, W. K., and D. McFadden, 1994, “Chapter 36 Large sample estimation and hypothesis testing,” , vol. 4 of *Handbook of Econometrics* . pp. 2111 – 2245, Elsevier.

Newey, W. K., and K. D. West, 1987, “A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55, 703–08.

Neyman, J., 1937, ““Smooth” tests for goodness of fit,” *Scandinavian Actuarial Journal*, 1937, 149–199.

Nyblom, J., 1989, “Testing for the Constancy of Parameters Over Time,” *Journal of the American Statistical Association*, 84, 223–230.

Ogaki, M., 1993, “17 Generalized method of moments: Econometric applications,” in *Econometrics*, vol. 11 of *Handbook of Statistics*, . pp. 455 – 488, Elsevier.

Pastor, L., and R. Stambaugh, 2001, “The Equity Premium and Structural Breaks,” *Journal of Finance*, 56, 1207–1239.

Pesaran, M. H., and A. Timmermann, 1995, “Predictability of Stock Returns: Robustness and Economic Significance,” *The Journal of Finance*, 50, 1201–1228.

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

Primiceri, G. E., 2005, “Time Varying Structural Vector Autoregressions and Monetary Policy,” *The Review of Economic Studies*, 72, 821–852.

Saikkonen, P., 1989, “Asymptotic relative efficiency of the classical test statistics under misspecification,” *Journal of Econometrics*, 42, 351 – 369.

Saracoglu, R., and T. J. Sargent, 1978, “Seasonality and portfolio balance under rational expectations,” *Journal of Monetary Economics*, 4, 435–458.

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

Schorfheide, F., and K. I. Wolpin, 2012, “On the Use of Holdout Samples for Model Selection,” *The American Economic Review*, 102, 477–481.

Schwarz, G., 1978, "Estimating the dimension of a model," *Ann. Statist.*, 6, 461–464.

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

Shen, X., and J. Ye, 2002, "Adaptive model selection," *J. Amer. Statist. Assoc.*, 97, 210–221.

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

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde, 2002, "Bayesian measures of model complexity and fit," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 64, 583–639.

Stock, J. H., and M. W. Watson, 1996, "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business & Economic Statistics*, 14, 11–30.

Stock, J. H., and M. W. Watson, 1998, "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model," *Journal of the American Statistical Association*, 93, 349–358.

Stock, J. H., and M. W. Watson, 2002, *Introduction to Econometrics*, Addison Wesley; United States, 4th edn.

Stock, J. H., and M. W. Watson, 2008, "Phillips Curve Inflation Forecasts," Nber working papers, National Bureau of Economic Research, Inc.

Stokey, N. L., and R. E. Lucas, 1989, *Recursive Methods in Economics Dynamics, 1st Edition*, The Harvard Press, 1 edn.

Sun, Y., A. Baricz, and S. Zhou, 2010, "On the monotonicity, log-concavity, and tight bounds of the generalized Marcum and Nuttall Q-functions," *IEEE Trans. Inform. Theory*, 56, 1166–1186.

Tibshirani, R., and K. Knight, 1999, "The Covariance Inflation Criterion for Adaptive Model Selection," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61, 529–546.

Valkanov, R., 2003, "Long-horizon regressions: theoretical results and applications," *Journal of Financial Economics*, 68, 201 – 232.

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.

Varian, H. R., 2014, "Big Data: New Tricks for Econometrics," *Journal of Economic Perspectives*, 28, 3–28.

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

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

Welch, I., and A. Goyal, 2008, "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 21, 1455–1508.

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

Ye, J., 1998, "On measuring and correcting the effects of data mining and model selection," *J. Amer. Statist. Assoc.*, 93, 120–131.

Zin, S. E., 2002, "Are behavioral asset-pricing models structural?," *Journal of Monetary Economics*, 49, 215–228.

Appendix

We first list all additional theoretical results in Appendices A and B. Then, we show the proofs of the theorems in Appendix C and those of the propositions and corollaries in Appendix D. The derivation of the disaster risk model can be found in Appendix F. Other miscellaneous derivations and proofs are collected in Appendix G. In the appendix, 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.

APPENDIX A: AUXILIARY RESULTS

A.1. *Auxiliary Results on Data-Generating Processes*

Here we introduce auxiliary propositions that characterize the useful properties of the data-generating processes under regularity conditions. Proposition 3 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 $\mathrm{Q}_{s,f}$ (i.e., the score of $\mathrm{Q}_{s,f}$ is f). Proposition 4 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 4, together with Hellinger-differentiability, is needed to ensure the local asymptotic normality of the local data-generating processes, as established in Proposition 5. 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.

PROPOSITION 3 (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:*

$$\frac{d\mu_{s,f}}{d\mu_0} = 1 + s\bar{f} + s\Delta_\mu(s) \quad and \quad \frac{dK_{s,f}(\cdot|y)}{dK_0(\cdot|y)} = 1 + s\tilde{f}(y,\cdot) + s\Delta_K(y,s) \ \forall\ y \in \mathcal{Y}, \tag{75}$$

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*

$$\bar{f}(\mathbf{y}) \equiv \mathrm{E}^{\mathrm{Q}_0}\left[f(\mathbf{y},\mathbf{y}')|\mathbf{y}\right] = \mathrm{E}^{\mathrm{Q}_0}\left[f(\mathbf{y}',\mathbf{y})|\mathbf{y}\right] \quad and \quad \tilde{f}(\mathbf{y},\mathbf{y}') \equiv f(\mathbf{y},\mathbf{y}') - \bar{f}(\mathbf{y}). \tag{76}$$

PROPOSITION 4 *Suppose Assumptions 2 – 3 hold. 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,

$$n^{-1}\sum_{t\leq \pi n} \tilde{f}_{n,t}^2 \xrightarrow{p} \Upsilon(\pi) \quad and \quad n^{-1}\sum_{t\leq \pi n} \mathrm{E}_{t-1}^{\mathrm{Q}}\left[\tilde{f}_{n,t}^2\right] \xrightarrow{p} \Upsilon(\pi), \quad where \tag{77}$$

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

Further, the asymptotic normality result follows:

(79) $$n^{-1/2} \sum_{t \le \pi n} \tilde{f}_{n,t} \xrightarrow{d} N\left(0, \Upsilon(\pi)\right).$$

PROPOSITION 5 (LAN of Unstable Parametric Submodels) *Suppose Assumptions 2 – 3 hold. 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 4, and $o_p(1)$ denotes a sequence of random variables that converge to zero in probability $\mathbb{P}_0$.

COROLLARY 2 (Contiguity) *Suppose Assumptions 2 – 3 hold. 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 $X_n : \mathcal{Y}^n \to \mathbb{R}$.*

A.2. *Auxiliary Results on Moment Functions*

Here we introduce the basic results (Proposition 6) 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

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

PROPOSITION 6 *Suppose Assumptions 1 – 5 hold. Then, under $\mathbb{P}_{1/\sqrt{n},g,b}$,*

(i) $$\begin{bmatrix} \frac{1}{\sqrt{\pi n}} \sum_{t \le \pi n} m_t(\theta_0) \\ \frac{1}{\sqrt{(1-\pi)n}} \sum_{t > \pi n} m_t(\theta_0) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \frac{1}{\sqrt{\pi}} W(\pi) \\ \frac{1}{\sqrt{1-\pi}} \left(W(1) - W(\pi)\right) \end{bmatrix} + \begin{bmatrix} \nu_e(g, b, \pi) \\ \nu_o(g, b, \pi) \end{bmatrix}$$ *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)
$$\begin{bmatrix} \frac{1}{\sqrt{\pi n}} \sum_{t\leq \pi n} m_t(\hat{\theta}_{e,n}) \\ \frac{1}{\sqrt{(1-\pi)n}} \sum_{t>\pi n} m_t(\hat{\theta}_{e,n}) \end{bmatrix} = \begin{bmatrix} \left[I - D(D^T D)^{-1} D^T\right] \frac{1}{\sqrt{\pi n}} \sum_{t\leq \pi n} m_t(\theta_0) \\ \frac{1}{\sqrt{(1-\pi)n}} \sum_{t>\pi n} m_t(\theta_0) - D(D^T D)^{-1} D^T \frac{1}{\sqrt{\pi n}} \sum_{t\leq \pi n} m_t(\theta_0) \end{bmatrix} + o_p(1),$$
where $\hat{\theta}_{e,n}$ *is the efficient GMM estimator based on estimation sample* $\mathbf{y}_e^n$*;*

(iv)
$$\begin{bmatrix} \frac{1}{\sqrt{\pi n}} \sum_{t\leq \pi n} m_t(\tilde{\theta}_{e,n}) \\ \frac{1}{\sqrt{(1-\pi)n}} \sum_{t>\pi n} m_t(\tilde{\theta}_{e,n}) \end{bmatrix} = \begin{bmatrix} \left[I - D(A^T D)^{-1} A^T\right] \frac{1}{\sqrt{\pi n}} \sum_{t\leq \pi n} m_t(\theta_0) \\ \frac{1}{\sqrt{(1-\pi)n}} \sum_{t>\pi n} m_t(\theta_0) - D(A^T D)^{-1} A^T \frac{1}{\sqrt{\pi n}} \sum_{t\leq \pi n} m_t(\theta_0) \end{bmatrix} + o_p(1),$$
where $\tilde{\theta}_{e,n}$ *is the recursive GMM estimator based on estimation sample* $\mathbf{y}_e^n$*.*

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 (81), we postulate the condition of asymptotic negligibility of innovations (Assumption 5 (iii)), which has been used to establish Gordin's CLT (Gordin, 1969).

Proposition 7 *Suppose Assumptions 1 – 5 hold. Then* $h(\cdot, \theta_0)$ *is defined as follows:*

$$h(\mathbf{y}, \mathbf{y}', \theta_0) = m(\mathbf{y}, \mathbf{y}', \theta_0) - \mathrm{E}^{\mathrm{Q}_0}\left[m_1(\theta_0)|\mathbf{y}_0 = \mathbf{y}\right] + \sum_{t=1}^{\infty}\left\{\mathrm{E}^{\mathrm{Q}_0}\left[m_{t+1}(\theta_0)|\mathbf{y}_1 = \mathbf{y}'\right] - \mathrm{E}^{\mathrm{Q}_0}\left[m_{t+1}(\theta_0)|\mathbf{y}_0 = \mathbf{y}\right]\right\}. \tag{81}$$

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*

$$\mathrm{E}^{\mathrm{Q}_0}\left[m(\cdot, \theta_0) f\right] = \mathrm{E}^{\mathrm{Q}_0}\left[h(\cdot, \theta_0) f\right] \text{ for all } f \in L_0^2(\mathrm{Q}_0). \tag{82}$$

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

$$\mathcal{T}(\mathrm{Q}_0) = \left\{f \in L_0^2(\mathrm{Q}_0) : \ \lambda(f) \in \mathit{lin}(D)\right\}, \tag{83}$$

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* $\mathit{lin}(D)$ *is spanned by columns of* D*, defined in (28).*

A.3. *Auxiliary Results on GMM Estimators Based on the Estimation Sample*

We now introduce the basic results, which extend the standard GMM approximations (Hansen, 1982) in Proposition 8. Then, we introduce a new set of GMM approximations in Proposition 9, which are new and unique to our paper.

Proposition 8 *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)* $\sqrt{\pi n}\left(\tilde{\theta}_{e,n}-\theta_0\right)=-(A^TD)^{-1}A^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n}m_t(\theta_0)\right]+o_p(1),$

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

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

PROPOSITION 9 *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) $\sqrt{\pi n}\begin{bmatrix}\tilde{\theta}_{e,n}^{(1)}-\theta_{e,n}^{(1)}\\ \psi_s(\tilde{\theta}_{e,n}^{(1)})-\psi_s(\theta_{e,n}^{(1)})\end{bmatrix}=-\mathbf{I}_{\mathrm{Q}}^{-1}\Gamma_{\theta,1}^T\mathbf{I}_F\mathbf{I}_B^{-1}D_{11}^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n}m_t^{(1)}(\theta_{n,t}^{(1)})\right]+o_p(1);$

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

Here the matrices L_B *and* L_F *are*

$$L_B\equiv\mathbf{I}_B^{-1}D_{11}^T\Gamma_{m,1}\quad\text{and}\quad L_F\equiv\Gamma_{\theta,1}\mathbf{I}_{\mathrm{Q}}^{-1}D^T,\tag{84}$$

and D_{11} *and* D *are Jacobian matrices defined in (28), the information matrices* $\mathbf{I}_B$ *and* $\mathbf{I}_{\mathrm{Q}}$ *are defined in (29) – (30), and the selection matrices* $\Gamma_{m,1}$ *and* $\Gamma_{\theta,1}$ *are defined in (18) and (30), respectively.*

PROPOSITION 10 *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 (55) – (56). 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) $\begin{bmatrix}\mathcal{L}(\tilde{\theta}_{e,n}^{(1)};\mathbf{y}_e^n)\\ \mathcal{L}(\tilde{\theta}_{e,n}^{(1)};\mathbf{y}_o^n)\end{bmatrix}=\begin{bmatrix}((L_B-2L_F)\zeta_{e,n}-2L_\Delta\nu_e)^T\,\mathbf{I}_F\,(L_B\zeta_{e,n})\\ (L_B\zeta_{e,n}-2L_F\zeta_{o,n}-2L_\Delta\nu_o)^T\,\mathbf{I}_F\,(L_B\zeta_{e,n})\end{bmatrix}+o_p(1),$ *and*

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

where $\nu_e(g,b,\pi)$ *and* $\nu_o(g,b,\pi)$ *are defined in (80), and the random vectors* $\zeta_{e,n}$ *and* $\zeta_{o,n}$ *are*

$$\zeta_{e,n}\equiv\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n}m_t(\theta_0)-\nu_e(g,b,\pi)\ \text{and}\ \zeta_{o,n}\equiv\frac{1}{\sqrt{\pi n}}\sum_{t>\pi n}m_t(\theta_0)-\nu_o(g,b,\pi),\tag{85}$$

and the matrices L_F*,* L_B *are defined in (84) and* $L_\Delta\equiv L_F-L_B$*. Further, using Proposition 6,*

$$\begin{bmatrix}\zeta_{e,n}\\ \zeta_{o,n}\end{bmatrix}\xrightarrow{d}\begin{bmatrix}\frac{1}{\sqrt{\pi}}W(\pi)\\ \frac{1}{\sqrt{1-\pi}}\left(W(1)-W(\pi)\right)\end{bmatrix}.\tag{86}$$

APPENDIX B: 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.[27] 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:

$$\vartheta(\mathbb{P}_{1/\sqrt{n},g,b}) \equiv \frac{1}{n}\sum_{t=1}^{n} \vartheta(\mathrm{Q}_{1/\sqrt{n},f_{n,t}}), \text{ with } f_{n,t} = g_1(\mathbf{y}_{t-1},\mathbf{y}_t) + g_2(\mathbf{y}_{t-1},\mathbf{y}_t)b(t/n). \tag{87}$$

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

THEOREM 4 *(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 5 *(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 4 and 5 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.

APPENDIX C: PROOFS OF THE MAIN THEOREMS

1. Proof of Theorem 1

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 6, 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 (1985a),

[27] 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.

it follows that for any GMM specification test $\check{\varphi}_n$ with an asymptotic level α and an asymptotic local power function ($\forall\ \check{\varphi}_n \in \Phi_\alpha(\mathrm{Q}_0)$),

$$\inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} \lim_{n\to\infty} \int \check{\varphi}_n d\mathbb{P}_{1/\sqrt{n},g,0} \leq \inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} \lim_{n\to\infty} \int \hat{\varphi}_n d\mathbb{P}_{1/\sqrt{n},g,0} \quad \text{(i.e., } C \text{ test is asymptotically optimal)} \tag{88}$$

$$= \inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} \lim_{n\to\infty} \mathbb{P}_{1/\sqrt{n},g,0} \left\{ \left|\widehat{\mathbb{G}}_n\right|^2 > c_{1-\alpha} \right\}, \tag{89}$$

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

$$\widehat{\mathbb{G}}_n = \left(\Lambda_2 - \Lambda_2 D_{21} \mathbf{I}_{\mathrm{F}}^{-1} D_{21}^T \Lambda_2\right)^{-1/2} \left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} m_t^{(2)}(\hat{\theta}_n)\right]; \tag{90}$$

see page 243 of Newey (1985a) and Appendix G.3 of this paper. Here $\Lambda_2 = I - D_{22}(D_{22}^T D_{22})^{-1} D_{22}^T$.

Now, we obtain (e.g., Newey, 1985a; Chen and Santos, 2018, or Proposition 6 of this paper)

$$|\widehat{\mathbb{G}}_n|^2 \xrightarrow{d} \chi^2_{d_{m,2}-d_{\theta,2}}(\mu_g), \tag{91}$$

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 $\mu_g = \lambda^{(2)}(g_1)^T \left(\Lambda_2 - \Lambda_2 D_{21} \mathbf{I}_{\mathrm{F}}^{-1} D_{21}^T \Lambda_2\right) \lambda^{(2)}(g_1)$.

From (49) and (88) – (89), we conclude that

$$\inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} q(g, \check{\varphi}) \leq \inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} \lim_{n\to\infty} \mathbb{P}_{1/\sqrt{n},g} \left\{ \left|\widehat{\mathbb{G}}_n\right|^2 > c_{1-\alpha} \right\} = \inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} P\left\{\chi^2_{d_m-d_\theta}(\mu_g) > c_{1-\alpha}\right\}. \tag{92}$$

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

$$\inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} q(g, \check{\varphi}) \leq \inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} M_{\frac{d_{m,2}-d_{\theta,2}}{2}}\left(\sqrt{\mu_g}, \sqrt{c_{1-\alpha}}\right) = M_{\frac{d_{m,2}-d_{\theta,2}}{2}}\left(\inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} \sqrt{\mu_g}, \sqrt{c_{1-\alpha}}\right), \tag{93}$$

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 $\mathbf{I}_{\mathrm{F}}^{-1/2} (\mathbf{I}_{\mathrm{F}} - \mathbf{I}_{\mathrm{B}}) \mathbf{I}_{\mathrm{F}}^{-1/2} = \mathbf{I}_{\mathrm{F}}^{-1/2} (D_{21}^T \Lambda_2 D_{21}) \mathbf{I}_{\mathrm{F}}^{-1/2}$, 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

$$\inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} q(g,\check{\varphi}) \leq M_{\frac{d_{m,2}-d_{\theta,2}}{2}}\left(\sqrt{\frac{\kappa^2}{1+\varrho(\theta_0)}}, \sqrt{c_{1-\alpha}}\right). \tag{94}$$

Proof of Theorem 2

According to Proposition 10 (ii), we can show that

$$\begin{aligned}\mathrm{E}\left[\operatorname*{wlim}_{n\to\infty}\frac{1}{2}\left(\mathcal{L}(\hat{\theta}_{\mathrm{e,n}};\mathbf{y}_{\mathrm{o}}^n)-\mathcal{L}(\hat{\theta}_{\mathrm{e,n}};\mathbf{y}_{\mathrm{e}}^n)\right)\right] &= \pi^{-1}\mathrm{E}\left[W(\pi)^T L_F^T \mathbf{I}_{\mathrm{F}} L_F W(\pi)\right] \\ &\quad + \left[\nu_e(g,b,\pi)-\nu_o(g,b,\pi)\right]^T L_\Delta^T \mathbf{I}_{\mathrm{F}} L_\Delta \nu_e(g,b,\pi),\end{aligned} \tag{95}$$

where $\mathrm{wlim}_{n\to\infty}$ is the weak convergence limit and $W(\cdot)$ is a d_m-dimensional Wiener process. The first term above is

$$\pi^{-1}\mathrm{E}\left[W(\pi)^T L_{\mathrm{F}}^T \mathbf{I}_{\mathrm{F}} L_{\mathrm{F}} W(\pi)\right] = \pi^{-1}\mathrm{E}\left[\mathbf{tr}\left(\mathbf{I}_{\mathrm{F}}^{1/2} L_{\mathrm{F}} W(\pi) W(\pi)^T L_{\mathrm{F}}^T \mathbf{I}_{\mathrm{F}}^{1/2}\right)\right] \tag{96}$$

$$= \mathbf{tr}\left(\mathbf{I}_{\mathrm{F}}^{1/2} L_{\mathrm{F}} L_{\mathrm{F}}^T \mathbf{I}_{\mathrm{F}}^{1/2}\right). \tag{97}$$

According to the definition of L_{F} in (84),

$$L_{\mathrm{F}} L_{\mathrm{F}}^T = \Gamma_{\theta,1}\mathbf{I}_{\mathrm{Q}}^{-1}\Gamma_{\theta,1}^T = \mathbf{I}_{\mathrm{F}}^{-1}. \tag{98}$$

Combining (97) and (98) yields

$$\pi^{-1}\mathrm{E}\left[W(\pi)^T L_{\mathrm{F}}^T \mathbf{I}_{\mathrm{F}} L_{\mathrm{F}} W(\pi)\right] = d_{\theta,1}. \tag{99}$$

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

$$\left[\nu_e-\nu_o\right]^T L_\Delta^T \mathbf{I}_{\mathrm{F}} L_\Delta \nu_e = \frac{1}{\sqrt{\pi}}\left(\frac{1}{\sqrt{\pi}}+\frac{1}{\sqrt{1-\pi}}\right)\left(\int_0^\pi b(u)\mathrm{d}u\right)^2 \lambda(g_2)^T L_\Delta^T \mathbf{I}_{\mathrm{F}} L_\Delta \lambda(g_2). \tag{100}$$

The left-hand side of (100) is bounded from above by

$$\begin{aligned}&\frac{1}{\sqrt{\pi}}\left(\frac{1}{\sqrt{\pi}}+\frac{1}{\sqrt{1-\pi}}\right)\left(\int_0^\pi b(u)\mathrm{d}u\right)^2 \lambda(g_2)^T L_\Delta^T \mathbf{I}_{\mathrm{F}} L_\Delta \lambda(g_2) \\ &\qquad \leq \pi\left(1+\sqrt{\frac{\pi}{1-\pi}}\right)|\lambda(g_2)|^2 \times \text{the largest eigenvalue of } L_\Delta^T \mathbf{I}_{\mathrm{F}} L_\Delta.\end{aligned} \tag{101}$$

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

According to Proposition 10 (i), we can show that

$$\mathrm{E}\left[\operatorname*{wlim}_{n\to\infty}\frac{1}{2}\left(\mathcal{L}(\tilde{\theta}_{\mathrm{e,n}};\mathbf{y}_{\mathrm{o}}^{n})-\mathcal{L}(\tilde{\theta}_{\mathrm{e,n}};\mathbf{y}_{\mathrm{e}}^{n})\right)\right]=\pi^{-1}\mathrm{E}\left[W(\pi)^{T}L_{\mathrm{F}}^{T}\mathbf{I}_{\mathrm{F}}L_{\mathrm{B}}W(\pi)\right], \tag{102}$$

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

$$\pi^{-1}\mathrm{E}\left[W(\pi)^{T}L_{\mathrm{F}}^{T}\mathbf{I}_{\mathrm{F}}L_{\mathrm{B}}W(\pi)\right]=\mathbf{tr}(\mathbf{I}_{\mathrm{F}}^{1/2}L_{\mathrm{B}}L_{\mathrm{F}}^{T}\mathbf{I}_{\mathrm{F}}^{1/2}). \tag{103}$$

Because $L_{\mathrm{B}}L_{\mathrm{F}}^{T}=\mathbf{I}_{\mathrm{B}}^{-1}D_{11}^{T}\left[D_{11},0_{d_{m,1}\times(d_{\theta}-d_{\theta,1})}\right]\mathbf{I}_{\mathbb{Q}}^{-1}\Gamma_{\theta,1}^{T}=\Gamma_{\theta,1}\mathbf{I}_{\mathbb{Q}}^{-1}\Gamma_{\theta,1}^{T}=\mathbf{I}_{\mathrm{F}}^{-1}$, the equality (103) can further be rewritten as

$$\pi^{-1}\mathrm{E}\left[W(\pi)^{T}L_{\mathrm{F}}^{T}\mathbf{I}_{\mathrm{F}}L_{\mathrm{B}}W(\pi)\right]=d_{\theta,1}. \tag{104}$$

Proof of Theorem 4

The local asymptotic normality (LAN) (see Proposition 5), 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 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

$$\lim_{s\to0}\frac{1}{s}\left[\vartheta(\mathrm{Q}_{s,f})-\vartheta(\mathrm{Q}_0)\right]=\mathrm{E}^{\mathrm{Q}_0}\left[\dot{\vartheta}f\right], \tag{105}$$

where $\dot{\vartheta}(\mathbf{y}_{t-1},\mathbf{y}_t)\equiv(D^TD)^{-1}D^Th(\mathbf{y}_{t-1},\mathbf{y}_t,\theta_0)$ with $h(\mathbf{y}_{t-1},\mathbf{y}_t,\theta_0)$ defined in Proposition 7 (e.g., Greenwood and Wefelmeyer, 1995). According to Proposition 7, $h(\cdot,\theta_0)$ 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 $g_1(\mathbf{y},\mathbf{y}')=v^T\dot{\vartheta}(\mathbf{y},\mathbf{y}')$, $g_2(\mathbf{y}_{t-1},\mathbf{y}_t)\equiv0$, and $b(u)\equiv0$ 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 $1\le t\le n$. 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

$$\left(\sqrt{n}\left(v^T\check{\theta}_n-v^T\theta_0\right),\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}g_1(\mathbf{y}_{t-1},\mathbf{y}_t)\right)\xrightarrow{d}(\Xi_0,U_0), \tag{106}$$

where $U_0\sim N(0,v^T(D^TD)^{-1}v)$ (see Proposition 4). 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 $\mathbb{P}_{1/\sqrt{n},g,0}$,

$$\sqrt{n}\left(v^T\check{\theta}_n - v^T\vartheta(\mathbb{P}_{1/\sqrt{n},g,0})\right) \xrightarrow{d} \Xi_g, \tag{107}$$

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 $\tau:\ \mathbb{R}^{d_\theta} \to \mathbb{R}$:

$$\begin{aligned}\Xi_g &= \tau(X_g) - v^T\xi \\ &= \tau(X_g) - \mathrm{E}^{\mathrm{Q}_0}\left[v^T\dot{\vartheta}f\right] \\ &= \tau(X_g) - \left[v^T(D^TD)^{-1}v\right].\end{aligned} \tag{108}$$

Here, the local estimation bias is $\xi \equiv (D^TD)^{-1}D^T\lambda(g_1) = (D^TD)^{-1}v$ (similar to Corollary 1 or the proof of Proposition 6 (ii)) and $X_g \sim N(\xi, (D^TD)^{-1})$. Based on Theorem 8.6 of van der Vaart (1998) for estimating normal means, it holds that for all measurable function τ,

$$\mathrm{E}^{\mathrm{Q}_{1/\sqrt{n},f}}\left[\Xi_g^2\right] \geq \mathrm{E}^{\mathrm{Q}_0}\left[\left(v^TX_0\right)^2\right] = v^T(D^TD)^{-1}v. \tag{109}$$

The key idea of (106) – (108) 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 (109) is a lower bound for the minimax risk R:

$$R \equiv \lim_{l\to\infty}\liminf_{n\to\infty}\int l \wedge \left[\sqrt{n}v^T\left(\check{\theta}_n - \vartheta(\mathbb{P}_{1/\sqrt{n},g,0})\right)\right]^2 d\mathbb{P}_{1/\sqrt{n},g,0}. \tag{110}$$

In fact, it holds that

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

Thus, the minimax risk can be bounded from below by

$$R \geq \lim_{l\to\infty}\mathrm{E}^{\mathrm{Q}_{1/\sqrt{n},f}}\left[l \wedge \Xi_g^2\right] \geq \lim_{l\to\infty}\mathrm{E}^{\mathrm{Q}_{1/\sqrt{n},f}}\left[l \wedge \Xi_g^2\right]. \tag{111}$$

According to the monotone convergence theorem, it follow that

$$R \geq \mathrm{E}^{\mathrm{Q}_{1/\sqrt{n},f}}\left[\Xi_g^2\right]. \tag{112}$$

Combining (109) and (112), the local asymptotic minimax lower bound result holds: $R \geq v^T(D^TD)^{-1}v$.

Proof of Theorem 5

We start with

$$\sqrt{n}\left[\hat{\theta}_n - \vartheta(\mathbb{P}_{1/\sqrt{n},g,b})\right] = \sqrt{n}\left(\hat{\theta}_n - \theta_0\right) - \sqrt{n}\left[\vartheta(\mathbb{P}_{1/\sqrt{n},g,b}) - \theta_0\right]. \tag{113}$$

According to Proposition 8 (ii), it follows that

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) = -(D^T D)^{-1} D^T \left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} m_t(\theta_0)\right] + o_p(1). \tag{114}$$

Consequently, similar to Corollary 1 or the proof of Proposition 6 (ii),

$$\sqrt{n}\left[\vartheta(\mathbb{P}_{1/\sqrt{n},g,b}) - \theta_0\right] = -(D^T D)^{-1} D^T \lambda(g_1) + o(1). \tag{115}$$

Thus, appealing to Proposition 6 (i), we can show that

$$\sqrt{n}\left[\hat{\theta}_n - \vartheta(\mathbb{P}_{1/\sqrt{n},g,b})\right] \xrightarrow{d} -(D^T D)^{-1} D^T W(1), \tag{116}$$

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

$$\liminf_{n\to\infty} \int l \wedge \left[\sqrt{n} v^T \left(\hat{\theta}_n - \vartheta(\mathbb{P}_{1/\sqrt{n},g,b})\right)\right]^2 d\mathbb{P}_{1/\sqrt{n},g,b} = \mathrm{E}\left[l \wedge X^2\right], \quad \text{with } X \sim N(0, v^T (D^T D)^{-1} v). \tag{117}$$

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

$$\lim_{l\to\infty} \liminf_{n\to\infty} \int l \wedge \left[\sqrt{n} v^T \left(\hat{\theta}_n - \vartheta(\mathbb{P}_{1/\sqrt{n},g,b})\right)\right]^2 d\mathbb{P}_{1/\sqrt{n},g,b} = \mathrm{E}\left[X^2\right] = v^T (D^T D)^{-1} v. \tag{118}$$

APPENDIX D: PROOFS OF PROPOSITIONS

1. Proof of Proposition 1

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,

$$\mathrm{E}^{\mathrm{Q}_0}[\Delta(s)] = \mathrm{E}^{\mathrm{Q}_0}\left[\frac{\mathrm{d}Q_{s,f}}{\mathrm{d}Q_0} - 1\right] = \int \mathrm{d}Q_{s,f} - \int \mathrm{d}Q_0 = 0. \tag{119}$$

According to Proposition 3, 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,

$$\mathrm{E}^{\mathrm{Q}_0}[f(\mathbf{y}, \mathbf{y}')|\mathbf{y}] = \mathrm{E}^{\mathrm{Q}_0}[f(\mathbf{y}', \mathbf{y})|\mathbf{y}]. \tag{120}$$

2. Proof of Proposition 2

According to Definition 4, it follows that

$$\mathrm{E}^{\mathrm{Q}_{1/\sqrt{n},f_{n,t}}}\left[m_t(\theta_0)\right] = \int m_t(\theta_0)\left[1 + f_{n,t}/\sqrt{n} + \Delta_n\right]\mathrm{dQ}_0. \tag{121}$$

Because $\mathrm{E}^{\mathrm{Q}_0}\left[m_t(\theta_0)\right] = 0$, the equality (121) above leads to

$$\mathrm{E}^{\mathrm{Q}_{1/\sqrt{n},f_{n,t}}}\left[m_t(\theta_0)\right] = \frac{\lambda(g_1) + \lambda(g_2)b(t/n)}{\sqrt{n}} + \int m_t(\theta_0)\Delta_n \mathrm{dQ}_0. \tag{122}$$

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

$$\left|\int m_t(\theta_0)\Delta_n \mathrm{dQ}_0\right| \leq \mathrm{E}^{\mathrm{Q}_0}\left[|m_t(\theta_0)|^2\right]^{1/2} \mathrm{E}^{\mathrm{Q}_0}\left[|\Delta_n|^2\right]^{1/2} = o\left(\frac{1}{\sqrt{n}}\right). \tag{123}$$

3. Proof of Proposition 3

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

$$\mathrm{d}\mu_{s,f}(\mathbf{y}) = \left[1 + s\bar{f}(\mathbf{y}) + s\Delta_\mu(s)\right]\mathrm{d}\mu(\mathbf{y}), \tag{124}$$

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

$$\mathrm{d}K_{s,f}(\mathbf{y}'|\mathbf{y}) = \left\{1 + s\left[f(\mathbf{y},\mathbf{y}') - \bar{f}(\mathbf{y})\right] + s\Delta_K(\mathbf{y},s)\right\}\mathrm{d}K(\mathbf{y}'|\mathbf{y}), \tag{125}$$

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 $\mathrm{E}^{\mathrm{Q}}\left[\Delta_K(\mathbf{y},s)|\mathbf{y}\right] = 0$.

4. Proof of Proposition 4

According to Assumption 3 (i),

$$n^{-1}\max_{1\leq t\leq n}|g(\mathbf{y}_{t-1},\mathbf{y}_t)|^2\xrightarrow{p}0. \tag{126}$$

According to simple algebra, we can show that

$$n^{-1}\sum_{t\leq\pi n}\tilde{f}_{n,t}^2=n^{-1}\sum_{t\leq\pi n}\left[\tilde{g}_1(\mathbf{y}_{t-1},\mathbf{y}_t)^2+2\tilde{g}_1(\mathbf{y}_{t-1},\mathbf{y}_t)\tilde{g}_2(\mathbf{y}_{t-1},\mathbf{y}_t)b(t/n)+\tilde{g}_2(\mathbf{y}_{t-1},\mathbf{y}_t)^2b(t/n)^2\right]. \tag{127}$$

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

$$n^{-1}\sum_{t\leq\pi n}\tilde{f}_{n,t}^2\to\Upsilon(\pi)\equiv\mathrm{E}^{\mathrm{Q}_0}\left[\tilde{g}^TB_\pi\tilde{g}\right]. \tag{128}$$

Using the same argument, we can show that

$$n^{-1}\sum_{t\leq\pi n}\mathrm{E}_{t-1}^{\mathrm{Q}}\left[\tilde{f}_{n,t}^2\right]\xrightarrow{p}\Upsilon(\pi)\equiv\mathrm{E}^{\mathrm{Q}_0}\left[\tilde{g}^TB_\pi\tilde{g}\right]. \tag{129}$$

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:

$$\frac{1}{\sqrt{n}}\sum_{t\leq\pi n}\tilde{f}_{n,t}\xrightarrow{d}N(0,\Upsilon(\pi)). \tag{130}$$

5. Proof of Proposition 5

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 (125), it follows that

$$\sum_{t\leq n}W_{n,t}=\frac{1}{\sqrt{n}}\sum_{t\leq n}\tilde{f}_{n,t}+\frac{1}{\sqrt{n}}\sum_{t\leq n}\tilde{\Delta}_{n,t}. \tag{131}$$

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 $\mathrm{E}^{\mathrm{Q}_0}\left[\tilde{\Delta}_{n,t}^2\right]\to0$ as $n\to\infty$ for all $t=1,\cdots,n$, it follows that

$$\mathrm{E}^{\mathrm{Q}_0}\left[\frac{1}{\sqrt{n}}\sum_{t\leq n}\tilde{\Delta}_{n,t}\right]=0\quad\text{and}\quad\mathrm{var}^{\mathrm{Q}_0}\left[\frac{1}{\sqrt{n}}\sum_{t\leq n}\tilde{\Delta}_{n,t}\right]\leq\frac{1}{n}\sum_{t\leq n}\mathrm{E}^{\mathrm{Q}_0}\left[\tilde{\Delta}_{n,t}^2\right]\to0. \tag{132}$$

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:

$$\sum_{t\leq n} W_{n,t} = \frac{1}{\sqrt{n}}\sum_{t\leq n}\tilde{f}_{n,t} + o_p(1). \tag{133}$$

By Taylor expansion, we have

$$\ln(1+x) = x - \frac{1}{2}x^2 + x^2 R(x), \tag{134}$$

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

$$\ln\prod_{t\leq n}\frac{\mathrm{d}K_{n,t}}{\mathrm{d}K_0} = \sum_{t\leq n}\ln(1+W_{n,t}) = \sum_{t\leq n}\left[W_{n,t} - \frac{1}{2}W_{n,t}^2 + W_{n,t}^2 R(W_{n,t})\right] \tag{135}$$

$$= \sum_{t\leq n} W_{n,t} - \frac{1}{2}\sum_{t\leq n} W_{n,t}^2 + \sum_{t\leq n} W_{n,t}^2 R(W_{n,t}). \tag{136}$$

Combining (133) and (136) yields

$$\ln\prod_{t\leq n}\frac{\mathrm{d}K_{n,t}}{\mathrm{d}K_0} = \frac{1}{\sqrt{n}}\sum_{t\leq n}\tilde{f}_{n,t} - \frac{1}{2}\sum_{t\leq n} W_{n,t}^2 + \sum_{t\leq n} W_{n,t}^2 R(W_{n,t}) + o_p(1). \tag{137}$$

We shall first show that

$$\sum_{t\leq n} W_{n,t}^2 = \frac{1}{n}\sum_{t\leq n}\tilde{f}_{n,t}^2 + o_p(1). \tag{138}$$

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

$$\left|\sum_{t\leq n} W_{n,t}^2 - \frac{1}{n}\sum_{t\leq n}\tilde{f}_{n,t}^2\right| \leq \sum_{t\leq n}\left|\frac{1}{\sqrt{n}}\tilde{\Delta}_{n,t}\left(\frac{2}{\sqrt{n}}\tilde{f}_{n,t} + \frac{1}{\sqrt{n}}\tilde{\Delta}_{n,t}\right)\right| \tag{139}$$

$$\leq \left(\frac{1}{n}\sum_{t\leq n}\tilde{\Delta}_{n,t}^2\right)^{1/2}\left[\frac{1}{n}\sum_{t\leq n}\left(2\tilde{f}_{n,t} + \tilde{\Delta}_{n,t}\right)^2\right]^{1/2}. \tag{140}$$

Based on (125), 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 (140) 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 (137) can be rewritten as

$$\ln\prod_{t\leq n}\frac{\mathrm{d}K_{n,t}}{\mathrm{d}K_0} = \frac{1}{\sqrt{n}}\sum_{t\leq n}\tilde{f}_{n,t} - \frac{1}{2n}\sum_{t\leq n}\tilde{f}_{n,t}^2 + \sum_{t\leq n} W_{n,t}^2 R(W_{n,t}) + o_p(1) \tag{141}$$

$$= \frac{1}{\sqrt{n}}\sum_{t\leq n}\tilde{f}_{n,t} - \frac{1}{2}\int_0^1 \Upsilon(u)\mathrm{d}u + \sum_{t\leq n} W_{n,t}^2 R(W_{n,t}) + o_p(1). \tag{142}$$

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

$$\sum_{t\le n} W_{n,t}^2 |R(W_{n,t})| \le \max_{1\le t\le n} |R(W_{n,t})| \sum_{t\le n} W_{n,t}^2, \tag{143}$$

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

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

$$\mathbb{P}_0\left(\max_{1\le t\le n} |R(W_{n,t})| > \epsilon\right) \le \sum_{t\le n} \mathbb{P}_0\left(|R(W_{n,t})| > \epsilon\right) \le \sum_{t\le n} \mathbb{P}_0\left(W_{n,t}^2 > \epsilon_R\right) \tag{144}$$

$$\le \sum_{t\le n} \mathbb{P}_0\left(\tilde{f}_{n,t}^2 > n\epsilon_R/4\right) + \sum_{t\le n} \mathbb{P}_0\left(\tilde{\Delta}_{n,t}^2 > n\epsilon_R/4\right). \tag{145}$$

By Markov's inequality, we can further show that

$$\mathbb{P}_0\left(\max_{1\le t\le n} |R(W_{n,t})| > \epsilon\right) \le \frac{4}{n\epsilon_R} \sum_{t\le n} \mathrm{E}^{\mathrm{Q}_0}\left[\tilde{f}_{n,t}^2 \mathbf{1}\{\tilde{f}_{n,t}^2 > n\epsilon_R/4\}\right] + \frac{4}{n\epsilon_R} \sum_{t\le n} \mathrm{E}^{\mathrm{Q}_0}\left[\tilde{\Delta}_{n,t}^2\right]. \tag{146}$$

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

$$\frac{1}{n} \sum_{t\le n} \mathrm{E}^{\mathrm{Q}_0}\left[\tilde{f}_{n,t}^2 \mathbf{1}\{\tilde{f}_{n,t}^2 > n\epsilon_R/4\}\right] \to 0 \text{ as } n \to \infty. \tag{147}$$

Further, according to (125), it holds that

$$\frac{1}{n} \sum_{t\le n} \mathrm{E}^{\mathrm{Q}_0}\left[\tilde{\Delta}_{n,t}^2\right] \to 0 \text{ as } n \to \infty. \tag{148}$$

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

6. Proof of Proposition 6

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

$$\mathrm{E}^{\mathrm{Q}_{1/\sqrt{n}, f_{n,t}}}\left[\tilde{m}_t(\theta_0)\right] = o\left(\frac{1}{\sqrt{n}}\right), \text{ with } f_{n,t} = g(\mathbf{y}_{t-1}, \mathbf{y}_t)^T \begin{bmatrix} 1 \\ b(t/n) \end{bmatrix}. \tag{149}$$

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, appealing to the functional central limit theorem (invariance principle) of McLeish (1975a) and Phillips and Durlauf (1986), we know that

$$\frac{1}{\sqrt{n}} \sum_{t\le \pi n} \tilde{m}(\theta_0) \xrightarrow{d} W(\pi), \quad \text{for all } \pi \in [0, 1]. \tag{150}$$

Thus,

$$\frac{1}{\sqrt{\pi n}}\sum_{t\leq \pi n} m_t(\theta_0) = \frac{1}{\sqrt{\pi n}}\sum_{t\leq \pi n} \tilde{m}_t(\theta_0) + \frac{1}{n}\sum_{t\leq \pi n}\frac{\lambda(g^T)}{\sqrt{\pi}}\begin{bmatrix} 1 \\ b(t/n)\end{bmatrix} \xrightarrow{d} \frac{W(\pi)}{\sqrt{\pi}} + \frac{\lambda(g^T)}{\sqrt{\pi}}\begin{bmatrix} \pi \\ \int_0^\pi b(u)\mathrm{d}u\end{bmatrix}. \tag{151}$$

Similarly, we can show that

$$\frac{1}{\sqrt{(1-\pi)n}}\sum_{t>\pi n} m_t(\theta_0) \xrightarrow{d} \frac{W(1)-W(\pi)}{\sqrt{1-\pi}} + \frac{\lambda(g^T)}{\sqrt{1-\pi}}\begin{bmatrix} 1-\pi \\ \int_\pi^1 b(u)\mathrm{d}u\end{bmatrix}. \tag{152}$$

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

$$0 = \int m_t(\theta_{n,t})\mathrm{d}\mathrm{Q}_{1/\sqrt{n}, f_{n,t}}, \text{ for all } t, n. \tag{153}$$

Using the Taylor expansion, we obtain

$$0 = \int \left[m_t(\theta_0) + \nabla_\theta m_t(\dot{\theta}_{n,t})(\theta_{n,t}-\theta_0)\right]\left[1 + f_{n,t}/\sqrt{n} + \Delta_{n,t}/\sqrt{n}\right]\mathrm{d}\mathrm{Q}_0, \text{ for all } t, n, \tag{154}$$

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

$$0 = \frac{1}{\sqrt{n}}\lambda(g^T)\begin{bmatrix} 1 \\ b(t/n)\end{bmatrix} + D(\theta_{n,t}-\theta_0) + o\left(\frac{1}{\sqrt{n}}\right), \text{ for all } t, n. \tag{155}$$

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

$$\theta_{n,t} - \theta_0 = -(D^TD)^{-1}D^T\frac{1}{\sqrt{n}}\lambda(g^T)\begin{bmatrix} 1 \\ b(t/n)\end{bmatrix} + o\left(\frac{1}{\sqrt{n}}\right), \text{ for all } t, n. \tag{156}$$

Hence, using the Taylor expansion again leads to

$$\frac{1}{\sqrt{\pi n}}\sum_{t\leq \pi n} m_t(\theta_{n,t}) = \frac{1}{\sqrt{\pi n}}\sum_{t\leq \pi n} m_t(\theta_0) - \frac{1}{n}\sum_{t\leq \pi n}\nabla_\theta m_t(\dot{\theta}_{n,t})(D^TD)^{-1}D^T\frac{\lambda(g^T)}{\sqrt{\pi}}\begin{bmatrix} 1 \\ b(t/n)\end{bmatrix} + o\left(1\right). \tag{157}$$

Due to Assumption 5, appealing to Lemma 4 of Li and Müller (2009) leads to

$$\frac{1}{\sqrt{\pi n}}\sum_{t\leq \pi n} m_t(\theta_{n,t}) = \frac{1}{\sqrt{\pi n}}\sum_{t\leq \pi n} m_t(\theta_0) - D(D^TD)^{-1}D^T\frac{\lambda(g^T)}{\sqrt{\pi}}\begin{bmatrix} \pi \\ \int_0^\pi b(u)\mathrm{d}u\end{bmatrix} + o\left(1\right). \tag{158}$$

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

$$\frac{1}{\sqrt{\pi n}}\sum_{t\leq \pi n} m_t(\theta_{n,t}) = \frac{1}{\sqrt{\pi n}}\sum_{t\leq \pi n} m_t(\theta_0) - \frac{\lambda(g^T)}{\sqrt{\pi}}\begin{bmatrix} \pi \\ \int_0^\pi b(u)\mathrm{d}u\end{bmatrix} + o\left(1\right). \tag{159}$$

Similarly, we can show that

$$\frac{1}{\sqrt{(1-\pi)n}}\sum_{t>\pi n} m_t(\theta_{n,t}) = \frac{1}{\sqrt{(1-\pi)n}}\sum_{t>\pi n} m_t(\theta_0) - \frac{\lambda(g^T)}{\sqrt{1-\pi}}\begin{bmatrix} 1-\pi \\ \int_\pi^1 b(u)\mathrm{d}u \end{bmatrix} + o(1)\,. \tag{160}$$

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

$$\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\hat\theta_{\mathrm{e,n}}) = \frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_0) + \frac{1}{\pi n}\sum_{t\le\pi n} \nabla_\theta m_t(\dot\theta_{\mathrm{e,n}})\left[\sqrt{\pi n}(\hat\theta_{\mathrm{e,n}}-\theta_0)\right] + o(1)\,, \tag{161}$$

where $\dot\theta_{\mathrm{e,n}}$ lies between $\hat\theta_{\mathrm{e,n}}$ and θ_0. According to Proposition 8 (ii),

$$\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\hat\theta_{\mathrm{e,n}}) = \frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_0) - D(D^TD)^{-1}D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_0)\right] + o_p(1) \tag{162}$$

Further rearranging the terms on the right-hand side of (162) leads to

$$\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\hat\theta_{\mathrm{e,n}}) = \left[I - D(D^TD)^{-1}D^T\right]\left[\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_0)\right] + o_p(1)\,. \tag{163}$$

Similarly,

$$\frac{1}{\sqrt{(1-\pi)n}}\sum_{t>\pi n} m_t(\hat\theta_{\mathrm{e,n}}) = \frac{1}{\sqrt{(1-\pi)n}}\sum_{t>\pi n} m_t(\theta_0) - D(D^TD)^{-1}D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_0)\right] + o_p(1)\,. \tag{164}$$

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

7. Proof of Proposition 7

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:

$$\mathcal{T}(\mathrm{Q}_0) = \left\{f \in L_0^2(\mathrm{Q}_0):\ \mathrm{E}^{\mathrm{Q}_0}\left[m(\cdot,\theta_0)f\right] \in \mathsf{lin}(D)\right\}, \tag{165}$$

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

$$\mathrm{E}^{\mathrm{Q}_0}\left[m(\cdot,\theta_0)f\right] = \mathrm{E}^{\mathrm{Q}_0}\left[h(\cdot,\theta_0)f\right]\ \text{ for all } f \in L_0^2(\mathrm{Q}_0). \tag{166}$$

Under the assumption, the following identity holds:

$$\mathrm{E}^{\mathrm{Q}_0}\left[h(\mathbf{y},\mathbf{y}',\theta_0)f(\mathbf{y},\mathbf{y}')\right] = \mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y},\mathbf{y}',\theta_0)f(\mathbf{y},\mathbf{y}')\right] - \sum_{k=1}^{\infty} A_k, \tag{167}$$

where

$$A_k = \mathrm{E}^{\mathrm{Q}_0}\left\{\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_{k-1}, \mathbf{y}_k, \theta_0)|\mathbf{y}_0 = \mathbf{y}\right] f(\mathbf{y}, \mathbf{y}')\right\} - \mathrm{E}^{\mathrm{Q}_0}\left\{\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_k, \mathbf{y}_{k+1}, \theta_0)|\mathbf{y}_1 = \mathbf{y}'\right] f(\mathbf{y}, \mathbf{y}')\right\}. \tag{168}$$

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

$$\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_k, \mathbf{y}_{k+1}, \theta_0)|\mathbf{y}_1 = \mathbf{y}'\right] f(\mathbf{y}, \mathbf{y}') = \mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_{k-1}, \mathbf{y}_k, \theta_0)|\mathbf{y}_0 = \mathbf{y}'\right] f(\mathbf{y}, \mathbf{y}'). \tag{169}$$

Thus, the equation (168) can be rewritten as

$$A_k = \mathrm{E}^{\mathrm{Q}_0}\left\{\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_{k-1}, \mathbf{y}_k, \theta_0)|\mathbf{y}_0 = \mathbf{y}\right] f(\mathbf{y}, \mathbf{y}')\right\} - \mathrm{E}^{\mathrm{Q}_0}\left\{\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_{k-1}, \mathbf{y}_k, \theta_0)|\mathbf{y}_0 = \mathbf{y}'\right] f(\mathbf{y}, \mathbf{y}')\right\}. \tag{170}$$

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

$$\begin{aligned}
&\mathrm{E}^{\mathrm{Q}_0}\left\{\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_{k-1}, \mathbf{y}_k, \theta_0)|\mathbf{y}_0 = \mathbf{y}'\right] f(\mathbf{y}, \mathbf{y}')\right\} \\
&\quad = \mathrm{E}^{\mathrm{Q}_0}\left\{\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_{k-1}, \mathbf{y}_k, \theta_0)|\mathbf{y}_0 = \mathbf{y}'\right] \mathrm{E}^{\mathrm{Q}_0}\left[f(\mathbf{y}, \mathbf{y}')|\mathbf{y}'\right]\right\} \text{ (Law of Iterated Projections)} \\
&\quad = \mathrm{E}^{\mathrm{Q}_0}\left\{\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_{k-1}, \mathbf{y}_k, \theta_0)|\mathbf{y}_0 = \mathbf{y}'\right] \mathrm{E}^{\mathrm{Q}_0}\left[f(\mathbf{y}', \mathbf{y})|\mathbf{y}'\right]\right\} \text{ (Proposition 3)} \\
&\quad = \mathrm{E}^{\mathrm{Q}_0}\left\{\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_{k-1}, \mathbf{y}_k, \theta_0)|\mathbf{y}_0 = \mathbf{y}'\right] f(\mathbf{y}', \mathbf{y})\right\} \text{ (Law of Iterated Projections)}
\end{aligned}$$

Therefore, $A_k = 0$ for all $k \geq 1$, and hence from (167), it follows that

$$\mathrm{E}^{\mathrm{Q}_0}\left[h(\mathbf{y}, \mathbf{y}', \theta_0) f(\mathbf{y}, \mathbf{y}')\right] = \mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}, \mathbf{y}', \theta_0) f(\mathbf{y}, \mathbf{y}')\right]. \tag{171}$$

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

$$\mathrm{E}^{\mathrm{Q}_0}\left[h(\mathbf{y}_0, \mathbf{y}_1, \theta_0) h(\mathbf{y}_0, \mathbf{y}_1, \theta_0)^T\right] = \sum_{\tau=-\infty}^{\infty} \mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_0, \mathbf{y}_1, \theta_0) m(\mathbf{y}_\tau, \mathbf{y}_{\tau+1}\theta_0)^T\right] = I. \tag{172}$$

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

$$\mathrm{E}^{\mathrm{Q}_0}\left\{\mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_k, \mathbf{y}_{k+1}, \theta_0)|\mathbf{y}_1\right] |\mathbf{y}_0\right\} = \mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_k, \mathbf{y}_{k+1}, \theta_0)|\mathbf{y}_0\right]. \tag{173}$$

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

8. Proof of Proposition 8.

The proof follows the standard GMM approximations in Hansen (1982), Hansen (2007b), and Hansen (2012).

9. Proof of Proposition 9.

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 $\theta_{\mathrm{e,n}}^{(1)}$, the constrained efficient GMM estimator $(\theta_{\mathrm{e,n}}^{(1)}, \psi_{\mathrm{e}}(\theta_{\mathrm{e,n}}^{(1)}))^T$ for the full model satisfies the first-order condition

$$\nabla J(\theta_{\rm e,n}^{(1)}, \psi_{\rm e}(\theta_{\rm e,n}^{(1)}); \mathbf{y}_{\rm e}^n) = \Gamma_{\theta,1}^T \Lambda_{\rm e,n}, \quad \text{with } \Gamma_{\theta,1} = [I, 0_{d_{\theta,1}\times d_{\theta,2}}], \tag{174}$$

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

$$\frac{1}{\sqrt{\pi n}}\Gamma_{\theta,1}^T \Lambda_{\rm e,n} = 2D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\le \pi n} m_t(\theta_0)\right] + 2\mathbf{I}_{\mathbb{Q}}\sqrt{\pi n}\begin{bmatrix} \theta_{\rm e,n}^{(1)} - \theta_0^{(1)} \\ \psi_{\rm e}(\theta_{\rm e,n}^{(1)}) - \theta_0^{(2)} \end{bmatrix} + o_p(1). \tag{175}$$

We first multiply both sides of (175) by $\Gamma_{\theta,1}\mathbf{I}_{\mathbb{Q}}^{-1}$, and then by $\left(\Gamma_{\theta,1}\mathbf{I}_{\mathbb{Q}}^{-1}\Gamma_{\theta,1}^T\right)^{-1}$. The optimal Lagrangian multipliers can be represented as

$$\begin{aligned}\frac{1}{\sqrt{\pi n}}\Lambda_{\rm e,n} = {} & 2\left(\Gamma_{\theta,1}\mathbf{I}_{\mathbb{Q}}^{-1}\Gamma_{\theta,1}^T\right)^{-1}\Gamma_{\theta,1}\mathbf{I}_{\mathbb{Q}}^{-1}D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_0)\right] \\ & + 2\left(\Gamma_{\theta,1}\mathbf{I}_{\mathbb{Q}}^{-1}\Gamma_{\theta,1}^T\right)^{-1}\sqrt{\pi n}(\theta_{\rm e,n}^{(1)} - \theta_0^{(1)}) + o_p(1).\end{aligned} \tag{176}$$

Substituting (175) and (176) into (174) yields

$$\begin{aligned}\frac{1}{\sqrt{\pi n}}\nabla J(\theta_{\rm e,n}^{(1)}, \psi_{\rm e}(\theta_{\rm e,n}^{(1)}); \mathbf{y}_{\rm e}^n) = {} & 2\Gamma_{\theta,1}^T\left(\Gamma_{\theta,1}\mathbf{I}_{\mathbb{Q}}^{-1}\Gamma_{\theta,1}^T\right)^{-1}\Gamma_{\theta,1}\mathbf{I}_{\mathbb{Q}}^{-1}D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_0)\right] \\ & + 2\Gamma_{\theta,1}^T(\Gamma_{\theta,1}\mathbf{I}_{\mathbb{Q}}^{-1}\Gamma_{\theta,1}^T)^{-1}\sqrt{\pi n}(\theta_{\rm e,n}^{(1)} - \theta_0^{(1)}) + o_p(1).\end{aligned} \tag{177}$$

According to Proposition 1, we substitute (46) into (177) and obtain

$$\begin{aligned}\frac{1}{\sqrt{\pi n}}\nabla J(\theta_{\rm e,n}^{(1)}, \psi_{\rm e}(\theta_{\rm e,n}^{(1)}); \mathbf{y}_{\rm e}^n) = {} & 2\Gamma_{\theta,1}^T\mathbf{I}_{\rm F}\Gamma_{\theta,1}\mathbf{I}_{\mathbb{Q}}^{-1}D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_0)\right] \\ & - 2\Gamma_{\theta,1}^T\mathbf{I}_{\rm F}\mathbf{I}_{\rm B}^{-1}D_{11}^T\left[\sqrt{\pi}\lambda^{(1)}(g_1) + \lambda^{(1)}(g_2)\int_0^\pi b(u)\mathrm{d}u/\sqrt{\pi}\right] + o_p(1).\end{aligned} \tag{178}$$

Based on (80), we have

$$\frac{1}{\sqrt{\pi n}}\nabla J(\theta_{\rm e,n}^{(1)}, \psi_{\rm e}(\theta_{\rm e,n}^{(1)}); \mathbf{y}_{\rm e}^n) = 2\Gamma_{\theta,1}^T\mathbf{I}_{\rm F}\left[\Gamma_{\theta,1}\mathbf{I}_{\mathbb{Q}}^{-1}D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_0)\right] - \mathbf{I}_{\rm B}^{-1}D_{11}^T\Gamma_{m,1}\nu(g,b,\pi)\right] + o_p(1). \tag{179}$$

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

$$\nabla J(\tilde{\theta}_{\rm e,n}^{(1)}, \psi_{\rm e}(\tilde{\theta}_{\rm e,n}^{(1)}); \mathbf{y}_{\rm e}^n) = \Gamma_{\theta,1}^T \Lambda_{\rm e,n}^{(1)}, \quad \text{with } \Gamma_{\theta,1} = [I, 0_{d_{\theta,1}\times d_{\theta,2}}], \tag{180}$$

and $\Lambda_{\rm e,n}^{(1)}$ is a $d_{\theta,1}\times 1$ vector of Lagrangian multipliers for the constraints $\Gamma_{\theta,1}\theta = \tilde{\theta}_{\rm e,n}^{(1)}$ in search of the constrained GMM estimator $(\tilde{\theta}_{\rm e,n}^{(1)}, \psi_{\rm e}(\tilde{\theta}_{\rm e,n}^{(1)}))^T$. The Taylor expansion of $\nabla J(\tilde{\theta}_{\rm e,n}^{(1)}, \psi_{\rm e}(\tilde{\theta}_{\rm e,n}^{(1)}); \mathbf{y}_{\rm e}^n)$ around $(\theta_{\rm e,n}^{(1)}, \psi_{\rm e}(\theta_{\rm e,n}^{(1)}))^T$, together with (180), leads to

$$\frac{1}{\sqrt{\pi n}}\Gamma_{\theta,1}^T\Lambda_{\mathrm{e,n}}^{(1)} = \frac{1}{\sqrt{\pi n}}\nabla J(\theta_{\mathrm{e,n}}^{(1)}, \psi_{\mathrm{e}}(\theta_{\mathrm{e,n}}^{(1)}); \mathbf{y}_{\mathrm{e}}^n) + 2\mathbf{I}_{\mathrm{Q}}\sqrt{\pi n}\begin{bmatrix} \tilde{\theta}_{\mathrm{e,n}}^{(1)} - \theta_{\mathrm{e,n}}^{(1)} \\ \psi_{\mathrm{e}}(\tilde{\theta}_{\mathrm{e,n}}^{(1)}) - \psi_{\mathrm{e}}(\theta_{\mathrm{e,n}}^{(1)}) \end{bmatrix} + o_p(1). \tag{181}$$

We first multiply both sides of (181) by $\Gamma_{\theta,1}\mathbf{I}_{\mathrm{Q}}^{-1}$, and then by $\left(\Gamma_{\theta,1}\mathbf{I}_{\mathrm{Q}}\Gamma_{\theta,1}^T\right)^{-1}$. The optimal Lagrangian multipliers can be represented as

$$\frac{1}{\sqrt{\pi n}}\Lambda_{\mathrm{e,n}}^{(1)} = \mathbf{I}_{\mathrm{F}}\Gamma_{\theta,1}\mathbf{I}_{\mathrm{Q}}^{-1}\frac{1}{\sqrt{\pi n}}\nabla J(\theta_{\mathrm{e,n}}^{(1)}, \psi_{\mathrm{e}}(\theta_{\mathrm{e,n}}^{(1)}); \mathbf{y}_{\mathrm{e}}^n) + 2\mathbf{I}_{\mathrm{F}}\sqrt{\pi n}(\tilde{\theta}_{\mathrm{e,n}}^{(1)} - \theta_{\mathrm{e,n}}^{(1)}) + o_p(1). \tag{182}$$

Further substituting (177) into equation (182) above yields

$$\frac{1}{\sqrt{\pi n}}\Lambda_{\mathrm{e,n}}^{(1)} = 2\mathbf{I}_{\mathrm{F}}\Gamma_{\theta,1}\mathbf{I}_{\mathrm{Q}}^{-1}D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_0)\right] + 2\mathbf{I}_{\mathrm{F}}\sqrt{\pi n}(\tilde{\theta}_{\mathrm{e,n}}^{(1)} - \theta_0^{(1)}) + o_p(1). \tag{183}$$

Based on Proposition 8, we obtain

$$\sqrt{\pi n}(\tilde{\theta}_{\mathrm{e,n}}^{(1)} - \theta_0^{(1)}) = -\mathbf{I}_{\mathrm{B}}^{-1}D_{11}^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t^{(1)}(\theta_0^{(1)})\right] + o_p(1). \tag{184}$$

Substituting (184) into (183) gives the following asymptotic representation of $\frac{1}{\sqrt{\pi n}}\Lambda_{\mathrm{e,n}}^{(1)}$:

$$\frac{1}{\sqrt{\pi n}}\Lambda_{\mathrm{e,n}}^{(1)} = 2\mathbf{I}_{\mathrm{F}}\Gamma_{\theta,1}\mathbf{I}_{\mathrm{Q}}^{-1}D^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t(\theta_0)\right] - 2\mathbf{I}_{\mathrm{F}}\mathbf{I}_{\mathrm{B}}^{-1}D_{11}^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t^{(1)}(\theta_0^{(1)})\right] + o_p(1). \tag{185}$$

We substitute (177) and (185) into (181) and multiply the both sides by $\mathbf{I}_{\mathrm{Q}}^{-1}/2$. The estimator can be represented by

$$\begin{aligned}\sqrt{\pi n}\begin{bmatrix} \tilde{\theta}_{\mathrm{e,n}}^{(1)} - \theta_{\mathrm{e,n}}^{(1)} \\ \psi_{\mathrm{e}}(\tilde{\theta}_{\mathrm{e,n}}^{(1)}) - \psi_{\mathrm{e}}(\theta_{\mathrm{e,n}}^{(1)}) \end{bmatrix} &= -\mathbf{I}_{\mathrm{Q}}^{-1}\Gamma_{\theta,1}^T\mathbf{I}_{\mathrm{F}}\mathbf{I}_{\mathrm{B}}^{-1}D_{11}^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t^{(1)}(\theta_0^{(1)}) - \nu_e^{(1)}(g,b,\pi)\right] + o_p(1) \\ &= -\mathbf{I}_{\mathrm{Q}}^{-1}\Gamma_{\theta,1}^T\mathbf{I}_{\mathrm{F}}\mathbf{I}_{\mathrm{B}}^{-1}D_{11}^T\left[\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n} m_t^{(1)}(\theta_{n,t}^{(1)})\right] + o_p(1).\end{aligned} \tag{186}$$

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

$$0 = \nabla J(\theta_{\mathrm{e,n}}^{(1)}, \psi_{\mathrm{e}}(\theta_{\mathrm{e,n}}^{(1)}); \mathbf{y}_{\mathrm{e}}^n) + 2\mathbf{I}_{\mathrm{Q}}\sqrt{\pi n}\begin{bmatrix} \hat{\theta}_{\mathrm{e,n}}^{(1)} - \theta_{\mathrm{e,n}}^{(1)} \\ \psi_{\mathrm{e}}(\hat{\theta}_{\mathrm{e,n}}^{(1)}) - \psi_{\mathrm{e}}(\theta_{\mathrm{e,n}}^{(1)}) \end{bmatrix} + o_p(1). \tag{187}$$

Substituting (177) into (187) and multiplying the both sides by $\mathbf{I}_{\mathrm{Q}}^{-1}/2$, we have

$$\sqrt{\pi n}\begin{bmatrix}\hat{\theta}^{(1)}_{\mathrm{e,n}}-\theta^{(1)}_{\mathrm{e,n}}\\ \psi_{\mathrm{e}}(\hat{\theta}^{(1)}_{\mathrm{e,n}})-\psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}})\end{bmatrix} = -\mathbf{I}_{\mathrm{Q}}^{-1}\Gamma_{\theta,1}^{T}\mathbf{I}_{\mathrm{F}}\left\{\Gamma_{\theta,1}\mathbf{I}_{\mathrm{Q}}^{-1}D^{T}\left[\frac{1}{\sqrt{\pi n}}\sum_{t\le\pi n}m_t(\theta_0)\right]-\mathbf{I}_{\mathrm{B}}^{-1}D_{11}^{T}\Gamma_{m,1}\nu_e\right\}+o_p(1). \tag{188}$$

10. Proof of Proposition 10

We first approximate $\mathcal{L}(\tilde{\theta}^{(1)}_{\mathrm{e,n}};\mathbf{y}^n_{\mathrm{e}})$. According to the second-order Taylor expansion around $(\theta^{(1)}_{\mathrm{e,n}},\psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}}))$, it follows that

$$\begin{aligned}\mathcal{L}(\tilde{\theta}^{(1)}_{\mathrm{e,n}};\mathbf{y}^n_{\mathrm{e}}) =& \left[\frac{1}{\sqrt{\pi n}}\nabla J(\theta^{(1)}_{\mathrm{e,n}},\psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}});\mathbf{y}^n_{\mathrm{e}})\right]^T\sqrt{\pi n}\begin{bmatrix}\tilde{\theta}^{(1)}_{\mathrm{e,n}}-\theta^{(1)}_{\mathrm{e,n}}\\ \psi_{\mathrm{e}}(\tilde{\theta}^{(1)}_{\mathrm{e,n}})-\psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}})\end{bmatrix}\\ &+\sqrt{\pi n}\begin{bmatrix}\tilde{\theta}^{(1)}_{\mathrm{e,n}}-\theta^{(1)}_{\mathrm{e,n}}\\ \psi_{\mathrm{e}}(\tilde{\theta}^{(1)}_{\mathrm{e,n}})-\psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}})\end{bmatrix}^T\mathbf{I}_{\mathrm{Q}}\sqrt{\pi n}\begin{bmatrix}\tilde{\theta}^{(1)}_{\mathrm{e,n}}-\theta^{(1)}_{\mathrm{e,n}}\\ \psi_{\mathrm{e}}(\tilde{\theta}^{(1)}_{\mathrm{e,n}})-\psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}})\end{bmatrix}+o_p(1).\end{aligned} \tag{189}$$

Thus, following (179) and (186),

$$\mathcal{L}(\tilde{\theta}^{(1)}_{\mathrm{e,n}};\mathbf{y}^n_{\mathrm{e}}) = -2\left[L_{\mathrm{F}}\zeta_{\mathrm{e,n}}+L_{\Delta}\nu_e\right]^T\mathbf{I}_{\mathrm{F}}L_{\mathrm{B}}\zeta_{\mathrm{e,n}}+\zeta_{\mathrm{e,n}}^TL_{\mathrm{B}}^T\mathbf{I}_{\mathrm{F}}L_{\mathrm{B}}\zeta_{\mathrm{e,n}}+o_p(1). \tag{190}$$

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

$$\begin{aligned}\mathcal{L}(\tilde{\theta}^{(1)}_{\mathrm{e,n}};\mathbf{y}^n_{\mathrm{o}}) =& \left[\frac{1}{\sqrt{\pi n}}\nabla J(\theta^{(1)}_{\mathrm{o,n}},\psi_{\mathrm{o}}(\theta^{(1)}_{\mathrm{o,n}});\mathbf{y}^n_{\mathrm{o}})\right]^T\sqrt{\pi n}\begin{bmatrix}\tilde{\theta}^{(1)}_{\mathrm{e,n}}-\theta^{(1)}_{\mathrm{o,n}}\\ \psi_{\mathrm{o}}(\tilde{\theta}^{(1)}_{\mathrm{e,n}})-\psi_{\mathrm{o}}(\theta^{(1)}_{\mathrm{o,n}})\end{bmatrix}\\ &+\sqrt{\pi n}\begin{bmatrix}\tilde{\theta}^{(1)}_{\mathrm{e,n}}-\theta^{(1)}_{\mathrm{o,n}}\\ \psi_{\mathrm{o}}(\tilde{\theta}^{(1)}_{\mathrm{e,n}})-\psi_{\mathrm{o}}(\theta^{(1)}_{\mathrm{o,n}})\end{bmatrix}^T\mathbf{I}_{\mathrm{Q}}\sqrt{\pi n}\begin{bmatrix}\tilde{\theta}^{(1)}_{\mathrm{e,n}}-\theta^{(1)}_{\mathrm{o,n}}\\ \psi_{\mathrm{o}}(\tilde{\theta}^{(1)}_{\mathrm{e,n}})-\psi_{\mathrm{o}}(\theta^{(1)}_{\mathrm{o,n}})\end{bmatrix}+o_p(1).\end{aligned} \tag{191}$$

Similarly,

$$\mathcal{L}(\tilde{\theta}^{(1)}_{\mathrm{e,n}};\mathbf{y}^n_{\mathrm{o}}) = -2\left[L_{\mathrm{F}}\zeta_{\mathrm{o,n}}+L_{\Delta}\nu_o\right]^T\mathbf{I}_{\mathrm{F}}L_{\mathrm{B}}\zeta_{\mathrm{e,n}}+\zeta_{\mathrm{e,n}}^TL_{\mathrm{B}}^T\mathbf{I}_{\mathrm{F}}L_{\mathrm{B}}\zeta_{\mathrm{e,n}}+o_p(1). \tag{192}$$

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

$$\begin{aligned}\mathcal{L}(\hat{\theta}^{(1)}_{\mathrm{e,n}};\mathbf{y}^n_{\mathrm{e}}) =& \left[\frac{1}{\sqrt{\pi n}}\nabla J(\theta^{(1)}_{\mathrm{e,n}},\psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}});\mathbf{y}^n_{\mathrm{e}})\right]^T\sqrt{\pi n}\begin{bmatrix}\hat{\theta}^{(1)}_{\mathrm{e,n}}-\theta^{(1)}_{\mathrm{e,n}}\\ \psi_{\mathrm{e}}(\hat{\theta}^{(1)}_{\mathrm{e,n}})-\psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}})\end{bmatrix}\\ &+\sqrt{\pi n}\begin{bmatrix}\hat{\theta}^{(1)}_{\mathrm{e,n}}-\theta^{(1)}_{\mathrm{e,n}}\\ \psi_{\mathrm{e}}(\hat{\theta}^{(1)}_{\mathrm{e,n}})-\psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}})\end{bmatrix}^T\mathbf{I}_{\mathrm{Q}}\sqrt{\pi n}\begin{bmatrix}\hat{\theta}^{(1)}_{\mathrm{e,n}}-\theta^{(1)}_{\mathrm{e,n}}\\ \psi_{\mathrm{e}}(\hat{\theta}^{(1)}_{\mathrm{e,n}})-\psi_{\mathrm{e}}(\theta^{(1)}_{\mathrm{e,n}})\end{bmatrix}+o_p(1).\end{aligned} \tag{193}$$

Similarly,

$$\mathcal{L}(\hat{\theta}^{(1)}_{\mathrm{e,n}};\mathbf{y}^n_{\mathrm{e}}) = -\left[L_{\mathrm{F}}\zeta_{\mathrm{e,n}}+L_{\Delta}\nu_e\right]^T\mathbf{I}_{\mathrm{F}}\left[L_{\mathrm{F}}\zeta_{\mathrm{e,n}}+L_{\Delta}\nu_e\right]+o_p(1). \tag{194}$$

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

$$\mathcal{L}(\hat{\theta}^{(1)}_{\mathrm{e,n}};\mathbf{y}^n_{\mathrm{o}}) = \left[\frac{1}{\sqrt{\pi n}}\nabla J(\theta^{(1)}_{\mathrm{o,n}},\psi_{\mathrm{o}}(\theta^{(1)}_{\mathrm{o,n}});\mathbf{y}^n_{\mathrm{o}})\right]^T \sqrt{\pi n}\begin{bmatrix}\hat{\theta}^{(1)}_{\mathrm{e,n}}-\theta^{(1)}_{\mathrm{o,n}}\\ \psi_{\mathrm{o}}(\hat{\theta}^{(1)}_{\mathrm{e,n}})-\psi_{\mathrm{o}}(\theta^{(1)}_{\mathrm{o,n}})\end{bmatrix} + \sqrt{\pi n}\begin{bmatrix}\hat{\theta}^{(1)}_{\mathrm{e,n}}-\theta^{(1)}_{\mathrm{o,n}}\\ \psi_{\mathrm{o}}(\hat{\theta}^{(1)}_{\mathrm{e,n}})-\psi_{\mathrm{o}}(\theta^{(1)}_{\mathrm{o,n}})\end{bmatrix}^T \mathbf{I}_{\mathcal{Q}}\sqrt{\pi n}\begin{bmatrix}\hat{\theta}^{(1)}_{\mathrm{e,n}}-\theta^{(1)}_{\mathrm{o,n}}\\ \psi_{\mathrm{o}}(\hat{\theta}^{(1)}_{\mathrm{e,n}})-\psi_{\mathrm{o}}(\theta^{(1)}_{\mathrm{o,n}})\end{bmatrix} + o_p(1). \tag{195}$$

Similarly,

$$\mathcal{L}(\tilde{\theta}^{(1)}_{\mathrm{e,n}};\mathbf{y}^n_{\mathrm{o}}) = -2\left[L_{\mathrm{F}}\zeta_{\mathrm{o,n}}+L_{\Delta}\nu_o\right]^T\mathbf{I}_{\mathrm{F}}\left[L_{\mathrm{F}}\zeta_{\mathrm{e,n}}+L_{\Delta}\nu_e\right] + \left[L_{\mathrm{F}}\zeta_{\mathrm{e,n}}+L_{\Delta}\nu_e\right]^T\mathbf{I}_{\mathrm{F}}\left[L_{\mathrm{F}}\zeta_{\mathrm{e,n}}+L_{\Delta}\nu_e\right] + o_p(1). \tag{196}$$

APPENDIX E: PROOFS OF COROLLARIES

1. Proof of Corollary 1

We can derive the result following the same derivations for (156) under the baseline GMM model $\mathcal{Q}^{(1)}$.

2. Proof of Corollary 2

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

APPENDIX F: DERIVATION OF THE DISASTER RISK MODEL

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 $\mathcal{M}_{t,t+1}=\delta_{\mathrm{D}}e^{-\gamma_{\mathrm{D}}g_{t+1}}$. The Euler equations for the risk-free rate and the market equity return are

$$1=\mathbb{E}_t\left[\mathcal{M}_{t,t+1}e^{r_{M,t+1}}\right] \quad \text{and} \quad e^{-r_{f,t}}=\mathbb{E}_t\left[\mathcal{M}_{t,t+1}\right]. \tag{197}$$

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

$$\mathbb{E}_t\left[\mathcal{M}_{t,t+1}\right]=\mathbb{E}_t\left[\mathcal{M}_{t,t+1}e^{r_{t+1}}\right]. \tag{198}$$

The left-hand side of (198) is equal to

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

and the right-hand side of (198) is equal to

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

Thus, the Euler equation (198) can be rewritten as

$$(199)\qquad (1-p)e^{-\gamma_{\mathrm{D}}\mu+\frac{1}{2}\gamma_{\mathrm{D}}^2\sigma^2}\left[e^{\eta+\frac{\tau^2}{2}-\gamma_{\mathrm{D}}\rho\sigma\tau}-1\right] = p\Delta(\xi),\ \text{ where } \Delta(\xi)=\xi\left(\frac{e^{\gamma_{\mathrm{D}}\underline{v}}}{\xi-\gamma_{\mathrm{D}}}-\frac{e^{\frac{\varsigma^2}{2}+(\gamma_{\mathrm{D}}-b)\underline{v}}}{\xi+b-\gamma_{\mathrm{D}}}\right).$$

Using the Taylor expansion, we obtain the approximation

$$(200)\qquad e^{\eta+\frac{\tau^2}{2}-\gamma_{\mathrm{D}}\rho\sigma\tau}-1\approx \eta+\frac{\tau^2}{2}-\gamma_{\mathrm{D}}\rho\sigma\tau,$$

which, combined with (199), gives the approximated Euler equation in (69).

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

$$(201)\qquad D_{11}=\begin{bmatrix}-1 & 0\\ 0 & -\frac{p}{\xi^2}\end{bmatrix}\text{ and }\Omega_{11}=\begin{bmatrix}p(1-p) & 0\\ 0 & (1-p)\sigma^2+\frac{p}{\xi^2}\end{bmatrix}\approx\begin{bmatrix}p(1-p) & 0\\ 0 & \frac{p}{\xi^2}\end{bmatrix},\text{ respectively.}$$

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

$$(202)\qquad \Sigma_1 = D_{11}^T\Omega_{11}^{-1}D_{11}\approx\begin{bmatrix}\frac{1}{p(1-p)} & 0\\ 0 & \frac{p}{\xi^2}\end{bmatrix}.$$

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

$$(203)\qquad D=\begin{bmatrix}-1 & 0\\ 0 & -\frac{p}{\xi^2}\\ -(1-p)\frac{\partial\eta(p,\xi)}{\partial p} & -(1-p)\frac{\partial\eta(p,\xi)}{\partial\xi}-\frac{pb}{\xi^2}\end{bmatrix},$$

and

$$(204)\qquad \Omega=\begin{bmatrix}p(1-p) & 0 & 0\\ 0 & (1-p)\sigma^2+\frac{p}{\xi^2} & (1-p)\rho\sigma\tau+bp/\xi^2\\ 0 & (1-p)\rho\sigma\tau+bp/\xi^2 & (1-p)\tau^2+pb^2/\xi^2\end{bmatrix},$$

where

$$(205)\qquad \eta(p,\xi)\equiv\gamma_{\mathrm{D}}\rho\sigma\tau-\frac{\tau^2}{2}+\ln\left[1+e^{\gamma_{\mathrm{D}}\mu-\frac{\gamma_{\mathrm{D}}^2\sigma^2}{2}}\xi\left(\frac{e^{\gamma_{\mathrm{D}}\underline{v}}}{\xi-\gamma_{\mathrm{D}}}-e^{\frac{1}{2}\varsigma^2}\frac{e^{(\gamma_{\mathrm{D}}-b)\underline{v}}}{\xi+b-\gamma_{\mathrm{D}}}\right)\frac{p}{1-p}\right].$$

We can also derive the closed-form solution for the dark matter measure in (71) if we use the approximate Euler equation in (69). In this case, using the notation introduced in (69) and (72), we can express the information matrix for (p,ξ) under the full GMM model as

$$\Sigma \approx \begin{bmatrix} \frac{1}{p(1-p)} + \frac{\Delta(\xi)^2}{(1-\rho^2)\tau^2}\frac{e^{2\gamma_{\mathrm{D}}\mu-\gamma_{\mathrm{D}}^2\sigma^2}}{(1-p)^3} & \frac{p}{(1-\rho^2)\tau^2}\frac{e^{2\gamma_{\mathrm{D}}\mu-\gamma_{\mathrm{D}}^2\sigma^2}}{(1-p)^2}\Delta(\xi)\dot{\Delta}(\xi) \\ \frac{p}{(1-\rho^2)\tau^2}\frac{e^{2\gamma_{\mathrm{D}}\mu-\gamma_{\mathrm{D}}^2\sigma^2}}{(1-p)^2}\Delta(\xi)\dot{\Delta}(\xi) & \frac{p}{\xi^2} + \frac{\dot{\Delta}(\xi)^2}{(1-\rho^2)\tau^2}e^{2\gamma_{\mathrm{D}}\mu-\gamma_{\mathrm{D}}^2\sigma^2}\frac{p^2}{1-p} \end{bmatrix}. \tag{206}$$

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 $\varrho(\theta)$ in (71).

APPENDIX G: MISCELLANEOUS PROOFS AND DERIVATIONS

G.1. *Moment Rotations*

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

$$\Omega^{-1} = L^T L. \tag{207}$$

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

$$\tilde{m}_t(\theta) = L m_t(\theta) = \begin{bmatrix} L_{11} m_t^{(1)}(\theta^{(1)}) \\ L_{21} m_t^{(1)}(\theta^{(1)}) + L_{22} m_t^{(2)}(\theta) \end{bmatrix} = \begin{bmatrix} \tilde{m}_t^{(1)}(\theta^{(1)}) \\ \tilde{m}_t^{(2)}(\theta) \end{bmatrix}. \tag{208}$$

Further, we let

$$\tilde{D} = LD = \begin{bmatrix} L_{11}D_{11} & 0 \\ L_{21}D_{11} + L_{22}D_{21} & L_{22}D_{22} \end{bmatrix} = \begin{bmatrix} \tilde{D}_{11} & 0 \\ \tilde{D}_{21} & \tilde{D}_{22} \end{bmatrix}. \tag{209}$$

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

G.2. *Hellinger-Differentiability Condition*

The condition (35) is equivalent to the condition

$$\left(\frac{\mathrm{dQ}_{s,g}}{\mathrm{dQ}}\right)^{1/2} = 1 + \frac{1}{2}sg + s\varepsilon(s), \tag{210}$$

where $\varepsilon(s)$ converges to zero in $L^2(\mathrm{Q})$ as $s \to 0$. Equation (210) is equivalent to

$$\lim_{s\to 0}\int\left[\frac{1}{s}\left(\left(\frac{\mathrm{dQ}_{s,g}}{\mathrm{dQ}}\right)^{1/2} - 1\right) - \frac{1}{2}g\right]^2 \mathrm{dQ} = \lim_{s\to 0}\int \varepsilon(s)^2 \mathrm{dQ} = 0. \tag{211}$$

G.3. *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[28]

$$P_2 = I - D_2 \left(D_2^T D_2\right)^{-1} D_2^T = \begin{bmatrix} I & 0 \\ 0 & \Lambda_2 \end{bmatrix}. \tag{212}$$

Using the rules for the inversion of partitioned matrices, we have

$$\left(D^T D\right)^{-1} = \begin{bmatrix} \left(D_1^T P_2 D_1\right)^{-1} & -\left(D_1^T P_2 D_1\right)^{-1} D_1^T D_2 \left(D_2^T D_2\right)^{-1} \\ -\left(D_2^T D_2\right)^{-1} D_2^T D_1 \left(D_1^T P_2 D_1\right)^{-1} & \left(D_2^T D_2\right)^{-1} + \left(D_2^T D_2\right)^{-1} D_2^T D_1 \left(D_1^T P_2 D_1\right)^{-1} D_1^T D_2 \left(D_2^T D_2\right)^{-1} \end{bmatrix}.$$

We can then show that

$$\begin{aligned} D\left(D^T D\right)^{-1} D^T = {} & D_1 \left(D_1^T P_2 D_1\right)^{-1} D_1^T - D_1 \left(D_1^T P_2 D_1\right)^{-1} D_1^T (I - P_2) \\ & - (I - P_2) D_1 \left(D_1^T P_2 D_1\right)^{-1} D_1^T \\ & + (I - P_2) + (I - M_2) D_1 \left(D_1^T P_2 D_1\right)^{-1} D_1^T (I - P_2) \\ = {} & I - P_2 + P_2 D_1 \left(D_1^T P_2 D_1\right)^{-1} D_1^T P_2. \end{aligned} \tag{213}$$

We conclude that

$$\Lambda = I - D\left(D^T D\right)^{-1} D^T = P_2 - P_2 D_1 \left(D_1^T P_2 D_1\right)^{-1} D_1^T P_2. \tag{214}$$

Recall that $\mathbf{I}_{\mathrm{F}} = D_1^T P_2 D_1$ (from Equation (31)). The matrix Λ can be rewritten as

$$\Lambda = \begin{bmatrix} I - D_{11} \mathbf{I}_{\mathrm{F}}^{-1} D_{11}^T & D_{11} \mathbf{I}_{\mathrm{F}}^{-1} D_{11}^T \Lambda_2 \\ \Lambda_2 D_{11} \mathbf{I}_{\mathrm{F}}^{-1} D_{11}^T & \Lambda_2 - \Lambda_2 D_{21} \mathbf{I}_{\mathrm{F}}^{-1} D_{21}^T \Lambda_2 \end{bmatrix}. \tag{215}$$

APPENDIX H: DARK MATTER OF LONG-RUN RISK MODELS

In the second example, we consider a long-run risk model similar to Bansal and Yaron (2004) and Bansal, Kiku, and Yaron (2012). In the model, the representative agent has recursive preferences as in Epstein and Zin (1989) and Weil (1989) and maximizes her lifetime utility,

$$V_t = \left[(1 - \delta_{\mathrm{L}}) C_t^{1-1/\psi_{\mathrm{L}}} + \delta_{\mathrm{L}} \left(\mathbb{E}_t \left[V_{t+1}^{1-\gamma_{\mathrm{L}}}\right]\right)^{\frac{1-1/\psi_{\mathrm{L}}}{1-\gamma_{\mathrm{L}}}}\right]^{\frac{1}{1-1/\psi_{\mathrm{L}}}}, \tag{216}$$

where C_t is consumption at time t, δ_{L} is the rate of time preference, γ_{L} is the coefficient of risk aversion for timeless gambles, and ψ_{L} is the elasticity of intertemporal substitution when there is perfect certainty. The log growth rate of consumption Δc_t, the expected consumption growth x_t, and the conditional volatility of

[28] The matrix inversion is the generalized inversion.

consumption growth σ_t evolve as follows:

$$\Delta c_{t+1} = \mu_c + x_t + \sigma_t \epsilon_{c,t+1}, \tag{217a}$$

$$x_{t+1} = \rho x_t + \varphi_x \sigma_t \epsilon_{x,t+1}, \tag{217b}$$

$$\widetilde{\sigma}^2_{t+1} = \overline{\sigma}^2 + \nu(\widetilde{\sigma}^2_t - \overline{\sigma}^2) + \sigma_w \epsilon_{\sigma,t+1}, \tag{217c}$$

$$\sigma^2_{t+1} = \max\left(\underline{\sigma}^2, \widetilde{\sigma}^2_{t+1}\right), \tag{217d}$$

where the shocks $\epsilon_{c,t}, \epsilon_{x,t}$, and $\epsilon_{\sigma,t}$ are i.i.d. $N(0,1)$ and mutually independent. The volatility process (217c) potentially allows for negative values of $\widetilde{\sigma}^2_t$. Following the literature, we impose a small positive lower bound $\underline{\sigma}$ (= 1 bps) on variance σ_t in solutions and simulations. Negative values of conditional variance can also be avoided by changing the specification. For example, the process of σ^2_t can be specified as a discrete-time version of the square root process.[29]

Next, the log dividend growth Δd_t follows

$$\Delta d_{t+1} = \mu_d + \phi_d x_t + \varphi_{d,c} \sigma_t \epsilon_{c,t+1} + \varphi_{d,d} \sigma_t \epsilon_{d,t+1}, \tag{218}$$

where the shocks $\epsilon_{d,t}$ are *i.i.d.* $N(0,1)$ and independent of the other shocks in (217a–217c). The equilibrium excess log return follows

$$r^e_{t+1} = \mu^e_{r,t} + \beta_c \sigma_t \epsilon_{c,t+1} + \beta_x \sigma_t \epsilon_{x,t+1} + \beta_\sigma \sigma_w \epsilon_{\sigma,t+1} + \varphi_{d,d} \sigma_t \epsilon_{d,t+1}, \tag{219}$$

where the conditional average log excess return is

$$\mu^e_{r,t} = \lambda_c \beta_c \sigma^2_t + \lambda_x \beta_x \varphi_x \sigma^2_t + \lambda_\sigma \beta_\sigma \sigma^2_w - \frac{1}{2} \sigma^2_{r_m,t}, \tag{220}$$

$$\text{where } \sigma^2_{r_m,t} = \beta^2_c \sigma^2_t + \beta^2_x \sigma^2_t + \beta^2_\sigma \sigma^2_w + \varphi^2_{d,d} \sigma^2_t. \tag{221}$$

The expressions for λ_c, λ_x, λ_σ, β_c, β_x, β_σ, and $A_{m,j}$ with $j = 0, 1, 2$ are presented in Online Appendix.

The model contains stochastic singularities. For instance, the excess log market return r^e_{t+1} is a deterministic function of $\Delta c_{t+1}, \Delta d_{t+1}, x_{t+1}, x_t$, σ^2_{t+1}, and σ^2_t. The log price-dividend ratio $z_{m,t}$ is a deterministic function of x_t and σ^2_t. To avoid the problems posed by stochastic singularities, we add noise shocks $\varphi_r \sigma_t \epsilon_{r,t+1}$ to stock returns, with $\epsilon_{r,t}$ being i.i.d. standard normal variables and mutually independent of other variables. This is a standard approach in the dynamic stochastic general equilibrium (DSGE) literature for dealing with stochastic singularity. The stochastic singularity is one of the main reasons why we adopt the moment-based method, rather than the likelihood-based method, to evaluate and characterize the structural models.

[29] To ensure that our analysis applies as closely as possible to the model as formulated in the literature, we deliberately choose to follow Bansal, Kiku, and Yaron (2012, 2016a). In particular, following these papers, we also solve the model using a local log-linear expansion around the steady state. Thus, the approximate price-dividend ratio is not affected by the presence of the lower bound on the conditional variance process. As Bansal, Kiku, and Yaron (2016a) show, the resulting approximation error, when compared to the global numerical solution, is negligible. When computing our asymptotic Fisher fragility measure, we impose the lower bound on conditional variance. Thus, our asymptotic measure reflects the specification of the conditional variance process with the lower bound.

TABLE I

Parameters of the Benchmark Long-Run Risk Model (M1).

Preferences	δ_{L}	γ_{L}	ψ_{L}			
	0.9989	10	1.5			
Consumption	μ_c	ρ	φ_x	$\overline{\sigma}$	ν	σ_w
	0.0015	0.975	0.038	0.0072	0.999	$2.8e-6$
Dividends	μ_d	ϕ_d	$\varphi_{d,c}$	$\varphi_{d,d}$		
	0.0015	2.5	2.6	5.96		
Returns	φ_r					
	3.0					

Note: Model M2 has $\nu = 0.98$ and $\gamma_{\mathrm{L}} = 27$, while the other parameters are the same as in Model M1.

Quantitative analysis

In our computation of the dark matter measure, we consider a system of moment restrictions based on the joint dynamics of time series $(\Delta c_{t+1}, x_t, \sigma_t^2, \Delta d_{t+1}, r_{t+1}^e)$.[30]

We choose the model of consumption (217a)–(217d) and dividend (218) as the baseline model $\mathbb{Q}^{(1)}$ with variables $(\Delta c_{t+1}, x_t, \sigma_t^2, \Delta d_{t+1})$. The moment restrictions associated with the baseline model constitute the set of baseline moment restrictions. We assume that the econometrician observes the processes of consumption and dividends, including the conditional mean and volatility x_t and σ_t^2, and the process of asset returns.

The simulated moments and sample moments are listed in Table II. The sample moments are based on annual data from 1930 to 2008, and the simulated moments are 80-year annual data aggregated from monthly simulated data.

TABLE II

Simulated and Sample Moments.

Moment	Data Estimate	Model 1 5%	Model 1 Median	Model 1 95%	Model 2 5%	Model 2 Median	Model 2 95%
$\mathbb{E}[r_M - r_f]$	7.09	2.33	5.88	10.58	3.65	6.78	10.05
$\mathbb{E}[r_M]$	7.66	2.91	6.66	11.20	4.42	7.75	11.20
$\sigma(r_M)$	20.28	12.10	20.99	29.11	15.01	17.55	20.33
$\mathbb{E}[r_f]$	0.57	-0.20	0.77	1.45	0.47	0.96	1.46
$\sigma(r_f)$	2.86	0.64	1.07	1.62	0.73	0.94	1.23
$\mathbb{E}[p-d]$	3.36	2.69	2.99	3.30	2.77	2.81	2.85
$\sigma(p-d)$	0.45	0.13	0.18	0.28	0.09	0.11	0.13

Accordingly, the baseline parameters are $\theta^{(1)} = (\mu_c, \rho, \varphi_x, \overline{\sigma}^2, \nu, \sigma_w, \mu_d, \phi_d, \varphi_{d,c}, \varphi_{d,d})$ with $d_{\theta,1} = 10$. By measuring the fragility of the long-run risk model relative to this particular baseline, we can interpret the fragility measure as quantifying the information that asset pricing restrictions provide for the consumption and dividend dynamics above and beyond the information contained in consumption and dividend data. We explicitly account for uncertainty about preference parameters γ_{L} and ψ_{L} by including them in the nuisance parameter vector $\theta^{(2)}$. Thus, $\theta^{(2)} = (\gamma_{\mathrm{L}}, \psi_{\mathrm{L}})$. The extra data investigated by the full structural model $\mathbb{Q}$ are the excess log market returns r_{t+1}^e. Other parameters, included in the auxiliary parameter vector $(\delta_{\mathrm{L}}, \varphi_r)$, are fixed at known

[30] Details can be found in the Online Appendix.

TABLE III

DARK MATTER MEASURES FOR THE LONG-RUN RISK MODELS

Model	$\varrho(\theta_0)$	$\theta^{(1)}$									
		μ_c	ρ	φ_x	$\overline{\sigma}^2$	ν	σ_w	μ_d	ϕ_d	$\varphi_{d,c}$	$\varphi_{d,d}$
	I. Nuisance parameter vector ψ: (γ_L, ψ_L)										
(M1)	196.3	1.0	1.1	1.0	48.9	97.8	1.0	1.0	3.4	1.0	1.0
(M2)	21.1	1.0	1.1	1.0	1.0	3.4	1.0	1.4	4.2	1.0	1.0
	II. Nuisance parameter vector ψ: empty										
(M1)	$3.57 \cdot 10^5$	1.0	2.1	1.1	115.6	117.5	1.3	1.1	7.1	1.0	1.0
(M2)	287.7	1.0	2.5	1.0	1.0	6.3	1.0	1.9	31.3	1.0	1.0

Note: The direction corresponding to the worst-case one-dimensional fragility measure $\varrho(\theta_0)$ for Model M1 is given by $v^*_{\max} = [0.000, 0.000, -0.000, 0.020, -0.001, 0.999, -0.001, 0.000, -0.000, 0.000]$. Model M2 has $\nu = 0.98$ and $\gamma_L = 27$ with other parameters unchanged. In Panel I, we account for the uncertainty of preference parameters $\theta^{(2)} = (\gamma_L, \psi_L)$ to compute the dark matter measure of the baseline parameters $\theta^{(1)}$. That is, $\theta^{(1)}$ and $\theta^{(2)}$ are treated as unknown parameters, while (δ_L, φ_r) are treated as auxiliary parameters with known fixed values in Panel I. In Panel II, these preference parameters are fixed as auxiliary parameters with the nuisance parameter vector $\theta^{(2)}$ empty. That is, $\theta^{(1)}$ is treated as unknown parameters, while (γ_L, ψ_L) and (δ_L, φ_r) are treated as auxiliary parameters with known fixed values in Panel II.

values. These values form a part of the imposed functional-form specification of the structural component that is under fragility assessment. Note that the baseline model covers the joint dynamics of consumption growth and dividend growth. The structural model adds the description of the distribution of stock returns in relation to the consumption and dividend growth processes.

The parameter values of Model M1 follow Bansal, Kiku, and Yaron (2012) and are summarized in Table I. As Bansal, Kiku, and Yaron (2012) (Table 2, page 194) show, the simulated first and second moments, based on the parametrization of Model M1, match the set of key asset pricing moments in the data reasonably well. The same is true for Model M2, whose parameter values are also reported in Table III (see Table II for the asset pricing moment matching). Our main purpose of presenting Table III is to compare the fragility of M1 and M2, two different calibrations of the LRR model, while the two panels are meant to further illustrate the fact that different treatments of the nuisance parameters can also affect the fragility measure.

First, consider Panel I of Table III. This panel contains fragility measures computed under the specification that treats preference parameters as unknown nuisance parameters whose uncertainty needs to be taken into account when computing the dark matter measure. The row (M1) of Panel I reports fragility measures for Model M1 when the unknown nuisance parameters are γ_L and ψ_L. The dark matter measure $\varrho(\theta_0) = 196.3$ is large. This implies that to match the precision of the estimator for the full structural model in all directions, the estimator based on the baseline model would require a time-series sample that is 196.3 times as long.

A high value of $\varrho(\theta_0)$ suggests that the asset pricing implications of the structural model are highly sensitive to plausible perturbations of parameter values. We compute the fragility measure for each individual parameter in the vector $\theta^{(1)}$. All of the univariate measures are much lower than the worst-case one-dimensional fragility measure (i.e. the dark matter measure) $\varrho(\theta_0)$, with a larger fragility measure for $\overline{\sigma}^2$ (the long-run variance of consumption growth) and ν (the persistence of conditional variance of consumption growth) than for the other individual parameters. This shows that it is not sufficient to consider perturbations of parameters one at a time to quantify model fragility; Müller (2012) highlighted a similar insight on sensitivity analysis.

In comparison, in Panel II of Table III we show fragility measures under the specification that ignores the uncertainty about preference parameters. This type of analysis is sensible if the model is not fully estimated, but rather the preference parameters are fixed at certain values. For instance, one may specifically design a model to capture the moments of asset returns with a low value of risk aversion. In that case, the choice of the preference parameters is effectively subsumed by the specification of the functional form of the model, and treating them as auxiliary parameters is in line with the logic of the model construction. The fragility measures in Panel II are higher. In particular, the worst-case one-dimensional fragility (i.e. the dark matter measure) $\varrho(\theta_0)$ increases dramatically from 196.3 to $3.57 \cdot 10^5$.

In our model we have assumed that the conditional mean and volatility of consumption growth, x_t and σ_t, are observable. An interesting question is whether the model becomes more or less fragile when agents observe x_t and σ_t but the econometrician does not (e.g., Schorfheide, Song, and Yaron, 2018). When the agents themselves need to learn about the latent states and potentially deal with model uncertainty (e.g., Collin-Dufresne, Johannes, and Lochstoer, 2016; Hansen and Sargent, 2010), the cross-equation restrictions implied by asset prices differ from the case of fully observable state variables. It is therefore difficult to establish the precise effect of limited observability on model fragility without further analysis, which is beyond the scope of this paper. Numerically, the assumption that x_t and σ_t^2 are observable means that we do not need to integrate out x_t or σ_t^2 in the moment restrictions when computing the model fragility measure. Furthermore, since we are examining the fragility of a specific calibration of the model, we can compute the fragility measure under the set of calibrated parameter values, instead of having to first filter out the values of x_t and σ_t^2 from the data, as in Constantinides and Ghosh (2011), Bansal, Kiku, and Yaron (2016b), and Schorfheide, Song, and Yaron (2018), and then estimate the corresponding parameter values.

Monte Carlo experiments

We use simulations to illustrate the connections between the dark matter measure, internal refutability, and external validity of long-run risk models in finite samples. In the simulation experiment, we assume that all the parameters except ν are treated as auxiliary parameters, fixed at known constant values and thus subsumed into the functional form of the moment function (i.e. model specifications). From the dark matter evaluation in Table III, we learn that the assumed identification of ν (i.e., the uncertainty of ν) is a major source of model fragility for long-run risk models. Focusing on ν simplifies our simulation illustration and increases the transparency by allowing us to consider a few key (transformed) moment restrictions (i.e. a small yet essential subset of the moment restrictions used in constructing Table III):

$$m_t(\theta) = \Omega(\theta)^{-1/2} \begin{bmatrix} (\tilde{\sigma}_t^2 - \overline{\sigma}^2)\epsilon_{\sigma,t+1} \\ r_{t+1}^e - \mu_{r,t}^e - \beta_c \sigma_t \epsilon_{c,t+1} - \beta_x \sigma_t \epsilon_{x,t+1} - \beta_\sigma \sigma_w \epsilon_{\sigma,t+1} \end{bmatrix} \text{ and } \theta = \nu, \tag{222}$$

where variables $\epsilon_{c,t+1}$, $\epsilon_{x,t+1}$, and $\epsilon_{\sigma,t+1}$ are the residuals in (217a) – (217d) depending on observed data and unknown parameters in θ, and $\mu^e_{r,t}$ is defined in (220) and also dependent on observed data and unknown parameters in θ. Here $\Omega(\theta)$ is the asymptotic covariance matrix of the untransformed moments, and it is a diagonal matrix $\Omega(\theta) = \text{diag}\{\sigma_w^2/(1-\nu^2), \varphi_{d,d}^2\overline{\sigma}^2\}$. In equation (222), the first matrix element is the baseline moment, and the second is the asset pricing moment. Clearly, the nuisance parameter vector $\theta^{(2)}$ is empty in this simulation example.

We assume that the true local data-generating process has a time-varying relation between the expected log excess return and the dynamic parameters:

$$(223) \qquad r^e_{t,n} = r^e_t + \frac{\iota_t \delta_r}{\sqrt{n}}, \text{ with } \iota_t = \begin{cases} 1, & \text{when } 1 \le t \le \lfloor \pi n \rfloor \\ -1, & \text{when } \lfloor \pi n \rfloor < t \le n, \end{cases}$$

where the time series ι_t captures the structural breaks. The corresponding moment biases, evaluated at θ_0, are

$$(224) \qquad \mathrm{E}^{\mathrm{Q}_0}\left[m_t(\theta_0)\right] = \begin{bmatrix} 0 \\ \frac{\lambda_t^{(2)}}{\sqrt{n}} \end{bmatrix} \text{ with } \lambda_t^{(2)} \equiv \frac{\iota_t \delta_r}{\varphi_{d,d}\overline{\sigma}}.$$

Therefore, under the data-generating processes M1 and M2 in Table I, the moments have identical local biases $\lambda_t^{(2)}$ after substituting the calibrated parameter values into (224). This guarantees that the comparisons across models in Panels A and B of Figure 6 are valid.

Figure 6 shows three different simulation experiments. Panel A displays the local power functions of C tests. The solid and dotted curves reflect the test powers when the data-generating processes are characterized by calibrations M1 and M2 in Table I, respectively. In this experiment, we vary the local misspecification δ_r in the risk premium. The data-generating process under calibration M1 features an excessively large amount of dark matter according to Table I, and thus it has low internal refutability (i.e. little test power) consistent with Theorem 1.

Panel B of Figure 6 displays the histograms of logged overfitting measures $\log \mathcal{O}(\hat{\theta}^{(1)}_{\mathrm{e,n}}, \mathbf{y}^n)$ of efficient GMM estimators for two data-generating processes under calibrations M1 and M2 in Table I. In this experiment, we specify a structural break in the risk premium in the middle of the time-series sample with $\delta_r = 0.02$. Panel B shows that the calibrated structural model with too much dark matter (model M1) is likely to have more severe overfitting concerns for the efficient GMM estimator, which is consistent with Theorem 2.

Panel C of Figure 6 compares the expected out-of-sample fits between recursive GMM estimators $\tilde{\theta}_{\mathrm{e,n}}$ and efficient GMM estimators $\hat{\theta}_{\mathrm{e,n}}$. The two types of estimators are defined in Section 5.2. Consistent with the conventional intuition, efficient GMM estimators outperform their recursive counterparts in terms of the expected out-of-sample fit under the data-generating process M2. This is because the additional identification information is reliable and meaningful when the amount of dark matter is not excessively large. On the contrary, recursive GMM estimators outperform their efficient counterparts in terms of the expected out-of-sample fit under the data-generating process M1 with too much dark matter. This means that the concern of misspecification and instability entirely offsets– and even reverses – the efficiency gain from the additional moment restrictions. Again, this experiment suggests that the econometrician should back off from efficiency to gain more robustness

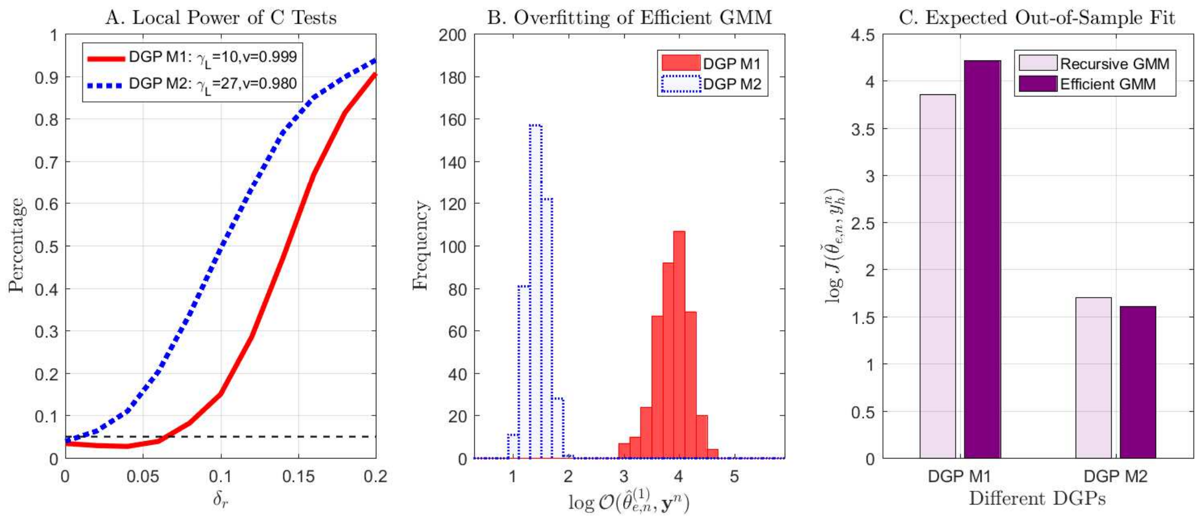

FIGURE 6.— Monte Carlo experiments for long-run risk models. In Panel A, we simulate 1000 independent monthly time series with length $n = 1200$ (i.e. 100 years). In Panels B and C, we simulate 400 independent monthly time series with length $n = 1200$ (i.e. 100 years) and break point $\pi = 1/2$. We set $\delta_r = 0.02$ for Panels B and C. In the simulation experiment, we assume that all the parameters except ν are treated as auxiliary parameters fixed at known constant values, subsumed into the functional form of the moment function (i.e. model specifications).

for the estimation results when the model contains a large amount of dark matter.