Measuring “Dark Matter” in Asset Pricing Models

Hui Chen, Winston Wei Dou, and Leonid Kogan*

December 23, 2021

Abstract

We formalize the concept of “dark matter” in asset pricing models by quantifying the additional informativeness of cross-equation restrictions about fundamental dynamics. The dark matter measure captures the degree of fragility for models that are potentially misspecified and unstable: a large dark matter measure signifies that the model lacks internal refutability (weak power of optimal specification tests) and external validity (high overfitting tendency and poor out-of-sample fit). The measure can be computed at low cost even for complex dynamic structural models. To illustrate its applications, we provide quantitative examples applying the measure to (time-varying) rare-disaster risk and long-run risk models.

Keywords: Fragile beliefs, Unstable models, Misspecification and robustness, Out-of-sample fit, Information bounds. (JEL Codes: C52, G12, D81, E32)

*Chen is at MIT (Sloan) and NBER. Dou is at University of Pennsylvania (Wharton). Kogan is at MIT (Sloan) and NBER. Emails: huichen@mit.edu, wdou@wharton.upenn.edu, lkogan@mit.edu. We are grateful to Stefan Nagel (the Editor), an Associate Editor, and two referees for their helpful suggestions. We thank Xiaohong Chen, Frank Schorfheide, Martin Schneider, and especially Lars Hansen for detailed comments and discussions. We also thank Torben Andersen, Bradyn Breon-Drish, Murray Carlson, John Cochrane, Ian Dew-Becker, Frank Diebold, Stefano Giglio, Robert Korajczyk, Anna Mikusheva, Tom Sargent, Laura Veldkamp, Jessica Wachter, Tan Wang, Amir Yaron, Fernando Zapatero, Stan Zin, and seminar and conference participants at University of Chicago, MIT Sloan, CKGSB, INSEAD, Northwestern, Harvard, NYU Stern, Yale, Washington University in St. Louis, University of Pennsylvania, Chicago CITE, ITAM Finance Conference, Macrofinance Society Workshop, Macro Financial Modeling Meeting, NBER Asset Pricing Meeting, NBER SI Meeting on Capital Markets and Economy, Red Rock Finance Conference, UBC Winter Finance Conference, Utah Winter Finance Conference, and WFA for their comments. We thank Di Tian for his superb research assistance. All errors are our own.

1. Introduction

In cosmology, dark matter is a form of substance 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." For instance, to generate desirable predictions, many asset pricing models require subtle dynamics for the fundamentals (the process of endowment or productivity, for example), which are difficult to identify directly using the fundamental data alone. Instead, the evidence can only be indirectly inferred from asset prices through the lens of model-implied restrictions. This phenomenon is present, for instance, in some of the models of rare disasters, to which John Campbell refers as the "dark matter for economists" in his 2008 Princeton Lecture in Finance (Campbell, 2018).

Intuitively, a model's heavy reliance on such "dark matter" features raises at least two concerns regarding its robustness. First, it is difficult to detect potential misspecification in the dark matter elements of a model due to our inability to measure them or test against them using direct evidence. As such, economic dark matter effectively raises a model's degrees of freedom. Second, as we estimate structural parameters from the data, the high effective degrees of freedom are likely to cause the model to overfit the data in sample, and lead to poor expected fit out of sample. We refer to these two problems as a lack of internal refutability and external validity, respectively.

Our paper makes two contributions. We propose a quantitative measure of the economic dark matter in structural models, which is relatively easy to compute even for models with complex dynamics. Then, under a general semiparametric framework, we formally show that models with larger dark matter measures tend to have lower in-sample refutability and less reliable out-of-sample performance owing to their higher tendency to overfit the data.

We define our measure of economic dark matter using a general GMM framework (see Hansen, 1982), under which a structural model is summarized by a fixed set of unconditional moment restrictions. We focus on a set of model parameters that can be identified by the full set of moment restrictions, referred to as the full model, and by a subset of the moment restrictions on the fundamentals, referred to as the baseline model. These parameters can be more precisely estimated through the additional moment restrictions implied by the structural model. These additional moment restrictions are often referred to as cross-equation restrictions. Our dark matter measure quantifies the additional informativeness of cross-equation restrictions about these model parameters. Specifically, it compares the asymptotic variances of two efficient GMM estimators based on the baseline model and the full model, which imposes cross-equation restrictions, and searches for the largest discrepancy between the two asymptotic variances along all directions. For a model that relies heavily on dark matter, cross-equation restrictions appear highly informative about the parameters of the fundamental dynamics relative to the fundamental data alone. It is important to note that, despite its close connection to asymptotic variances of efficient GMM estimators, economic dark matter is a model property and not a small-sample problem that can be ignored if one has sufficiently many observations.[1]

We formally connect our measure of economic dark matter to model fragility. Our notion of model fragility is concerned with testing power and expected out-of-sample fit in the presence of potential misspecification and local instability of the data-generating process (DGP). Intuitively, a structural model that can fit the data well in sample because of its high effective degrees of freedom, with some key model parameters insufficiently disciplined by direct evidence, is difficult to refute in statistical tests.[2] If a model is correctly specified, imposing highly informative cross-equation restrictions can significantly improve the precision of model parameter estimates, and thus the accuracy of the model's fit in and out of sample. However, if cross-equation restrictions are potentially misspecified, excessive reliance on the information derived from such restrictions tends to degrade the out-of-sample fit of the model despite its apparently strong in-sample performance.

We analyze the consequences of misspecification and local instability of the DGP by generalizing the framework of Li and Müller (2009) to the semiparametric setting.[3] We show that models with large dark matter measures are difficult to reject even when they are misspecified – thus, such models lack refutability. In fact, we prove that the power of the optimal specification tests vanishes as the dark matter measure approaches infinity. Moreover, under the worst cases of local instability of the DGP, models with larger dark matter measures have a higher expected degree of overfitting, i.e., a larger gap between the expected in-sample and out-of-sample model fit based on the Sargan-Hansen J statistic. This model property is not captured fully by the traditional measures of overfitting that depend on the number of free model parameters (e.g., AIC, BIC, among many others). In contrast, our measure focuses on the sensitivity of the model fit to perturbations in the parameters of the DGP.

In practice, sensitivity analysis is a popular method to examine model fragility. A model is considered fragile if its key implications are excessively sensitive to small perturbations of the DGP. However, to formalize this notion of model fragility, one needs to specify the relevant magnitude of "small perturbations" and define what constitutes "excessive sensitivity." To assess the full scope of model fragility in multivariate settings, one must further consider perturbations in the DGP that simultaneously affect multiple model parameters.

[1] Cheng, Dou, and Liao (2021) show that dark matter in asset pricing models is inherently connected but not limited to weak identification of model parameters. Similarly, as explicitly stressed by Stock, Wright, and Yogo (2002), weak identification as a model property is not a small-sample problem either.

[2] Kocherlakota (2007) similarly cautions about the "fallacy of fit" in structural economic models.

[3] We need a general semiparametric framework for three main reasons: (i) it provides a formal general description of local perturbations in the space of DGPs; (ii) it allows us to justify the information-matrix interpretation of our dark matter measure using semiparametric efficiency bounds; and (iii) it clarifies the relation between a structural economic model and its representation through moment restrictions.

Our dark matter measure can be applied to formalize the sensitivity analysis. It does so by (i) benchmarking the local perturbation in the DGP against the uncertainty of model parameters derived from the baseline model, and (ii) defining excessive sensitivity of the model to the local perturbations in the DGP based on the sampling variability of the efficient GMM estimators of the model parameters. Naturally, we require the baseline model to be a correct benchmark similar to the setup of Eichenbaum, Hansen, and Singleton (1988) because we need the identification provided by the baseline model to define reasonable perturbations in the DGP. In addition, for the multivariate setting, our measure identifies the worst-case direction of the perturbations as the direction in the model parameter space along which the cross-equation restrictions are most sensitive to the perturbation.

As another application, our dark matter measure can help reveal when the rational expectations assumption becomes tenuous in a model. A key assumption of rational expectations econometrics is that the agents in an economic model know more about model parameters than conveyed by the primitive data. As a large dark matter measure implies that the primary source of information about model parameters is the cross-equation restrictions rather than the primitive data, it may be increasingly difficult to argue that, as an approximation, economic agents have learned about the parameters from rich histories of primitive data. Moreover, information derived from cross-equation restrictions is a unreliable in the presence of potential model misspecification or instability. Taken together, a large dark matter measure suggests that the rational-expectations economist effectively circumvents the statistical challenges of validating the model and estimating its parameters by postulating fragile beliefs onto the economic agents inside the model.

As examples, we use our measure to evaluate the fragility of three prominent 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 fundamental data unless asset pricing information is used. We derive the dark matter measure in this example analytically, which helps convey the main intuition behind the measure. The other examples are a time-varying disaster risk model and a long-run risk model. Each model has at least nine key parameters. We use these examples to show that different calibrations of the same model with similar quality of in-sample fit can differ vastly in terms of their refutability and overfitting tendency.

We conduct Monte Carlo simulation experiments for all three examples and show that the calibrated models with large dark matter measures lack refutability, tend to overfit the data in sample, and have poor out-of-sample fit, consistent with our theoretical results. The simulation studies also show that robust estimation methods, such as the recursive GMM estimation procedure (e.g., Hansen, 2007b, 2012), are less susceptible to overfitting when a model contains a large amount of dark matter. By definition, the recursive GMM procedure first estimates the baseline parameters only using the baseline moments, then estimates the nuisance parameters using the asset pricing cross-equation restrictions with the baseline parameter values set equal to their estimates from the baseline moments. Although the recursive GMM estimator has worse in-sample fit compared to the efficient GMM estimator, it can deliver better out-of-sample performance when the dark matter measure is excessively large.

Our numerical examples, drawn from three influential asset pricing models, illustrate the properties of the dark matter measure, and its negative association with internal refutability of a model and its external validity. There are several ways in which one could use the dark matter measure to facilitate model construction and estimation. For example, different calibrations within the same model class (i.e., the same parametric functional form) can have drastically different levels of dark matter. Thus, our dark matter measure can be used to compare and select robust calibrations within a general model class. As another application, the recursive GMM procedure above offers a robust approach to parameter estimation in fragile models, as identified by our dark matter measure. Our dark matter measure can be used to detect whether a robust estimation procedure should be used at the cost of potential estimation efficiency. Finally, even for the models that may be recognized as potentially prone to fragility, our measure offers additional insight into which model components (and the associated model parameters) are most responsible for the fragility. In particular, the time-varying disaster risk model shares important features with both the constant disaster risk model and the long-run risk model. Our measure helps reveal that, relative to disaster size or average disaster frequency, the persistence and ultimately the long-run variance of conditional disaster probability is a more important source of model fragility. Such diagnostic information helps guide model development decisions and related data collection strategies to reduce fragility of the models that exhibit high levels of dark matter.

Related Literature. The idea that a model's fragility is connected to its degrees of freedom (i.e., model complexity) dates back at least to Fisher (1922). Traditionally, degrees of freedom of a model are measured by the number of parameters 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 a model's degrees of freedom have been well documented. New methods have been developed in the statistics literature to measure the sensitivity of model implications to parameter perturbations, as is in our notion of model fragility.[5] A common feature of these proposals is that they rely on

[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 has covered several alternative approaches to measuring "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).

the very model under evaluation to determine the degree of parameter perturbations; this is potentially problematic when the model under evaluation is itself fragile and possibly severely misspecified.

Our dark matter measure is different from the extant model fragility measures in three aspects. First, we allow for general local perturbations of potentially misspecified DGPs using a semiparametric framework, similar to Hansen and Sargent (2001), and not just for local parameter perturbations for which the model's functional form is correctly specified. Second, the reasonable local perturbations of DGPs are generated by the baseline model, which serves as a benchmark to assess the fragility of the structural model. Third, we connect our dark matter measure to a model's internal refutability and out-of-sample fit. Out-of-sample fit has been discussed by Hansen and Heckman (1996) for calibration and estimation analysis, and more recently, emphasized by Schorfheide and Wolpin (2012) and Athey and Imbens (2017, 2019) as an important criterion for assessing economic models.

It is important to note that the dark matter measure alone doesn't constitute a full-fledged tool for model selection. The measure focuses on the informativeness of the cross-equation restrictions about model parameters that are identified by both the baseline model and the full model, which are in contrast to the "nuisance parameters" that are only involved in the cross-equation restrictions. It is thus possible that, by adding more nuisance parameters to the structural model (and making the model more complex), one reduces the informativeness of the cross-equation restrictions and in turn lowers the dark matter measure. This shows that it is not appropriate to compare the dark matter measures across models with different functional forms. Instead, it should be used for comparing different calibrations within the same parametric framework.

In the structural estimation literature, testing a model against "untargeted moments" is a common yet ad hoc approach for judging a model's external validity.[6] Our dark matter measure provides a complementary approach for assessing a model's external validity, whose econometric properties are rigorously justified. It compares a model's out-of-sample and in-sample fit, but the notion of "out-of-sample" is quite different from that in the approach of untargeted moments. Instead of requiring new moments for testing, our dark matter measure is concerned with the expected out-of-sample fit for the same set of moments in new data, which can be evaluated using the split sample approach with the estimation sample and holdout sample, a standard method in financial and macroeconomic time-series analysis.

Another related literature focuses on the low refutability issue for linear asset pricing models.

[6] In the structural estimation literature, there is a long tradition to use one set of moments to estimate a model and then use another set of untargeted moments to test the model's out-of-sample fit. Recent examples that emphasize untargeted moments include Li, Taylor, and Wang (2018) and Dou, Taylor, Wang, and Wang (2020). Two challenges with this approach are: (i) all models will eventually be rejected after including sufficiently many untargeted moments; (ii) there is no clear guidance on the proper choice of untargeted moments.

Kan and Zhang (1999), Lewellen, Nagel, and Shanken (2010), and Gospodinov, Kan, and Robotti (2017), among others, analyze why standard statistical tests may have little power to reject misspecified pricing models. In the analysis of Lewellen, Nagel, and Shanken (2010), low power stems from the low effective dimensionality of the set of test assets compared to the number of risk factors in the stochastic discount factor (SDF). In Kan and Zhang (1999) and Gospodinov, Kan, and Robotti (2017), low power is caused by the weak correlation between the risk factors and the test assets.

Relative to this literature, we consider a broader set of nonlinear structural models and uncover a new source of low refutability, which has to do with an imbalance in information content between baseline moment restrictions and cross-equation restrictions imposed by the asset pricing model. Our focus is on model specification, with a large dark matter measure indicating that the power of the optimal specification test is low – in contrast to previous work, which has been concerned with point estimates and tests of the asset pricing models. Furthermore, we show that a large dark matter measure implies that the standard efficient inference procedure tends to overfit the model in sample with a poor out-of-sample fit. Follow-up work on linear asset pricing models (e.g., Kleibergen, 2009; Gospodinov, Kan, and Robotti, 2014; Kleibergen and Zhan, 2019; Giglio and Xiu, 2021) has proposed statistical inference procedures robust to identification failure and potential model misspecification. Development of new econometric procedures to deal with the dark matter feature of nonlinear structural economic models is a promising direction for future research (e.g., Cheng, Dou, and Liao, 2021).

2. A Motivating Example

In this section, we use a simple Gordon growth model to illustrate the main idea behind the concept of economic dark matter and motivate our dark matter measure. As we will show, higher value of the dark matter measure indicates diminishing ability to detect model misspecification.

2.1. *Model Setup*

Asset Pricing Model. Suppose the dividend Y_t for a stock evolves as follows:

$$\frac{Y_t}{Y_{t-1}} = 1 + \theta + \sigma_Y \varepsilon_{Y,t}, \quad \text{where } \varepsilon_{Y,t} \text{ is i.i.d., with } \mathrm{E}[\varepsilon_{Y,t}] = 0,\ \mathrm{E}[\varepsilon_{Y,t}^2] = 1. \tag{1}$$

The parameters $\theta \geq 0$ and $\sigma_Y > 0$ are the mean and volatility of dividend's net growth rates, respectively. According to the Gordon growth model, the price of the stock is the discounted value of expected future dividends. Assuming the risk-adjusted discount rate is r, then the stock price is

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

which implies a constant price-dividend ratio:

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

where $\theta \in \Theta \equiv [0, r)$ to ensure the existence of the equilibrium.

The econometrician evaluates this model in a sample of size n under a GMM framework,[7] in which a set of pre-specified moment restrictions summarize the implications of the model. To avoid stochastic singularity, we add i.i.d. shocks to the price-dividend ratio such that (3) only holds on average,

$$\frac{P_t}{Y_t} = F(\theta) + \sigma_P \varepsilon_{P,t}, \quad \text{where } \varepsilon_{P,t} \text{ is i.i.d., with } \mathrm{E}[\varepsilon_{P,t}] = 0,\, \mathrm{E}[\varepsilon_{P,t}^2] = 1. \tag{4}$$

The parameter $\sigma_P > 0$ is the standard deviation of price-dividend ratios. The shocks $\varepsilon_{P,t}$ could be due to measurement errors or noise trading, and we assume that $\varepsilon_{P,t}$ and $\varepsilon_{Y,t}$ are mutually independent. For simplicity, we assume that the econometrician knows the values of all the parameters except for the average dividend growth rate θ and focuses on the following pre-specified moment restrictions:

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

where $\Sigma \equiv \mathrm{diag}\{\sigma_Y, \sigma_P\}$ and $\mathbf{y}_t \equiv (Y_t, P_t)^T$. We refer to the first element of $m(\mathbf{y}_{t-1}, \mathbf{y}_t, \theta)$, denoted by $m^{(1)}(\mathbf{y}_{t-1}, \mathbf{y}_t, \theta)$, as the baseline moment. The corresponding moment restriction on dividend growth summarizes the baseline model. The full model then adds an additional restriction on the average price-dividend ratio. We refer to this additional moment, denoted by $m^{(2)}(\mathbf{y}_{t-1}, \mathbf{y}_t, \theta)$, as the asset pricing moment.

Underlying DGP and Model Misspecification. Our main objective in this example is to analyze the extent to which misspecifications in the asset pricing model can be successfully detected. For this purpose, we assume that the true DGP for the price-dividend ratio potentially differs from the asset pricing model in (4) while the dynamics of the dividend growth in (1) is

[7] Although we set up this simple example in a small sample, we want to stress the fact that the economic dark matter is a model property and not a small-sample problem.

correctly specified. More precisely, we assume that there exists $\theta_0 \in \Theta$ such that

$$\begin{bmatrix} Y_t/Y_{t-1} \\ P_t/Y_t \end{bmatrix} = \frac{1}{\sqrt{n}} \begin{bmatrix} 0 \\ \lambda \end{bmatrix} + \begin{bmatrix} 1+\theta_0 \\ F(\theta_0) \end{bmatrix} + \Sigma \begin{bmatrix} \varepsilon_{Y,t} \\ \varepsilon_{P,t} \end{bmatrix}. \tag{6}$$

Under the true DGP, the asset pricing moment restriction in (5) is locally misspecified when $\lambda \neq 0$. The correct restriction should be

$$\mathrm{E}\left[m^{(2)}(\mathbf{y}_{t-1}, \mathbf{y}_t, \theta_0)\right] = \frac{\lambda}{\sigma_P \sqrt{n}}. \tag{7}$$

Misspecification $\lambda/(\sigma_P\sqrt{n})$ shrinks with sample size n, and thus the difficulty in detecting the misspecification is not disappearing asymptotically as $n \to \infty$. This approach to modelling misspecification, using local asymptotics, has been applied successfully to approximate the finite-sample behavior of financial time series when testing and estimating linear asset pricing models under an asymptotic framework (e.g., Campbell and Yogo, 2006; Adrian, Crump, and Moench, 2015; Giglio and Xiu, 2021),[8] and is a natural approach for us to study model fragility because the dark matter problem cannot be ignored even with many observations.

2.2. *Testing for Misspecification*

The econometrician is interested in testing the validity of the restriction on the price-dividend ratio (i.e., whether $\lambda = 0$ in (6)).

Consider the set of alternatives $\mathcal{A}_\kappa(\mathrm{Q}_0) \equiv \{\lambda \in \mathbb{R} : |\lambda| \geq \kappa\sigma_P\}$, where the constant $\kappa > 0$ characterizes the minimum level of misspecification. According to Eichenbaum, Hansen, and Singleton (1988), the C test (or incremental J test) statistic is

$$C_n \equiv \min_{\theta\in\Theta} n\left|\widehat{m}_n(\theta)\right|^2 - \min_{\theta\in\Theta} n\left|\widehat{m}_n^{(1)}(\theta)\right|^2, \tag{8}$$

where $\widehat{m}_n^{(1)}(\theta)$ and $\widehat{m}_n(\theta)$ are the sample means for $m^{(1)}(\mathbf{y}_{t-1}, \mathbf{y}_t, \theta)$ and $m(\mathbf{y}_{t-1}, \mathbf{y}_t, \theta)$, respectively. The test statistic C_n will have large value when the model struggles to fit the additional moment restrictions in the full model.[9] Newey (1985a) shows that the C test is asymptotically optimal among GMM specification tests.

One can show that the C test statistic has a noncentral chi-square asymptotic distribution

[8] This approach is standard in the statistics and econometrics (e.g., van der Vaart, 1998; Lehmann, 1999).

[9] In this example, the C test is equivalent to the J test (Hansen, 1982) because $\min_{\theta\in\Theta} n\left|\widehat{m}_n^{(1)}(\theta)\right|^2 = 0$.

with one degree of freedom:

$$\underset{n\to\infty}{\text{wlim}}\, C_n = \left[Z + \frac{\lambda}{\sigma_P\sqrt{1+\varrho(\theta_0)}} \right]^2, \tag{9}$$

where $\text{wlim}_{n\to\infty}$ represents the weak convergence limit, Z is a standard normal variable, and

$$\varrho(\theta_0) \equiv \mathbf{I}_{\text{F}}/\mathbf{I}_{\text{B}} - 1, \text{ with } \mathbf{I}_{\text{B}} \equiv D_{11}^T D_{11} \text{ and } \mathbf{I}_{\text{F}} \equiv D^T D. \tag{10}$$

Here, $D_{11} \equiv \text{E}\left[\partial m^{(1)}(\mathbf{y}_{t-1}, \mathbf{y}_t, \theta_0)/\partial\theta\right]$ and $D \equiv \text{E}\left[\partial m(\mathbf{y}_{t-1}, \mathbf{y}_t, \theta_0)/\partial\theta\right]$ are the respective Jacobian matrices, evaluated at θ_0, for the baseline and full moments restrictions.

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

$$\inf_{\lambda\in\mathcal{A}_\kappa(\mathrm{Q}_0)} \lim_{n\to\infty} \mathbb{P}\left\{C_n > c_{1-\alpha}\right\} = \inf_{\lambda\in\mathcal{A}_\kappa(\mathrm{Q}_0)} \mathbb{P}\left\{ \left[Z + \frac{\lambda}{\sigma_P\sqrt{1+\varrho(\theta_0)}} \right]^2 > c_{1-\alpha} \right\} \tag{11}$$

$$\leq \mathbb{P}\left\{ \left[Z + \frac{\kappa}{\sqrt{1+\varrho(\theta_0)}} \right]^2 > c_{1-\alpha} \right\}, \tag{12}$$

where $c_{1-\alpha}$ is the $1-\alpha$ quantile of a chi-square distribution with 1 degree of freedom. The right-hand side of (12) is an upper bound on the maximin asymptotic power of the GMM specification tests, achieved by choosing $\lambda = \kappa\sigma_P$.

Condition (12) shows that our ability to detect model misspecification crucially depends on the value of $\varrho(\theta_0)$, which we refer to as the dark matter measure. As $\varrho(\theta_0)$ increases, the noncentrality parameter $\kappa^2/[1+\varrho(\theta_0)]$ approaches 0, and thus the upper bound $\mathbb{P}\left\{\left[Z+\kappa/\sqrt{1+\varrho(\theta_0)}\right]^2 > c_{1-\alpha}\right\}$ approaches α, meaning that the power of the specification test vanishes.

To get a sense of the magnitude of the effects, we calibrate the simple Gordon growth model by setting $\theta_0 = 0$, $r = 3\%$, $\sigma_Y = 4\%$, and $\sigma_P = 5$. Under this calibration, the value of the dark matter measure is $\varrho(\theta_0) = 83.9$, which signifies a very limited power to detect the misspecification of the model for the price-dividend ratio. Even when the model-implied average price-dividend ratio is severely misspecified with $\kappa = 6$,[10] the power of the test is at most 0.089. Moreover, using the formula in (12), the upper bound on the test power only starts to approach 1 when the misspecification is as large as $\kappa = 40$. Intuitively, the baseline moment restrictions have limited ability to refute the asset pricing cross-equation restrictions implied by the structural model when the dark matter measure $\varrho(\theta_0)$ is large. This is because, by tuning the parameter

[10] Setting $\kappa = 6$ means that the worst-case misspecification is 6 standard deviations away from 0 according to the population standard deviation of average price-dividend ratios $n^{-1}\sum_{t=1}^n P_t/Y_t$, which is $\sigma_P/\sqrt{n}$.

value of θ inside the "acceptable region" imposed by the baseline model (i.e., within the 95% confidence interval of θ inferred from the baseline moment restrictions), the econometrician can fit the model-implied average price-dividend ratio over an immensely wide range, meaning that the model-implied cross-equation restriction can hardly be rejected by the data. Formally, we establish the relation between model refutability and its dark matter measure under a formal and general econometric framework in Section 5.

2.3. Information Interpretation for the Dark Matter Measure

Even though we motivate and derive the dark matter measure in the examination of the test power for model misspecification (i.e., model refutability), it has a natural and intuitive information interpretation because the information matrices $\mathbf{I}_{\text{B}}$ and $\mathbf{I}_{\text{F}}$ in (10) gauge the informativeness of the baseline and full moment restrictions, respectively, about the model parameter θ (e.g., Chamberlain, 1987, Theorem 2). The dark matter measure $\varrho(\theta_0)$ effectively captures the incremental informativeness of the asset pricing moment restrictions relative to the baseline moment restrictions. A large value for $\varrho(\theta_0)$ indicates a severe information imbalance between the baseline moment restrictions and the asset pricing cross-equation restrictions. In our motivating example, we can further trace the incremental informativeness of the asset pricing cross-equation restriction according to the following expression:

$$\varrho(\theta_0) \equiv \mathbf{I}_{\text{F}}/\mathbf{I}_{\text{B}} - 1 = \left[F'(\theta_0)\frac{\sigma_Y}{\sigma_P}\right]^2. \tag{13}$$

Intuitively, $\varrho(\theta_0)$ increases when the asset pricing cross-equation restriction is more sensitive to the parameter value θ and is more precisely measured. This incremental informativeness of the asset pricing cross-equation restriction is essentially the focus of our dark matter measure.

Furthermore, because the inverse information matrices, denoted by $\mathbf{I}_{\text{B}}^{-1}$ and $\mathbf{I}_{\text{F}}^{-1}$, capture the asymptotic variances of the respective efficient GMM estimators based on the baseline and full moment restrictions, the dark matter measure has a natural "relative-sample-size" interpretation in an asymptotic sense. Specifically, $\varrho(\theta_0)$ gives the relative sample size required for the efficient GMM estimator of the baseline model to match the asymptotic precision about the model parameter θ provided by that of the full model. Because asymptotic variance scales inversely with the sample size, the sample size required for the baseline model to match the asymptotic precision of the full model is $83.9\,(= \varrho(\theta_0))$ times larger as $n \to \infty$. Importantly, the measure does not depend on sample size n, which highlights the fact that economic dark matter is not a small-sample problem that can be ignored if one has many observations. Like weak identification, the amount of economic dark matter is a model property rather than a small-sample problem (e.g., Stock, Wright, and Yogo, 2002).

We generalize the definition of the dark matter measure to the multivariate setting in Section 3. There, we compare the baseline and full model information matrices in all directions. The direction in which the two information matrices differ the most is also the one in which the dark matter measure is defined.

In addition to the low refutability issue for models with large dark matter measures, we further justify the importance of information imbalances (i.e., the amount of dark matter) as a primitive model property by showing that models with large dark matter measures tend to have poor out-of-sample fit numerically in the quantitative examples of Section 4 and mathematically in the formal econometric analysis of Section 5.

3. A Formal Definition: Dark Matter

In this section we set up a general econometric framework and formally define the dark matter measure. The technical regularity assumptions for econometric results are stated and explained in detail in the Appendix.

3.1. Econometric Setup

Let $\mathcal{Y} = \mathbb{R}^{d_y}$, which denotes 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 and Structural Models. 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 $\mathbb{P}_n$.

Consider a stable structural model denoted by $\mathcal{Q}$, which aims to capture certain pre-determined 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).[12] We assume that a set of pre-specified moment restrictions on the data summarize the model's key implications and that the model's performance in a given data sample can be measured by the degree to which these pre-specified

[11]The set of Markov processes satisfying the Doeblin condition includes a broad class of time series commonly used in finance and macroeconomics. 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).

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

moment restrictions are violated (i.e., the fit of moment restrictions). Notice that we are not focusing on checking whether a particular model is rejected by the data or not; instead, as emphasized in Section 2, we are focusing on evaluating model fragility, and our notion of model fragility is based on the sensitivity of the pre-specified moment restrictions to local perturbations of the underlying DGP.[13] We follow the literature (e.g., Li and Müller, 2009; Chen and Santos, 2018) and use the GMM objective function to evaluate the structural models. As reflected in the original applications of GMM in asset pricing (see Hansen and Singleton, 1982, 1983) and recently emphasized by Hansen (2014), structural models are typically partially specified in the sense that ultimately the model is statistically rejected once enough moments are added. Instead, GMM has proven particularly valuable for analyzing structural models by focusing on key moment restrictions without being overly influenced by all the unimportant details and potential singularities of the remainder of the structural model.

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

$$\mathcal{Q} = \left\{\mathrm{Q} \in \mathcal{H} : \mathrm{E}^{\mathrm{Q}}\left[m(\cdot,\theta)\right] = 0 \text{ for some } \theta \in \Theta\right\}, \tag{14}$$

which is a collection of probability measures in $\mathcal{H}$ 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 fewer than that of the moment restrictions ($d_\theta < d_m$).[14]

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

$$m(\mathbf{y}_{t-1},\mathbf{y}_t,\theta) = \left[\begin{array}{c} m^{(1)}(\mathbf{y}_{t-1},\mathbf{y}_t,\theta^{(1)}) \\ m^{(2)}(\mathbf{y}_{t-1},\mathbf{y}_t,\theta^{(1)},\theta^{(2)}) \end{array}\right], \quad \text{with} \quad \theta = \left[\begin{array}{c} \theta^{(1)} \\ \theta^{(2)} \end{array}\right]. \tag{15}$$

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

$$m^{(1)}(\mathbf{y}_{t-1},\mathbf{y}_t,\theta^{(1)}) = \Gamma_{m,1} m(\mathbf{y}_{t-1},\mathbf{y}_t,\theta), \text{ with } \Gamma_{m,1} \equiv \left[I, 0_{d_{m,1}\times(d_m-d_{m,1})}\right]. \tag{16}$$

The assumption of the recursive structure for moment functions enables us to examine the fragility of a subset of moment restrictions, namely those in $m^{(2)}(\cdot,\theta)$. Such recursive structures

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

[14] Although dynamic stochastic equilibrium models often feature conditional moment restrictions, for estimation and testing, it is common to focus on a finite number of 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.

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 or output), while the additional moments $m^{(2)}(\cdot, \theta)$ may apply to the joint dynamics of the real quantities and asset returns.

Because the first coordinate block of the moment function $m(\cdot, \theta)$ only depends on $\theta^{(1)}$, the Jacobian matrix $D(\theta)$ is a block lower triangular matrix defined as follows:

$$D(\theta) = \begin{bmatrix} D_{11}(\theta) & 0 \\ D_{12}(\theta) & D_{22}(\theta) \end{bmatrix}, \text{ where } D_{ij}(\theta) \equiv \mathrm{E}^{\mathrm{Q}}\left[\nabla_{\theta^{(j)}} m^{(i)}(\cdot, \theta)\right] \text{ and } i, j = 1, 2. \tag{17}$$

Corresponding to the first coordinate block $m^{(1)}(\cdot, \theta^{(1)})$ of the moment function $m(\cdot, \theta)$ in (15), we define the baseline structural model $\mathcal{Q}^{(1)}$ as:

$$\mathcal{Q}^{(1)} = \left\{Q \in \mathcal{H} : \mathrm{E}^{\mathrm{Q}}\left[m^{(1)}(\cdot, \theta^{(1)})\right] = 0 \text{ for some } \theta^{(1)} \in \Theta^{(1)}\right\}, \tag{18}$$

where $\Theta^{(1)}$ is the baseline parameter set of $\theta^{(1)}$. Thus, the baseline structural model $\mathcal{Q}^{(1)}$ is a collection of probability measures under which the first block of moment restrictions, hereafter referred to as the baseline moment restrictions, hold for some parameter vector $\theta^{(1)}$. This definition is analogous to the definition of the full model, and clearly $\mathcal{Q} \subset \mathcal{Q}^{(1)}$. According to (15), 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.

Following the definition of the full structural model in (14), we define a mapping from the probability measure of the bivariate marginal distribution $\mathrm{Q} \in \mathcal{Q}$ to model parameters θ, denoted by $\theta = \vartheta(\mathrm{Q})$, such that

$$\mathrm{E}^{\mathrm{Q}}\left[m(\cdot, \vartheta(\mathrm{Q}))\right] = 0. \tag{19}$$

Similarly, for the baseline structural model in (18), we define a mapping from the probability measure of the bivariate marginal distribution $\mathrm{Q} \in \mathcal{Q}^{(1)}$ to model parameters θ, $\theta^{(1)} = \vartheta^{(1)}(\mathrm{Q})$, such that

$$\mathrm{E}^{\mathrm{Q}}\left[m^{(1)}(\cdot, \vartheta^{(1)}(\mathrm{Q}))\right] = 0. \tag{20}$$

Calibrated Models. Consider a calibrated model parameter value $\theta_0 \in int(\Theta)$, the interior of Θ. The calibrated full and baseline structural model are sets of probability measures satisfying

$$\mathcal{Q}(\theta_0) \equiv \left\{\mathrm{Q} \in \mathcal{H} : \mathrm{E}^{\mathrm{Q}}\left[m(\cdot, \theta_0)\right] = 0\right\}, \text{ and} \tag{21}$$

$$\mathcal{Q}^{(1)}(\theta_0^{(1)}) \equiv \left\{\mathrm{Q} \in \mathcal{H} : \mathrm{E}^{\mathrm{Q}}\left[m^{(1)}(\cdot, \theta_0^{(1)})\right] = 0\right\}. \tag{22}$$

By definition, $\mathcal{Q}(\theta_0) \subset \mathcal{Q}$. We assume that $\mathcal{Q}(\theta_0)$ is non-empty, and 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. Note that Q_0 remains unknown to the econometrician, even though θ_0 may be known.

J Statistics. Under distribution Q_0, we denote the Jacobian matrices evaluated at the calibrated model parameter value θ_0 by $D_{11} \equiv D_{11}(\theta_0)$ and $D \equiv D(\theta_0)$ for the baseline and full model, respectively, and we denote the spectral density matrices (at zero frequency) for the baseline and full model by

$$\Omega_{11} \equiv \sum_{t=-\infty}^{\infty} \mathrm{E}^{\mathrm{Q}_0}\left[m^{(1)}(\mathbf{y}_0, \mathbf{y}_1, \theta_0) m^{(1)}(\mathbf{y}_{t-1}, \mathbf{y}_t, \theta_0)^T\right], \text{ and} \tag{23}$$

$$\Omega \equiv \sum_{t=-\infty}^{\infty} \mathrm{E}^{\mathrm{Q}_0}\left[m(\mathbf{y}_0, \mathbf{y}_1, \theta_0) m(\mathbf{y}_{t-1}, \mathbf{y}_t, \theta_0)^T\right], \text{ respectively,} \tag{24}$$

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.[15] Without loss of generality, we further assume that $\Omega = I$, which is innocuous because we can always rotate the system of moment restrictions without altering the structure of the model (Hansen, 2007b). More details are provided in Online Appendix 5.2.

For any given θ, we define

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

and we define the J statistics for the baseline and full models as

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

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

3.2. Information Matrices

We now introduce information matrices for the structural 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,

[15] For example, several consistent estimators 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.

Ritov, and Wellner, 1993).

One of our theoretical contributions is to formalize the information interpretation of dark matter. In the Online Appendix, we extend the semiparametric efficiency bounds for unconditional moment restrictions, established by Levit (1976), Nevelson (1977), and Chamberlain (1987, Theorem 2), from i.i.d. DGPs to Markov processes with local instability.[16] Particularly, in Online Appendix 3, we show that the optimal GMM covariance matrix derived by Hansen (1982) achieves the semiparametric minimax efficiency bound for unconditional moment restrictions with Markov DGPs that are locally unstable.

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

$$\mathbf{I}_{\mathrm{B}} = D_{11}^T D_{11} \quad \text{and} \quad \mathbf{I}_{\Omega} = D^T D = \begin{bmatrix} D_{11}^T D_{11} + D_{21}^T D_{21} & D_{21}^T D_{22} \\ D_{22}^T D_{21} & D_{22}^T D_{22} \end{bmatrix}, \tag{25}$$

where $D \equiv D(\theta_0)$ and $D_{ij} \equiv D_{ij}(\theta_0)$ for $i, j = 1, 2$. From $\mathbf{I}_{\Omega}$ we can also define the marginal information matrix for $\theta^{(1)}$ in the full model, evaluated at $\theta_0^{(1)}$, as

$$\mathbf{I}_{\mathrm{F}} = \left[\Gamma_{\theta,1} \mathbf{I}_{\Omega}^{-1} \Gamma_{\theta,1}^T\right]^{-1}, \quad \text{where } \Gamma_{\theta,1} \equiv \left[I, 0_{d_{\theta,1} \times (d_\theta - d_{\theta,1})}\right], \tag{26}$$

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

$$\mathbf{I}_{\mathrm{F}} = D_{11}^T D_{11} + D_{21}^T \Lambda_2 D_{21}, \quad \text{with } \Lambda_2 \equiv I - D_{22}(D_{22}^T D_{22})^{-1} D_{22}^T. \tag{27}$$

3.3. *Dark Matter Measure*

We construct the dark matter measure by quantifying the incremental informativeness of the additional moment restrictions in the full model (but not in the baseline model) regarding the model parameters $\theta^{(1)}$. Because $\theta^{(1)}$ appears in both the baseline moment restrictions and the additional moment restrictions in the full model, the cross-equation restrictions provide additional information about $\theta^{(1)}$ beyond the baseline model. As the illustrative example in Section 2.3 demonstrates, the incremental informativeness of the additional moment restrictions

[16] 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 and study its efficiency for conditional moment restrictions, which contain finite dimensional unknown parameters and infinite dimensional unknown functions.

naturally depends on the sensitivity of these moments to changes in model parameters. If a small change in the parameter values can dramatically change the value of the moments (i.e., high sensitivity), then imposing the additional moment restrictions empirically will tend to greatly restrict the parameter estimates; i.e., these moment restrictions will appear informative. As we demonstrate formally in Sections 4 and 5, in the presence of misspecification concerns about the asset pricing model, such information imposed by the additional moment restrictions implied by the structural model can be very problematic.

We now introduce our dark matter measure.

Definition 1 (Dark Matter Measure)**.** *Let the incremental information matrix of the full model relative to the baseline models be*

$$\Pi \equiv \mathbf{I}_{F}^{1/2}\mathbf{I}_{B}^{-1}\mathbf{I}_{F}^{1/2} - I. \tag{28}$$

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

$$\varrho(\theta_0) \equiv \max_{|\mathbf{v}|=1} \mathbf{v}^T \Pi \mathbf{v}. \tag{29}$$

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

$$\varrho(\theta_0) = \max_{|\mathbf{v}|=1} \frac{\mathbf{v}^T \mathbf{I}_{\mathrm{B}}^{-1} \mathbf{v}}{\mathbf{v}^T \mathbf{I}_{\mathrm{F}}^{-1} \mathbf{v}} - 1. \tag{30}$$

As (30) shows, our measure effectively compares the asymptotic covariance matrices of the two estimators of $\theta^{(1)}$, one based on the baseline model and the other based on the full model. It is the largest ratio of the two asymptotic variances of the efficient GMM estimator under the baseline and full model for a linear combination of model parameters $\mathbf{v}^T\theta^{(1)}$ over all possible directions $\mathbf{v} \in \mathbb{R}^{d_{\theta,1}}$. We focus on the one-dimensional worst-case fragility. There are straightforward extensions to the cases in which $\mathbf{v}$ is a matrix.

While the amount of dark matter, like weak identification, is a model property rather than a small-sample problem, the expression in (30) shows that the dark matter measure has a natural "relative-sample-size" interpretation in an asymptotic sense. This equation gives the minimum relative sample size required for the efficient GMM estimator of the baseline model to match the asymptotic precision about the baseline parameter $\theta^{(1)}$ provided by that of the full model in all directions. Because asymptotic variance scales inversely with the sample size, the sample size required for the baseline model to match the asymptotic precision of the full model is $[1 + \varrho(\theta_0)]$ times the sample size n for the full model as $n \to \infty$. In sum, the dark matter measure, $\varrho(\theta_0)$, should be interpreted in an asymptotic sense.

Our dark matter measure isolates the information provided by the full model above and beyond the baseline model. For the same full 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 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 structural model are targeted by the fragility analysis.

3.4. *Test Power and Overfitting Measure*

Before presenting the examples in Section 4, we first introduce the definitions of local power of specification tests and overfitting measure based on the econometric setup above.

Local Power of Specification Tests. A specification test for a structural model $\mathcal{Q}$ against its baseline 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 DGP) against the alternative that there only exists some parameter for which the baseline moment restrictions hold but none would satisfy all the moment restrictions of the full model; that is,

$$H_0 : \mathrm{Q}_n \in \mathcal{Q} \quad vs. \quad H_{\mathcal{A}} : \mathrm{Q}_n \in \mathcal{Q}^{(1)} \setminus \mathcal{Q}, \tag{31}$$

where Q_n is the bivariate marginal distribution of the true local DGP.

Let $\check{\varphi}_n$ be an arbitrary GMM test statistic that maps $\mathbf{y}^n$ to $[0, 1]$, as defined in Hansen (1982) and 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.[17] More precisely, we consider the local DGP $\mathbb{P}_n$ for $\mathbf{y}^n$ with a bivariate marginal distribution Q_n that converges to $\mathrm{Q}_0 \in \mathcal{Q}(\theta_0)$ as $n \to \infty$, such that

$$\mathrm{E}^{\mathrm{Q}_n}\left[m_t(\theta_0)\right] = \delta/\sqrt{n} + o\left(1/\sqrt{n}\right), \tag{32}$$

where $\delta \in \mathbb{R}^{d_m}$ is the local bias in moment restrictions under $\mathbb{P}_n$ when evaluated at θ_0. The test $\check{\varphi}_n$ has a local asymptotic power function $q(\delta, \check{\varphi})$ if

$$q(\delta, \check{\varphi}) \equiv \lim_{n\to\infty} \int \check{\varphi}_n \mathrm{d}\mathbb{P}_n, \quad \forall\ \delta \in \mathbb{R}^{d_m}, \tag{33}$$

where $\check{\varphi} \equiv \{\check{\varphi}_n\}_{n\geq 1}$ is the sequence of test statistics and δ is the local bias in moment restrictions associated with $\mathbb{P}_n$ in (32).

[17] As the sample size n approaches infinity, the distance between the null hypothesis and the DGP necessarily diminishes according to $n^{-1/2}$. If this distance were held fixed, then the power of all consistent tests would tend to 1 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; Campbell and Yogo, 2006).

Out-of-Sample Fit and Robust Estimation. A common method adopted by economists and statisticians for assessing the external validity of models is to hold out data from the model estimation. The assessment of external validity serves two important purposes: it mitigates the concern of in-sample overfitting,[18] and 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 is useful to account for model uncertainty, model instability, calibration uncertainty, and estimation uncertainty, in addition to the usual uncertainty of future events (see Stock and Watson, 2008).

The holdout approach 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 $\pi \in (0, 1/2]$.[19] 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).[20] This approach has been commonly adopted in the literature not only in the context of statistical forecasting and model selection, but also in the context of calibration selection (e.g., Hansen and Heckman, 1996, Pages 92–94). Further, the holdout approach is 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 (see Müller and Watson, 2016, Section 5.2).

The idea is to quantify the overfitting tendency as a model property by focusing on the J statistic as the loss function. Let $Q_{n,t}$ be the marginal bivariate distribution of $(\mathbf{y}_{t-1}, \mathbf{y}_t)$. We define $\theta^{(1)}_{n,t} \equiv \vartheta^{(1)}(Q_{n,t})$ for $t = 1, \cdots, n$, and

$$\theta^{(1)}_{\mathrm{e,n}} \equiv \frac{1}{\lfloor \pi n \rfloor} \sum_{t=1}^{\lfloor \pi n \rfloor} \theta^{(1)}_{n,t} \quad \text{and} \quad \theta^{(1)}_{\mathrm{o,n}} \equiv \frac{1}{\lfloor (1-\pi) n \rfloor} \sum_{t=\lfloor \pi n \rfloor + 1}^{n} \theta^{(1)}_{n,t}, \tag{34}$$

where the mapping $\vartheta^{(1)}$ is defined in (20).

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

$$\mathcal{L}(\theta^{(1)}; \mathbf{y}^n_s) \equiv J(\theta^{(1)}, \psi_s(\theta^{(1)}), \mathbf{y}^n_s) - J(\theta^{(1)}_{\mathrm{s,n}}, \psi_s(\theta^{(1)}_{\mathrm{s,n}}), \mathbf{y}^n_s), \text{ with } s \in \{\mathrm{e}, \mathrm{o}\}, \tag{35}$$

[18] For example, see Foster, Smith, and Whaley (1997); Lettau and Van Nieuwerburgh (2008); Welch and Goyal (2008), as well as recent work by Athey and Imbens (2017, 2019); Kozak, Nagel, and Santosh (2019), among others.

[19] 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.

[20] 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).

where $\theta^{(1)}_{\mathrm{s,n}}$ is the average of those correct baseline parameter values that perfectly fit baseline moment restrictions as defined in (34), and $\psi_s(\theta^{(1)})$ is chosen to minimize the J statistic while taking $\theta^{(1)}$ as given:[21]

$$\psi_s(\theta^{(1)}) \equiv \underset{\theta^{(2)}}{\operatorname{argmin}}\, J(\theta^{(1)}, \theta^{(2)}, \mathbf{y}^n_s) \quad \text{for any fixed } \theta^{(1)} \text{ with } s \in \{\mathrm{e}, \mathrm{o}\}. \tag{36}$$

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 isolate the effect of parameter deviation from the average of the correct baseline parameter values from the effect on instability. 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 $s \in \{\mathrm{e}, \mathrm{o}\}$. 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:

$$\mathcal{O}(\check{\theta}^{(1)}_{\mathrm{e,n}}, \mathbf{y}^n) \equiv \frac{1}{2}\left[\mathcal{L}(\check{\theta}^{(1)}_{\mathrm{e,n}}, \mathbf{y}^n_{\mathrm{o}}) - \mathcal{L}(\check{\theta}^{(1)}_{\mathrm{e,n}}, \mathbf{y}^n_{\mathrm{e}})\right]. \tag{37}$$

Recursive GMM Estimation Procedure. We focus on two particular estimation procedures — the efficient and recursive GMM estimation procedure. The former is designed to use the identification strength provided by the additional asset pricing moment restrictions $\mathrm{E}^{\mathrm{Q}_0}\left[m^{(2)}_t(\theta)\right] = 0$ as much as possible, while the latter does not use any identification assumptions imposed by the additional asset pricing moment restrictions $\mathrm{E}^{\mathrm{Q}_0}\left[m^{(2)}_t(\theta)\right] = 0$. 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, 2007b, 2012). While the original impetus of the recursive GMM estimation was primarily computational, we show that the recursive GMM procedure is more robust against potential instability and misspecification because it barely relies on the nontestable assumption of identification strength of the additional moment restrictions $m^{(2)}_t(\theta)$; the robustness of the recursive GMM estimation procedure is especially valuable when the dark matter measure is large.

[21] Mathematically, (35) and (36) follow the generic recursive GMM estimation procedure in Hansen (2007b) and Hansen (2012).

Characterized by selection matrices, the efficient GMM estimator and the recursive GMM estimator based on the estimation sample $\mathbf{y}_{\mathrm{e}}^{n}$, 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 $A_{22} \equiv \left[D_{21}(D_{11}^T D_{11})^{-1} D_{21}^T + I\right]^{-1} D_{22}$ (see Hansen, 2007b).

Summary of the Main Theoretical Results. We now summarize informally our main theoretical results on the relations between the dark matter measure, model refutability, and overfitting tendency, which we develop fully in Section 5 below. We discuss these theoretical findings further in the simulation studies of Section 4, where we apply our dark matter measure to several prominent asset pricing models.

Our first main theoretical result (formally stated in Theorem 1 in Section 5) connects the dark matter measure and the power of an optimal maximin specification test. Specifically, Theorem 1 shows that a structural model with a larger dark matter measure, as a model property, tends to have a lower local asymptotic power of an optimal maximin specification test. Theorem 1 also shows that, as the dark matter measure approaches infinity, the specification test has vanishing ability to detect misspecification. Our second result (formally stated in Theorem 2 in Section 5) shows that, when the structural model has a larger dark matter measure, as a model property, the efficient GMM estimator is expected to have a larger overfitting measure. Our third theoretical result (see Theorem 3 in Section 5) shows that, regardless of the magnitude of the dark matter measure, the overfitting of the recursive GMM estimator is constant, which is equal to the dimension of the baseline parameter vector.

4. Quantitative Examples

We now use the dark matter measure to analyze three of the leading consumption-based asset pricing models: a rare disaster model, a time-varying disaster risk model, and a long-run risk model. The primary goal of these examples is to illustrate the computation of the dark matter measure and show the connection among a model's dark matter measure, in-sample refutability, and out-of-sample performance. We formally establish the connections in Section 5 under a general semiparametric framework.

4.1. Dark Matter of Rare Disaster Risk Models

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 rare disasters can have large ex-ante effects on asset prices. In this subsection, we use our dark matter measure to analyze a disaster risk model similar to those of Rietz (1988) and Barro (2006).

The model specifies the joint dynamics of log growth rate of aggregate consumption g_t and 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 1 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 the disaster regime, $g_t = -v_t$, where v_t follows a truncated exponential distribution with density

$$v_t \overset{\text{i.i.d.}}{\sim} \mathbf{1}\{v_t > \underline{v}\}\xi e^{-\xi(v_t - \underline{v})}, \quad \text{with } \underline{v} > 0,\ \xi > 0. \tag{38}$$

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. In the normal regime ($z_t = 0$), g_t and r_t are jointly normal, and

$$r_t = \eta + \rho\frac{\tau}{\sigma}(g_t - \mu) + \sqrt{1-\rho^2}\tau\varepsilon_{0,t}, \tag{39}$$

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

$$r_t = \ell g_t + \varsigma\varepsilon_{1,t}, \tag{40}$$

where $\varepsilon_{1,t}$ is i.i.d. standard normal. The shocks z_t, v_t, $\varepsilon_{0,t}$, and $\varepsilon_{1,t}$ are mutually independent.

Next, we assume that the representative agent has a constant relative risk aversion utility function $u_t(c_t) = \delta^t c_t^{1-\gamma}/(1-\gamma)$, where $\gamma \in (0, +\infty)$ is the coefficient of relative risk aversion and $\delta \in (0, 1)$ is the time preference parameter. The log equity premium, $\overline{r}(p, \xi) \equiv \mathbb{E}[r_t]$, is available in closed form (see Online Appendix 6 for details) as follows:

$$\overline{r}(p,\xi) = (1-p)\eta - p\ell\left(\underline{v} + 1/\xi\right), \quad \text{where} \tag{41}$$

$$\eta = \gamma\rho\sigma\tau - \frac{\tau^2}{2} + e^{\gamma\mu - \frac{\gamma^2\sigma^2}{2}}\Delta(\xi)\frac{p}{1-p}, \text{ with } \Delta(\xi) = \xi\left[\frac{e^{\gamma\underline{v}}}{\xi-\gamma} - \frac{e^{\frac{\varsigma^2}{2} + (\gamma-\ell)\underline{v}}}{\xi+\ell-\gamma}\right]. \tag{42}$$

The term η in (41) is the log equity premium in the normal regime (see equation (39)). The first two terms of η in (42) describe the market risk premia due to Gaussian consumption shocks; the third term is the disaster risk premium, which explodes as ξ approaches γ from above. In other words, there is an upper bound on the average disaster size for the equity premium to remain finite, which also limits how heavy the tail of the disaster size distribution can be.

The fact that the equity premium explodes as ξ approaches γ is an important feature of our version of the disaster risk model. No matter how rare the disasters are (i.e., no matter how small p is), an arbitrarily large equity premium can be generated when the average disaster size is sufficiently large (i.e., ξ is sufficiently small). Extremely rare but large disasters are difficult to refute based on the observable data and standard statistical tests when model misspecification is a concern; moreover, asset pricing models that rely heavily on such extremely rare but large disasters tend to perform poorly out of sample when instability is a concern. Below, we illustrate how our dark matter measure can detect and quantify in-sample refutability and out-of-sample performance of these models.

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

$$m_t(\theta) = \Omega(\theta)^{-1/2} \begin{bmatrix} z_t - p \\ g_t - (1 - z_t)\mu + z_t(\underline{v} + 1/\xi) \\ r_t - (1 - z_t) \left[\eta + \rho\frac{\tau}{\sigma}(g_t - \mu)\right] - z_t \ell g_t \end{bmatrix}. \tag{43}$$

The first two moments in $m_t(\theta)$ are for the baseline model. The full model adds a third moment on the equity premium, and $\Omega(\theta)$ is the asymptotic covariance matrix of the untransformed moments. To ensure analytical tractability and highlight the key idea, we focus on the parameters $\theta = (p, \xi)^T$ when constructing the dark matter measure, while treating the parameters $(\gamma, \mu, \sigma, \underline{v}, \tau, \rho, \ell, \varsigma)$ as auxiliary parameters fixed at known values, making them a part of the functional-form specification. Thus, the uncertainty in $(\gamma, \mu, \sigma, \underline{v}, \tau, \rho, \ell, \varsigma)$ is not accounted for in the dark matter measure. The nuisance parameter vector $\theta^{(2)}$ is empty in this example.

Based on the relation in (42), the dark matter measure is (see Online Appendix 6):

$$\varrho(\theta) = 1 + \frac{p\Delta\left(\xi\right)^2 + p\left(1 - p\right)\xi^2\dot{\Delta}\left(\xi\right)^2}{\left(1 - \rho^2\right)\tau^2\left(1 - p\right)^2} e^{2\gamma\mu - \gamma^2\sigma^2}, \tag{44}$$

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

$$\dot{\Delta}(\xi) = -\frac{e^{\gamma\underline{v}}\gamma}{(\xi - \gamma)^2} + \frac{e^{(\gamma-\ell)\underline{v}}(\gamma - \ell)}{(\xi - \gamma + \ell)^2} e^{\varsigma^2/2}. \tag{45}$$

All else equal, when ξ approaches γ, both $\Delta(\xi)$ and $\dot{\Delta}(\xi)$ approach infinity, suggesting that disaster risk models featuring large but rare disasters (i.e., small ξ and small p) will have large dark matter measures.

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

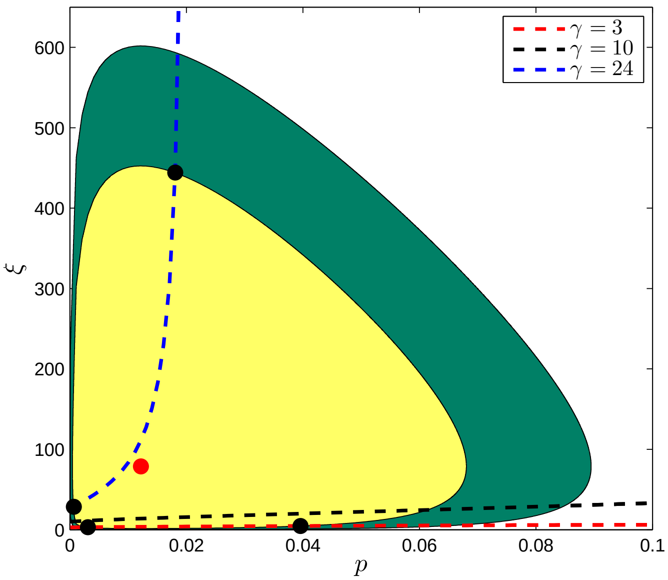

Figure 1: This figure shows the 95% and 99% confidence regions of (p, ξ) for the baseline model and the equity premium isoquants implied by the asset pricing moment restriction (41) for $\gamma = 3, 10, 24$. p is disaster probability, and ξ characterizes the inverse of average disaster size. The efficient 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 $\gamma = 3$ 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, as a part of the functional-form specification.

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 1, we plot the 95% and 99% confidence regions for (p, ξ) based on the baseline model. As expected, the confidence regions are large, as the baseline model provides very limited information about p and ξ. We also plot the equity premium isoquants: for a given level of risk aversion γ, each dashed line in Figure 1 shows the different combinations of p and ξ that match the unconditional average equity premium of 5.09%. Even for low risk aversion (e.g., $\gamma = 3$), there exist model calibrations 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.[22]

While it is difficult to distinguish among a wide range of calibrations based on the fit with the macro data, these calibrated models differ vastly based on the dark matter measure. For illustration, we focus on the following four calibrations, which are the four points where the

[22] 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.

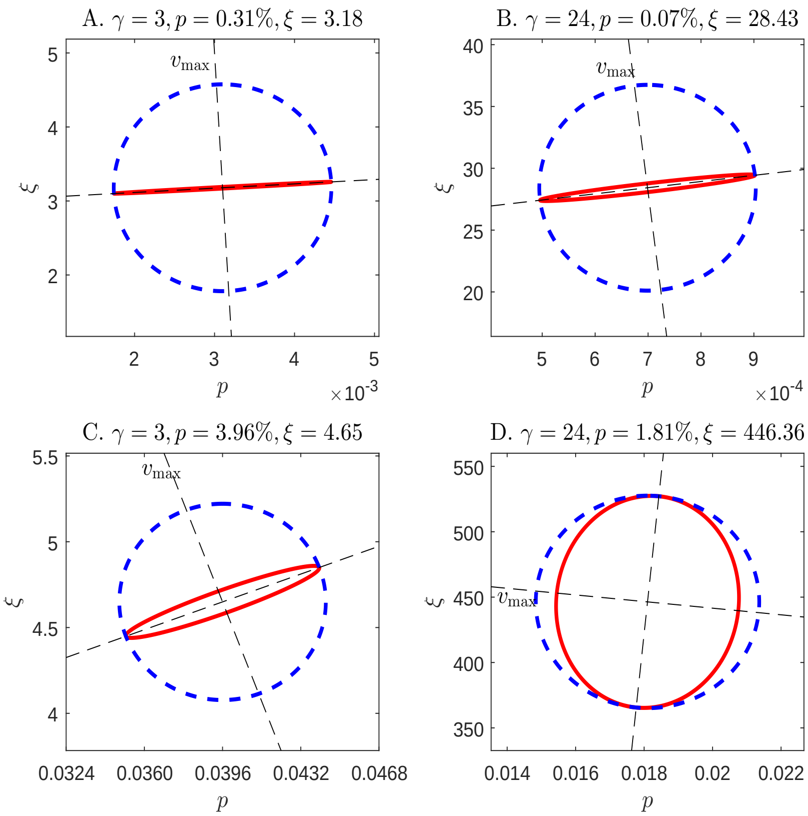

Figure 2: This figure visualizes the dark matter measure through the 95% confidence regions for the asymptotic distribution of the efficient GMM estimators for four "acceptable" calibrations. In panels A through D, the dark matter measures are $\varrho(\theta) = 1.78 \cdot 10^4$, $5.60 \cdot 10^2$, 74.03, and 1.49, 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; all other parameters are auxiliary parameters with fixed known values as a part of the functional-form specification. Therefore, the dark matter measure is defined only based on $\theta = (p, \xi)^T$.

equity premium isoquants for $\gamma = 3$ and 24 intersect the boundary of the 95% confidence region in Figure 1. For $\gamma = 3$, the two points are $(p = 3.96\%, \xi = 4.65)$ and $(p = 0.31\%, \xi = 3.179)$. For $\gamma = 24$, the two points are $(p = 1.81\%, \xi = 446.36)$ and $(p = 0.07\%, \xi = 28.43)$.

With just two parameters in $\theta = (p, \xi)^T$, we can visualize the dark matter measure by plotting the asymptotic confidence regions for (p, ξ) in the baseline and full model, as determined by the respective information matrices $\mathbf{I}_{\mathrm{B}}$ and $\mathbf{I}_{\mathrm{F}}$. In each panel of Figure 2, the larger dashed-line circle marks the 95% asymptotic confidence region for (p, ξ) under the baseline model, which only uses the first two macro moments in (43). The smaller solid-line ellipse indicates the 95% asymptotic confidence region for (p, ξ) under the full model, which adds the asset pricing moment. Intuitively, the direction in Figure 2 along which the asset pricing restriction does not provide additional information about the parameters $\theta = (p, \xi)^T$, is parallel to the tangent direction of the equity premium isoquants in Figure 1, evaluated at the black dots. This is due to the fact that the asset pricing moment restriction does not add more information in the direction along which the equity premium does not change.

Panels A and B of Figure 2 correspond to the calibrations with "extremely rare and large disasters." The dark matter measure $\varrho(\theta)$ is $1.78 \cdot 10^4$ and $5.60 \cdot 10^2$ for $\gamma = 3$ and 24, respectively. In panel C, the dark matter measure is $\varrho(\theta) = 74.07$. This means that under the baseline model in the literature, we need to increase the amount of consumption data by a factor of 74.07 to match or exceed the precision afforded by the equity premium constraint in the estimation of any linear combination of p and ξ. Finally, in panel D, with relatively small but more frequent disasters (annual probability of 1.81%, average disaster size of 7%) and high risk aversion ($\gamma = 24$), the dark matter measure $\varrho(\theta)$ decreases to 1.49.

Monte Carlo Experiments. We now use simulations to illustrate the connections among the dark matter measure, the internal refutability, and the external validity of disaster risk models under different calibrations. We assume that the true local DGP has a time-varying relation between the expected log excess return and other dynamic parameters:

$$\overline{r}_n = \overline{r}(p_0, \xi_0) + \frac{\iota_t \delta_r}{\sqrt{n}}, \text{ with } \iota_t = \begin{cases} 1, & \text{when } 1 \leq t \leq \lfloor \pi n \rfloor \\ -1, & \text{when } \lfloor \pi n \rfloor < t \leq n, \end{cases} \tag{46}$$

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

$$\mathrm{E}^{\mathrm{Q}_0}\left[m_t(\theta_0)\right] = \frac{1}{\sqrt{n}}\left[0, 0, \lambda_t^{(2)}\right]^T \text{ with } \lambda_t^{(2)} \equiv \frac{\iota_t \delta_r}{\sqrt{(1-p_0)(1-\rho^2)\tau^2 + p_0 \varsigma^2}}. \tag{47}$$

Intuitively, $\lambda_t^{(2)}$ captures local moment instability. Instability in the DGP, such as structural breaks and nonstationarity, is a prevalent feature in the dynamics of macro quantities and asset prices.[23] Our specification of $\iota_t \delta_r$ in (46) follows the literature on local instability (e.g., Andrews, 1993; Sowell, 1996; Li and Müller, 2009).

Figure 3 shows three simulation experiments. Panel A displays the local power functions of the standard C tests based on the moment restrictions in (43), which is defined as the probability of rejecting the null hypothesis (i.e., the moment restrictions are correct) for a given δ_r. The local power function is defined in Section 3.4. The solid and dotted curves reflect the test powers when the DGP are characterized by calibrations A and C in Figure 2, respectively. In this experiment, we vary the local misspecification δ_r in the risk premium moment restriction. The DGP under calibration A features an excessively large amount of dark matter according to panel A of Figure 2, and not surprisingly, it has little internal refutability (i.e. low test power). This

[23] See, for example, Pesaran and Timmermann (1995); Pastor and Stambaugh (2001); Lettau, Ludvigson, and Wachter (2008); Lettau and Van Nieuwerburgh (2008).

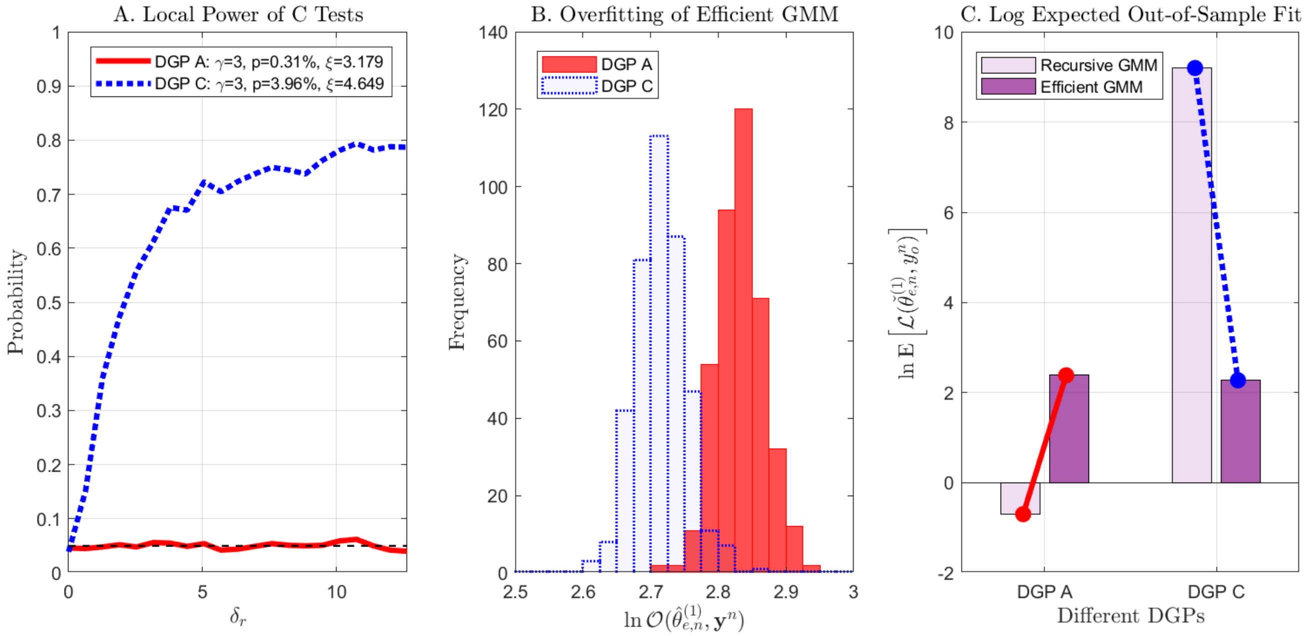

Figure 3: This figure shows Monte Carlo experiments for disaster risk models. In panel A, we simulate 1000 independent yearly time series with length $n = 2200$ to capture a pooled sample of 100 years for 22 countries, similar to Wachter (2013). In panels B and C, we set $\delta_r = 0.4$, and simulate 400 independent yearly time series with length $n = 100$ (i.e., 100 years) and break point $\pi = 1/2$. Only the uncertainties about p and ξ are accounted for, and all other parameters are auxiliary parameters as a part of the functional-form specification.

finding suggests that a larger dark matter measure signifies lower internal refutability of the model, which we confirm below in Theorem 1.

Panel B of Figure 3 displays the histograms of log overfitting measures $\ln \mathcal{O}(\hat{\theta}^{(1)}_{\mathrm{e,n}}, \mathbf{y}^n)$ of efficient GMM estimators for two DGPs under calibrations A and C in Figure 2. The estimator $\hat{\theta}_{\mathrm{e,n}}$ is based on the estimation sample $\mathbf{y}^n_{\mathrm{e}}$. In this experiment, we specify a structural break in the risk premium in the middle of the time-series sample ($\pi = 1/2$) with $\delta_r = 0.4$. Panel B shows that the efficient GMM estimator is likely to overfit the data in the calibrated structural model with a large dark matter measure. Thus, larger dark matter measure indicates higher overfitting tendency of the efficient GMM estimator. We address this relation in Theorem 2.

Panel C of Figure 3 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 3.4. Consistent with the conventional intuition, under the DGP C, the efficient GMM estimator yields a better expected out-of-sample fit than the recursive GMM estimator. This is because the identification information is more reliable and data-driven when the amount of dark matter is less excessively large. In contrast, the recursive GMM estimator delivers a better expected out-of-sample fit under the DGP A, which exhibits a much larger dark matter measure (we show in Theorem 3 below that this is a general property of the recursive estimator in our context). 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 DGP A suggests that the econometrician should prioritize robustness over efficiency when estimating models with an excessively large dark matter measure $\varrho(\theta)$.

4.2. Dark Matter of Time-Varying Disaster Risk Models

In the second example, we consider a time-varying disaster risk model similar to Gabaix (2012) and Wachter (2013). By making the probability of disasters time-varying, the model can generate time-varying risk premium. This additional flexibility can help explain much richer asset pricing dynamics beyond the equity premium.[24] We use this example to show how the dark matter can change as we extend a model.

Similar to Wachter (2013), we assume that the representative agent has recursive preferences with unit elasticity of intertemporal substitution (EIS), and maximizes her utility V_t as follows:

$$\ln V_t = (1-\delta)\ln C_t + \delta(1-\gamma)^{-1}\ln \mathbb{E}_t\left[V_{t+1}^{1-\gamma}\right], \tag{48}$$

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

$$\Delta c_{t+1} = \mu + \sigma_c \varepsilon_{c,t+1} - \zeta_{t+1}, \tag{49}$$

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

$$\zeta_{t+1} = x_{t+1}v_{t+1}, \tag{50}$$

with v_{t+1} following a truncated exponential distribution with lower bound $\underline{v}$, the same as in (38):

$$v_{t+1} \sim \mathbf{1}\{v_{t+1} > \underline{v}\}\xi e^{-\xi(v_{t+1}-\underline{v})}, \tag{51}$$

and x_{t+1} is a Bernoulli variable that is equal to 1 with probability $p_t = \max(\underline{p}, \widetilde{p}_t)$ and $\widetilde{p}_t$ evolving according to

$$\widetilde{p}_{t+1} = (1-\rho)\overline{p} + \rho\widetilde{p}_t + \sigma_p\sqrt{p_t}\varepsilon_{p,t+1}. \tag{52}$$

We impose a small positive lower bound $\underline{p}$ (= 1 bps) on the disaster probability p_t in solutions and simulations. The dividend D_t is modeled as levered consumption with log growth $\Delta d_{t+1} \equiv$

[24] The time-varying disaster risk has been used to explain the observed dynamics in macroeconomic quantities (e.g., Gourio, 2012), prices of derivatives (e.g., Seo and Wachter, 2018, 2019), and international exchange rates and capital flows (e.g., Gourio, Siemer, and Verdelhan, 2013; Dou and Verdelhan, 2017).

$\ln D_{t+1} - \ln D_t$ following the evolution:

$$\Delta d_{t+1} = \mu - \frac{1}{2}\varphi^2\sigma_c^2 + \phi\sigma_c\varepsilon_{c,t+1} - \phi\zeta_{t+1} + \varphi\sigma_c\varepsilon_{d,t+1}. \tag{53}$$

Following Barro (2006) and Wachter (2013), we assume the government bill has one-period maturity and can default when a disaster occurs; specifically, the return on the government bill is

$$r_{b,t+1} = \begin{cases} y_{b,t}, & \text{if } x_{b,t+1} = 0 \;\; \text{(i.e., no default);} \\ y_{b,t} - v_{t+1}, & \text{if } x_{b,t+1} = 1 \;\; \text{(i.e., default),} \end{cases} \tag{54}$$

where the Bernoulli variable $x_{b,t+1}$ captures the possibility of a government bill default with conditional probability q in the event of disaster.

The shock vectors $(v_{t+1}, \varepsilon_{c,t+1}, \varepsilon_{p,t+1}, \varepsilon_{d,t+1})^T$ are i.i.d. random vectors. The Bernoulli variable x_{t+1} is independent with the contemporaneous jump probability shock $\varepsilon_{p,t+1}$ and its leads in the time series, but x_{t+1} and the lags of $\varepsilon_{p,t+1}$ are dependent through the jump probability p_t. The default of government bills would never occur (i.e., $x_{b,t+1} = 0$) in the normal state (i.e., $x_{t+1} = 0$), while the default would occur (i.e., $x_{b,t+1} = 1$) with conditional probability q in the disaster state (i.e., conditioning on $x_{t+1} = 1$).

We denote by $z_{m,t}$ the log price-dividend ratio of the market portfolio. Using the log-linearization approximation, we search for equilibrium characterized by

$$z_{m,t} = A_{m,0} + A_{m,1}p_t, \tag{55}$$

where the expressions for constants $A_{m,0}$ and $A_{m,1}$, as functions of model parameters, are presented in Online Appendix 7. Let $r_{m,t+1}$ and $r_{f,t}$ be the log market return and log risk-free rate, respectively. The equilibrium excess log return of market portfolio and government bill are

$$r_{m,t+1} - r_{f,t} = \mu^e_{m,t} + \phi\sigma\varepsilon_{c,t+1} + \beta_p\sigma_p\sqrt{p_t}\varepsilon_{p,t+1} - \phi\zeta_{t+1} + \varphi\sigma_c\varepsilon_{d,t+1}, \tag{56}$$

$$r_{b,t+1} - r_{f,t} = \mu^e_{b,t} - x_{b,t+1}\zeta_{t+1}, \tag{57}$$

where $\mu^e_{m,t}$ and $\mu^e_{b,t}$ are characterized by

$$\mu^e_{m,t} = \phi\gamma\sigma_c^2 + \beta_p\lambda_p\sigma_p^2 p_t + \Delta_{\xi,\gamma}(\phi)p_t - \frac{1}{2}\left[\left(\beta_c^2 + \varphi^2\right)\sigma_c^2 + \beta_p^2\sigma_p^2 p_t\right], \tag{58}$$

$$\mu^e_{b,t} = q\Delta_{\xi,\gamma}(1)p_t, \tag{59}$$

with $\Delta_{\xi,\gamma}(\phi) \equiv \Xi(\gamma) - \Xi(\gamma - \phi) - \phi(\underline{v} + 1/\xi)$ for any ϕ and $\Xi(x) \equiv e^{x\underline{v}}\dfrac{\xi}{\xi - x} - 1$ for any x. The expressions for λ_p and β_p as functions of the model parameters, as well as the derivations of (56) and (57), are presented in Online Appendix 7.

Table 1: Parameters of time-varying disaster risk models

Panel A: Two sets of calibrations

	Benchmark model (M1)				Alternative model (M2)			
Preferences	δ	γ			δ	γ		
	0.97	3			0.97	3		
Consumption	μ	σ_c	ξ	$\underline{v}$	μ	σ_c	ξ	$\underline{v}$
	0.0252	0.02	5.2	0.1	0.0252	0.02	5.2	0.1
Dividend	ϕ	φ	q		ϕ	φ	q	
	2.6	6.8	0.4		2.6	6.8	0.4	
Disaster	$\overline{p}$	ρ	σ_p		$\overline{p}$	ρ	σ_p	
	0.0394	0.936	0.067		0.0394	0.85	0.10	

Panel B: Model and data moments of time-varying disaster risk model

Moment	Data estimate	Benchmark model (M1) 5%	Median	95%	Alternative model (M2) 5%	Median	95%
$\mathbb{E}[r_m - r_b]$	7.06	1.88	6.89	15.59	2.35	6.26	10.46
$\sigma(r_m)$	17.72	14.77	18.23	23.69	14.29	17.47	20.98
$\mathbb{E}[r_b]$	1.34	−0.80	3.04	4.52	0.45	2.81	4.12
$\sigma(r_b)$	2.66	0.78	1.73	3.79	1.19	2.08	3.87
$AC1(p-d)$	0.90	0.68	0.89	0.95	0.57	0.79	0.92
$\beta_1(r_m - r_b, p-d)$	−0.13	−0.49	−0.22	−0.09	−0.62	−0.30	−0.06

We focus on the parameters $\theta = (\mu, \sigma_c^2, \overline{p}, \rho, \sigma_p^2, \xi, \phi, \varphi, q, \gamma)^T$ with $d_\theta = 10$ when constructing the dark matter measure, while treating the parameters δ and $\underline{v}$ as auxiliary parameters as a part of the functional-form specification. The baseline parameters are $\theta^{(1)} = (\mu, \sigma_c^2, \overline{p}, \rho, \sigma_p^2, \xi, \phi, \varphi, q)^T$ with $d_{\theta,1} = 9$, while the nuisance parameters include $\theta^{(2)} = \gamma$. The observed variables $\mathbf{y}_t$ include $\{\Delta c_t, x_t, x_{b,t}, v_t, \Delta d_t\}$ for the baseline model and $\{z_{m,t}, r_{m,t}, r_{b,t}\}$ for the asset pricing component.

To compute the dark matter measure for the time-varying disaster risk model, we first formulate the economic model above as a structural model $\mathcal{Q}$ based on the baseline moments that capture the key dynamic features of (49) – (54) and the asset pricing moments that capture the dynamic features of (55) – (58), including the equity premium, return volatility, and predictability of excess returns using price-dividend ratios. The moments are detailed in Online Appendix 7.

Quantitative Analysis. The parameter values of the benchmark model M1 are taken from Wachter (2013) and are summarized in panel A of Table 1. In model M2, we reduce the disaster probability p_t's persistence ρ and increase its conditional volatility σ_p. To take the model to the data, we use annual real per-capita consumption growth (nondurables and services) from NIPA and returns on the CRSP value-weighted market portfolio. The sample moments are based on annual data from 1947 to 2010, and the model-implied moments use 60-year simulated annual data. For close comparison, we use a sample period of postwar data similar to that of Wachter (2013), and similarly, the moments of the model are calculated conditional on no disasters having occurred. For both models, the simulated first and second moments match the set of key asset pricing moments in the data reasonably well (see panel B of Table 1). Table 2 compares the dark matter measures of M1 and M2, two different calibrations of the time-varying disaster risk model. The two panels of Table 2 further illustrate the fact that different treatments of the nuisance parameters can affect the dark matter measure.

First, panel A of Table 2 contains fragility measures computed under the specification that treats γ as a nuisance parameter whose uncertainty needs to be taken into account. The row M1 of panel A in Table 2 reports the dark matter measure for model M1 when the unknown nuisance parameter is $\theta^{(2)} = \gamma$. The dark matter measure, $\varrho(\theta) = 1.49 \cdot 10^3$, is large. This implies that to match the precision of the efficient GMM estimator for the full structural model in all directions, the efficient GMM estimator based on the baseline model would approximately require a time-series sample that is $1.49 \cdot 10^3$ times as long.

A high value of $\varrho(\theta)$ suggests that the asset pricing implications of the structural model are highly sensitive to plausible perturbations of parameter values in θ. We compute the dark matter measure for each individual parameter in the vector $\theta^{(1)}$. All of the univariate measures are much lower than the worst-case 1-dimensional dark matter measure, $\varrho(\theta)$, with a larger dark matter measure for σ_p^2 (the variance of time-varying disaster probability shocks) and ρ (the persistence of time-varying disaster probability) 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 B of Table 2, we show dark matter measures when the preference parameters are fixed at certain values as a part of the functional-form specification. 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 dark matter measures in panel B are higher than those in panel A. In particular, the worst-case 1-dimensional dark matter $\varrho(\theta)$ increases dramatically from $1.4 \cdot 10^3$ to $4.46 \cdot 10^5$.

Monte Carlo Experiments. We use simulations to illustrate the connections among the dark matter measure, internal refutability, and external validity of time-varying disaster risk models. In this simulation experiment, we assume that all the parameters except ρ are treated as auxiliary parameters, whose uncertainty is not accounted for and whose values are subsumed into the functional form of the moment function (i.e. model specifications). From the dark matter

Table 2: Dark matter measures for the time-varying disaster risk models

Model	$\varrho(\theta)$	$\theta^{(1)}$								
		μ	σ_c^2	$\overline{p}$	ρ	σ_p^2	ξ	ϕ	φ	q
	Panel A: Accounting for uncertainty on nuisance parameters: $\theta^{(2)} = \gamma$									
M1	$1.49 \cdot 10^3$	1.99	1.98	1.48	261.9	164.9	2.10	2.04	1.97	1.01
M2	80.99	2.35	2.33	1.98	1.31	1.26	2.88	2.43	2.33	1.29
	Panel B: Not accounting for uncertainty on nuisance parameters: $\theta^{(2)}$ is empty									
M1	$4.46 \cdot 10^5$	3.79	3.83	3.20	271.0	166.7	7.53	2.74	3.82	1.05
M2	$4.69 \cdot 10^3$	5.65	5.49	2.59	7.13	5.28	7.04	6.93	5.45	1.37

Note: The direction corresponding to the worst-case 1-dimensional dark matter measure $\varrho(\theta)$ for benchmark model M1 is given by $v^*_{\max} = [0.20, -0.97, -0.15, 0.01, -0.05, 0.00, -0.00, -0.00, 0.01]$ in panel A when accounting for the uncertainty of nuisance parameter $\theta^{(2)} = \gamma$. In panel B, we treat γ as an auxiliary parameter like $\underline{v}$ and δ as a part of the functional-form specification, whose uncertainty is not accounted for when quantifying the dark matter.

evaluation in Table 2, we learn that the assumed identification of ρ based on the potentially misspecified asset pricing moment restrictions is a major source of model fragility for time-varying disaster risk models. Focusing on ρ 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 moments used in Table 2):

$$m_t(\theta) = \Omega(\theta)^{-1/2} \begin{bmatrix} x_{t-2}\left[x_t - \rho x_{t-1} - (1-\rho)\overline{p}\right] \\ r_t^e - \chi_3(\theta) z_{m,t} - \chi_4(\theta) z_{m,t-1} - \chi_5(\theta) \end{bmatrix} \text{ and } \theta = \rho, \tag{60}$$

where $r_t^e \equiv r_{m,t} - r_{b,t} + (\phi - x_{b,t})\zeta_t$ is the excess log return conditional on no disaster. Here, $\Omega(\theta)$ is the asymptotic covariance matrix of the untransformed moments, and it is a diagonal matrix whose elements are computed in Online Appendix 5.1. In (60), the first moment is the baseline moment, and the second the asset pricing moment. Clearly, the nuisance parameter vector $\theta^{(2)}$ is empty in this simulation example.

We assume that the true local DGP features a time-varying relation between the expected excess log return relative to the defaultable government bill:

$$r_{t,n}^e = r_t^e + \frac{\iota_t \delta_r}{\sqrt{n}}, \text{ with } \iota_t = \begin{cases} 1, & \text{when } 1 \le t \le \lfloor \pi n \rfloor \\ -1, & \text{when } \lfloor \pi n \rfloor < t \le n, \end{cases} \tag{61}$$

where the time series ι_t captures the structural breaks and the theoretical excess log return conditional on no disasters, $r_t^e \equiv r_{m,t} - r_{b,t} + (\phi - x_{b,t})\zeta_t$, with $r_{m,t}$ and $r_{b,t}$ evolving according to the model-implied relation in (56) and (57). The corresponding moment misspecification,

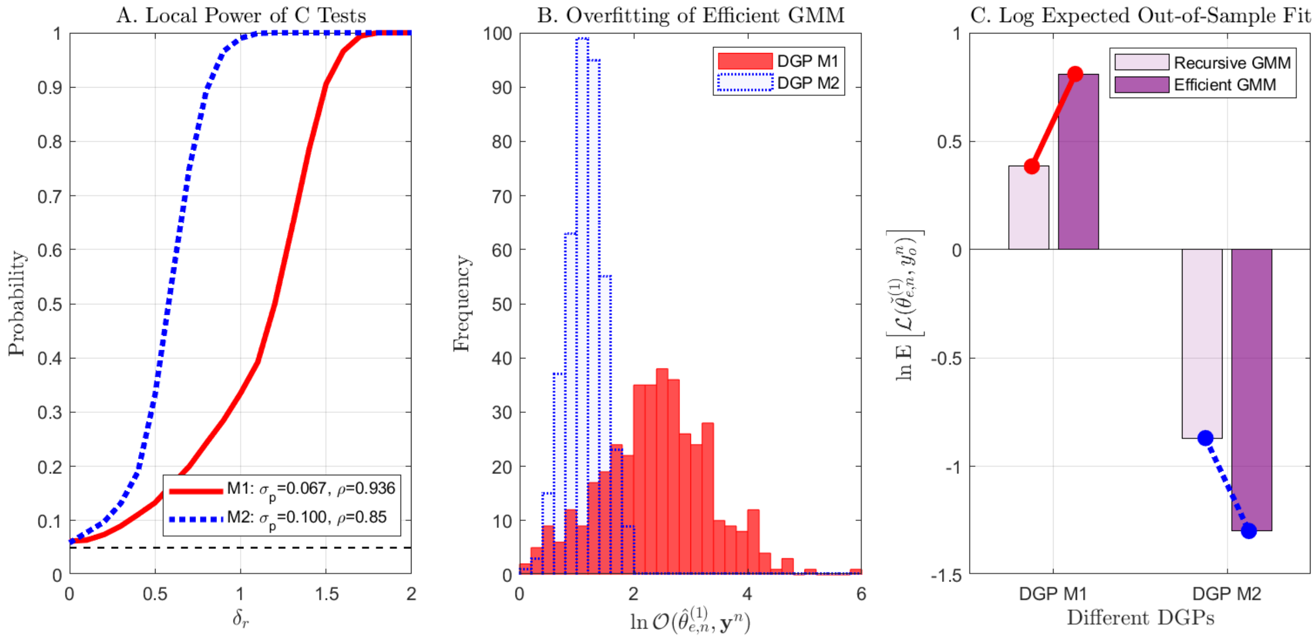

Figure 4: This figure shows Monte Carlo experiments for time-varying disaster risk models. In this simulation experiment where all the parameters except ρ are treated as auxiliary parameters, the dark matter measure is $\varrho(\theta) = 1.5 \cdot 10^5$ and $\varrho(\theta) = 4.3 \cdot 10^2$ for the model M1 and M2, respectively. In panel A, we simulate 1000 independent yearly time series with length $n = 2200$ to capture a pooled sample of 100 years for 22 countries similar to Wachter (2013). In panels B and C, we set $\delta_r = 0.05$ and simulate 400 independent yearly time series with length $n = 100$ (i.e., 100 years) and break point $\pi = 1/2$. Only the uncertainty about ρ is accounted for; all other parameters are treated as auxiliary parameters as a part of the functional-form specification.

evaluated at θ_0, are

$$\mathrm{E}^{\mathrm{Q}_0}\left[m_t(\theta_0)\right] = \frac{1}{\sqrt{n}}\left[0, \lambda_t^{(2)}\right]^T \text{ with } \lambda_t^{(2)} \equiv \frac{\iota_t \delta_r}{\sigma_c\sqrt{\phi^2+\varphi^2}}. \tag{62}$$

Figure 4 shows three different simulation experiments. The main takeaway is similar to that of the simulation experiments displayed in Figure 3 for rare disaster risk models. Panel A shows that the calibrated time-varying disaster risk model with a larger dark matter measure (benchmark model M1) has lower internal refutability (i.e., lower test power). Panel B shows that the calibrated model with a larger dark matter measure (the benchmark model M1) is likely to have more severe overfitting concerns for the efficient GMM estimator (than the alternative model M2). Panel C suggests that the econometrician should back off from efficiency to gain more robustness for the estimation results when the model contains a large amount of dark matter; specifically, the out-of-sample fit of the recursive GMM estimator may be superior to that of the efficient GMM estimator when the model has a large amount of dark matter.

Our analyses of the (time-varying) disaster-risk model here and in Section 4.1 relate to studies that have highlighted 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 (Stock and Wright, 2000). Our analysis highlights that in these 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.

4.3. Dark Matter of Long-Run Risk Models

In the third 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 and maximizes her lifetime utility:

$$V_t = \left[(1-\delta)C_t^{1-1/\psi} + \delta\left(\mathbb{E}_t\left[V_{t+1}^{1-\gamma}\right]\right)^{\frac{1-1/\psi}{1-\gamma}}\right]^{\frac{1}{1-1/\psi}}, \tag{63}$$

where C_t is consumption at time t, δ is the rate of time preference, γ is the coefficient of risk aversion for timeless gambles, and ψ is the elasticity of intertemporal substitution when there is perfect certainty. Log growth rate of consumption Δc_t, expected consumption growth x_t, and conditional volatility of consumption growth σ_t evolve as follows:

$$\Delta c_{t+1} = \mu_c + x_t + \sigma_t \epsilon_{c,t+1}, \tag{64a}$$

$$x_{t+1} = \rho x_t + \varphi_x \sigma_t \epsilon_{x,t+1}, \tag{64b}$$

$$\widetilde{\sigma}_{t+1}^2 = \overline{\sigma}^2 + \nu(\widetilde{\sigma}_t^2 - \overline{\sigma}^2) + \sigma_w \epsilon_{\sigma,t+1}, \tag{64c}$$

$$\sigma_{t+1}^2 = \max\left(\underline{\sigma}^2, \widetilde{\sigma}_{t+1}^2\right), \tag{64d}$$

where the shocks $\epsilon_{c,t}$, $\epsilon_{x,t}$, and $\epsilon_{\sigma,t}$ are i.i.d. standard normal variables and mutually independent. The volatility process (64c) potentially allows for negative values of $\widetilde{\sigma}_t^2$. 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_t^2 can be specified as a discrete-time version of the square root process.[25]

[25]To ensure that our analysis applies as closely as possible to the model widely used in the literature, we deliberately choose to follow Bansal, Kiku, and Yaron (2012, 2016). Particularly, following these papers, we also solve the model using a log-linearization approximation. 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 (2016) show, the resulting approximation error, when compared to the global numerical solution, is negligible. When computing the dark matter measure, we impose the lower bound on conditional variance σ_t^2 to reflect the specification of the conditional variance process.

Next, log dividend growth Δd_t follows

$$\Delta d_{t+1} = \mu_d + \phi_d x_t + \varphi_{d,c}\sigma_t\epsilon_{c,t+1} + \varphi_{d,d}\sigma_t\epsilon_{d,t+1}, \tag{65}$$

where $\epsilon_{d,t}$ is *i.i.d.* standard normal $N(0,1)$ and independent of the other shocks in (64a) – (64c). The equilibrium excess log return follows

$$r^e_{t+1} = \mu^e_{r,t} + \beta_c\sigma_t\epsilon_{c,t+1} + \beta_x\sigma_t\epsilon_{x,t+1} + \beta_\sigma\sigma_w\epsilon_{\sigma,t+1} + \varphi_{d,d}\sigma_t\epsilon_{d,t+1}, \tag{66}$$

where the conditional average log excess return is

$$\mu^e_{r,t} \approx \lambda_c\beta_c\sigma_t^2 + \lambda_x\beta_x\varphi_x\sigma_t^2 + \lambda_\sigma\beta_\sigma\sigma_w^2 - \frac{1}{2}\sigma^2_{r_m,t}, \tag{67}$$

and the conditional volatility of log excess returns $\sigma_{r_m,t}$ satisfies

$$\sigma^2_{r_m,t} \approx \beta_c^2\sigma_t^2 + \beta_x^2\sigma_t^2 + \beta_\sigma^2\sigma_w^2 + \varphi_{d,d}^2\sigma_t^2. \tag{68}$$

The expressions for λ_c, λ_x, λ_σ, β_c, β_x, and β_σ, as functions of the model parameters, are presented in Online Appendix 8.

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.

We focus on the parameters $\theta = (\mu_c, \rho, \varphi_x, \overline{\sigma}^2, \nu, \sigma_w, \mu_d, \phi_d, \varphi_{d,c}, \varphi_{d,d}, \gamma, \psi)^T$ with $d_\theta = 12$ when constructing the dark matter measure, while treating the parameters δ and $\underline{\sigma}$ as auxiliary parameters which are part of the functional-form specification. The baseline parameters are $\theta^{(1)} = (\mu_c, \rho, \varphi_x, \overline{\sigma}^2, \nu, \sigma_w, \mu_d, \phi_d, \varphi_{d,c}, \varphi_{d,d})^T$ with $d_{\theta,1} = 10$. We explicitly account for uncertainty about preference parameters γ and ψ by including them in the nuisance parameter vector $\theta^{(2)} = (\gamma, \psi)^T$. The observed variables $\mathbf{y}_t$ include $\{\Delta c_t, x_t, \sigma^2_t, \Delta d_t\}$ for the baseline model and $\{z_{m,t}, r^e_t\}$ for the asset pricing component. In our computation of the dark matter measure, we consider a system of moment restrictions based on the joint dynamics of time series $\mathbf{y}_t = (\Delta c_t, x_t, \sigma^2_t, \Delta d_t, z_{m,t}, r^e_t)^T$. The moments are detailed in Online Appendix 8.

Table 3: Parameters of long-run risk models

Panel A: Two sets of calibrations

	Benchmark model (M1)				Alternative model (M2)			
Preferences	δ	γ	ψ		δ	γ	ψ	
	0.9989	10	1.5		0.9989	27	1.5	
Consumption	μ_c	ρ	φ_x		μ_c	ρ	φ_x	
	0.0015	0.975	0.038		0.0015	0.975	0.038	
Dividend	μ_d	ϕ_d	$\varphi_{d,c}$	$\varphi_{d,d}$	μ_d	ϕ_d	$\varphi_{d,c}$	$\varphi_{d,d}$
	0.0015	2.5	2.6	5.96	0.0015	2.5	2.6	5.96
Volatility	$\overline{\sigma}$	ν	σ_w	φ_r	$\overline{\sigma}$	ν	σ_w	φ_r
	0.0072	0.999	$2.8e-6$	3.0	0.0072	0.98	$2.8e-6$	3.0

Panel B: Model and data moments of long-run risk model

	Data	Benchmark model (M1)			Alternative model (M2)		
Moment	estimate	5%	Median	95%	5%	Median	95%
$\mathbb{E}\left[r_f\right]$	0.57	−0.20	0.77	1.45	0.47	0.96	1.46
$\mathbb{E}\left[r_m - r_f\right]$	7.09	2.33	5.88	10.58	3.65	6.78	10.05
$\sigma\left(r_m\right)$	20.28	12.10	20.99	29.11	15.01	17.55	20.33
$AC1(p-d)$	0.87	0.39	0.68	0.89	0.29	0.63	0.83
$\beta_1\left(r_m - r_f, p-d\right)$	−0.09	−0.39	−0.11	0.09	−0.19	−0.06	0.04

Quantitative Analysis. The parameter values of benchmark model M1 follow Bansal, Kiku, and Yaron (2012) and are summarized in panel A of Table 3. As Bansal, Kiku, and Yaron (2012, page 194) show, the simulated first and second moments, based on the parametrization of benchmark model M1, match the set of key asset pricing moments in the data reasonably well. The same is true for alternative model M2, whose parameter values are also reported in panel A of Table 3. The performance of both models on matching asset pricing moments is reported in panel B of Table 3. Table 4 compares the dark matter measures of M1 and M2, two different calibrations of the long-run risk model. Echoing Table 1 for time-varying disaster risk models, the two panels of Table 4 show that different treatments of the nuisance parameters can affect the dark matter measure.

The simulated moments and sample moments are listed in panel B of Table 3. 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. Panel A of Table 4 contains dark matter measures computed when treating preference parameters as nuisance parameters whose uncertainty is taken into account in the construction of dark matter measures. The row M1 of panel A reports dark matter measures for benchmark model M1 when the nuisance parameters are $\theta^{(2)} = (\gamma, \psi)^T$. The dark matter measure, $\varrho(\theta) = 196.3$, is large, implying that to match the

Table 4: Dark matter measures for the time-varying disaster risk models

Model	$\varrho(\theta)$	$\theta^{(1)}$									
		μ_c	ρ	φ_x	$\overline{\sigma}^2$	ν	σ_w	μ_d	ϕ_d	$\varphi_{d,c}$	$\varphi_{d,d}$
	Panel A: Accounting for uncertainty on nuisance parameters: $\theta^{(2)} = (\gamma, \psi)^T$										
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
	Panel B: Not accounting for uncertainty on nuisance parameters: $\theta^{(2)}$ is 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 1-dimensional dark matter measure $\varrho(\theta)$ for benchmark model M1 is given by $v^*_{\max} = [0.000, 0.000, -0.000, 0.020, -0.001, 0.999, -0.001, 0.000, -0.000, 0.000]$ in panel A when accounting for the uncertainty of nuisance parameter $\theta^{(2)} = (\gamma, \psi)^T$. In panel B, we treat γ and ψ as auxiliary parameters like $\underline{\sigma}$ and δ as a part of the functional-form specification, whose uncertainty is not accounted for when calculating the dark matter measure.

precision of the estimator for the full model in all directions, the estimator based on the baseline model would approximately require a time-series sample that is 196.3 times as long.

In Table 4, we compute the dark matter measure for each individual parameter in the vector $\theta^{(1)}$. All of the univariate measures are much lower than the worst-case 1-dimensional fragility measure, $\varrho(\theta)$, with larger fragility measures 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. Similar to an important message from Table 2 for time-varying disaster risk models, we also learn here that it is not sufficient to consider perturbations of parameters one at a time to quantify model fragility based on the long-run risk examples. By contrast, panel B of Table 4 shows dark matter measures when the preference parameters are fixed at certain values as a part of the functional-form specification. The dark matter measures in panel B are larger than those in panel A. Specifically, the worst-case 1-dimensional dark matter measure, $\varrho(\theta)$, 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, respectively, 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 filter out x_t or σ_t^2 when computing the model fragility measure. Furthermore, because 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 and then estimate the corresponding parameter values as in Constantinides and Ghosh (2011), Bansal, Kiku, and Yaron (2016), and Schorfheide, Song, and Yaron (2018).

Monte Carlo Experiments. We use simulations to illustrate the connections among the dark matter measure, internal refutability, and external validity of long-run risk models. In this 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 model specifications. From Table 4, we learn that the assumed identification of ν is a major source of model fragility for long-run risk models. Focusing on ν simplifies the demonstration and increases the transparency by allowing us to focus on a few key (transformed) moment restrictions (i.e., a small yet essential subset of the moment restrictions used in constructing Table 4):

$$m_t(\theta) = \Omega(\theta)^{-1/2} \begin{bmatrix} (\tilde{\sigma}_t^2 - \overline{\sigma}^2)\epsilon_{\sigma,t+1} \\ r_{t+1}^e - \mu_{r,t}^e - \beta_c \sigma_t \epsilon_{c,t+1} - \beta_x \sigma_t \epsilon_{x,t+1} - \beta_\sigma \sigma_w \epsilon_{\sigma,t+1} \end{bmatrix} \text{ and } \theta = \nu, \tag{69}$$

where variables $\epsilon_{c,t+1}$, $\epsilon_{x,t+1}$, and $\epsilon_{\sigma,t+1}$ are the residuals in (64a) – (64d), depending on observed data and unknown parameters in θ, and $\mu_{r,t}^e$ is defined in (67) 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 (69), 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 DGP features a time-varying relation between the expected log excess return and the dynamic parameters:

$$r_{t,n}^e = r_t^e + \frac{\iota_t \delta_r}{\sqrt{n}}, \text{ with } \iota_t = \begin{cases} 1, & \text{when } 1 \le t \le \lfloor \pi n \rfloor \\ -1, & \text{when } \lfloor \pi n \rfloor < t \le n, \end{cases} \tag{70}$$

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

$$\mathrm{E}^{\mathrm{Q}_0}\left[m_t(\theta_0)\right] = \frac{1}{\sqrt{n}}\left[0, \lambda_t^{(2)}\right]^T \text{ with } \lambda_t^{(2)} \equiv \frac{\iota_t \delta_r}{\varphi_{d,d}\overline{\sigma}}. \tag{71}$$

Figure 5 shows three different simulation experiments, which reinforce the messages from

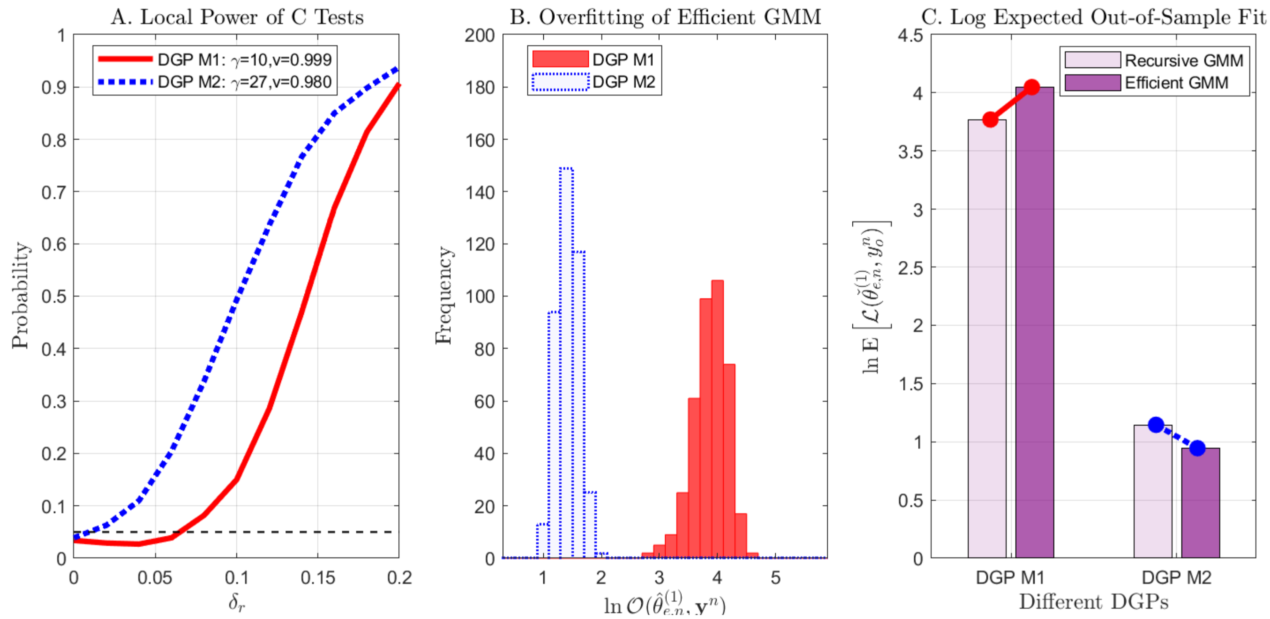

Figure 5: Monte Carlo experiments for long-run risk models. In this simulation experiment where all the parameters except ν are treated as auxiliary parameters, the dark matter measure is $\varrho(\theta) = 1.03 \cdot 10^5$ and $\varrho(\theta) = 22.34$ for the model M1 and M2, respectively. In panel 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 and subsumed into the functional form of the moment function (i.e., model specifications).

Figures 3 and 4 for rare-disaster risk and time-varying disaster risk models, respectively. Panel A shows that the DGP for the long-run risk model M1 features an excessively large amount of dark matter according to Table 4, and thus it has low internal refutability. Panel B displays the histograms of log overfitting measures $\ln \mathcal{O}(\hat{\theta}^{(1)}_{\mathrm{e,n}}, \mathbf{y}^n)$ of the efficient GMM estimators for the two calibrated models M1 and M2 in Table 4, showing that the calibrated structural model with too much dark matter (benchmark model M1) is likely to have more severe overfitting concerns for the efficient GMM estimator. Panel C compares the expected out-of-sample fit between recursive GMM estimators $\tilde{\theta}_{\mathrm{e,n}}$ and efficient GMM estimators $\hat{\theta}_{\mathrm{e,n}}$, suggesting that the econometrician should back off from efficiency to gain more robustness when estimating a model that exhibits a high level of dark matter.

5. Why Is Dark Matter a Concern

We first 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 DGPs in a semiparametric framework. We then formally establish the connection between the dark matter measure, model refutability, and out-of-sample fit. Using the semiparametric framework in Section 5.1, we show that a model with excessive dark matter lacks internal refutability and external validity. 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 5.2 and 5.3, respectively.

5.1. *Misspecification and Instability*

Here, we develop the econometric foundation needed for our analysis of economic dark matter. We set up a formal and general semiparametric framework to capture local misspecification and local instability. While misspecification and instability ultimately manifest themselves as perturbations in the moment restrictions imposed by the structural economic model, we start with a more fundamental concept — local misspecification of the structural model and local instability of the DGP — and derive the local moment misspecifications that follow. To model local misspecification, we adopt a statistical method similar to that of Hansen and Sargent (2001). The econometricians treat the reference probability measure Q_0 as an approximation of the measure corresponding to the true DGP, and they assume that the true process lies within a collection of local alternative DGPs that are statistically difficult to distinguish from Q_0 (i.e., a neighborhood of Q_0 in the space of probability measures). To model local instability, we generalize the local instability framework of Li and Müller (2009) to the semiparametric setting.

We first specify the true local DGP below, and we then introduce the concept of model misspecification and local instability.

Local DGPs. Our analysis is local in nature. We focus on a calibrated model with model parameter θ_0 as defined in (21), whose corresponding bivariate marginal distribution is Q_0. Next, we define the collection of local perturbations of Q_0, denoted by $\mathcal{N}(\mathrm{Q}_0)$, as follows.

Definition 2. *The collection $\mathcal{N}(\mathrm{Q}_0)$ is a collection of subsets of $\mathrm{L}^2(\mathrm{Q}_0)$, the space of square-integrable random variables on the probability space $(\mathcal{Y}\times\mathcal{Y}, \mathcal{F}\otimes\mathcal{F}, \mathrm{Q}_0)$. The collection $\mathcal{N}(\mathrm{Q}_0)$ consists of 1-dimensional parametric families 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):*[26]

$$\frac{d\mathrm{Q}_{s,f}}{d\mathrm{Q}_0} = 1 + sf + s\Delta(s), \tag{72}$$

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 $s \mapsto \mathrm{Q}_{s,f}$.

[26] The smoothness condition (72) is equivalent to the Hellinger-differentiability, shown in Appendix 5.3. It is a common regularity condition adopted for (semi)parametric inference (e.g., van der Vaart, 1988).

The following proposition summarizes the properties of the score f defined above. The proof of Proposition 1 can be found in Online Appendix 4.2.

Proposition 1 (Properties of Scores)**.** *If $f \in L^2(\mathrm{Q}_0)$ satisfies (72), 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, we specify the true local DGP by appealing to the concept in Definition 2. 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 DGP for $\mathbf{y}^n$ has 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

$$\frac{\mathrm{dQ}_{1/\sqrt{n}, f_{n,t}^*}}{\mathrm{dQ}_0} = 1 + \frac{f_{n,t}^*}{\sqrt{n}} + \Delta_n, \text{ where} \tag{73}$$

$$f_{n,t}^* \equiv \begin{bmatrix} 1 \\ b^*(t/n) \end{bmatrix}^T g^*(\mathbf{y}_{t-1}, \mathbf{y}_t) \text{ and } \sqrt{n}\Delta_n \to 0 \text{ in } L^2(\mathrm{Q}_0). \tag{74}$$

Vector g^* has two elements: $g^* = [g_1^*, g_2^*]^T$ with $g_1^*, g_2^* \in L^2(\mathrm{Q}_0)$. Thus, score function $f_{n,t}^*$ can be rewritten as

$$f_{n,t}^* = g_1^*(\mathbf{y}_{t-1}, \mathbf{y}_t) + g_2^*(\mathbf{y}_{t-1}, \mathbf{y}_t) b^*(t/n), \tag{75}$$

where $g_1^*(\mathbf{y}_{t-1}, \mathbf{y}_t)$ represents time-invariant perturbation and $g_2^*(\mathbf{y}_{t-1}, \mathbf{y}_t) b^*(t/n)$ represents time-varying perturbation. The unknown function $b^*(\cdot)$ is a deterministic function on $[0, 1]$ that generates local instability.[27] 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.

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

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

$$L_0^2(\mathrm{Q}_0) \equiv \left\{ \varsigma \in L^2(\mathrm{Q}_0) : \ \mathrm{E}^{\mathrm{Q}_0}[\varsigma(\mathbf{y}, \mathbf{y}')] = 0 \textit{ and } \mathrm{E}^{\mathrm{Q}_0}[\varsigma(\mathbf{y}, \mathbf{y}')|\mathbf{y}] = \mathrm{E}^{\mathrm{Q}_0}[\varsigma(\mathbf{y}', \mathbf{y})|\mathbf{y}] \right\}. \tag{76}$$

Then, the properties of scores in Proposition 1 can be simply restated as $f \in L_0^2(\mathrm{Q}_0)$. The true local DGP in (73), characterized by $f_{n,t}^*$, satisfies the following conditions:

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

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

[27] Similar to, for example, Andrews (1993), Sowell (1996), and Li and Müller (2009), we assume instability to be non-stochastic. The assumption is for technical simplicity. We can extend to stochastic instability following the arguments in Li and Müller (2009).

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

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

The first part of condition (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 (see Proposition 1 in Online Appendix 2.1). The second part of condition (i) that $g^*_1, g^*_2 \in L^2_0(\mathrm{Q}_0)$ is not restrictive, because guaranteed by Proposition 1.

Misspecification of Structural Models. To characterize the misspecification of the structural model with respect to the DGP $\mathrm{Q}_{s,f}$, we only need to characterize the relation between the moment function $m(\cdot, \theta_0)$ and the score f. Following the literature (e.g., Chen and Santos, 2018, and references therein), we consider the so-called tangent set for scores based on a given system of moments. For given moments $m(\cdot, \theta)$, the *tangent set* of the structural model $\mathcal{Q}$ at Q_0, denoted by $\mathcal{T}(\mathrm{Q}_0)$, consists of all scores $f \in L^2_0(\mathrm{Q}_0)$ such that the paths of locally perturbed distributions corresponding to f satisfy $\mathrm{Q}_{s,f} \in \mathcal{Q}$ defined in (14), and such that $\vartheta(\mathrm{Q}_{s,f})$, defined in (19) as a function of s, is differentiable with respect to s at $s = 0$. The tangent set $\mathcal{T}(\mathrm{Q}_0)$ is mathematically defined as follows:

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

We can further characterize the tangent set $\mathcal{T}(\mathrm{Q}_0)$ as follows:

$$\mathcal{T}(\mathrm{Q}_0) = \left\{ f \in L^2_0(\mathrm{Q}_0): \ \lambda(f) \in \mathsf{lin}(D) \right\}, \text{ with } \lambda(f) \equiv \mathrm{E}^{\mathrm{Q}_0}\left[m(\cdot, \theta_0) f\right], \tag{77}$$

where $\lambda(f)$ is a linear operator on $L^2_0(\mathrm{Q}_0)$ and linear space $\mathsf{lin}(D)$ is spanned by the columns of the Jacobian matrix D defined in (17) and (25). 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.[28] Intuitively, the inner product of moment and score functions $\lambda(f)$ is the local bias of moment restrictions evaluated at θ_0:

$$\mathrm{E}^{\mathrm{Q}_{s,f}}\left[m(\cdot, \theta_0)\right] = s\lambda(f) + o(s), \quad \text{as } s \to 0, \tag{78}$$

and thus, when $\lambda(f) \in \mathsf{lin}(D)$, there exists a local perturbation from θ_0 to make the moment

[28] One direct implication of (77) is that if $d_\theta < d_m$, then $\mathcal{T}(\mathrm{Q}_0) \neq L^2_0(\mathrm{Q}_0)$, and thus the distribution Q_0 is *locally overidentified* by $\mathcal{Q}$ (see Chen and Santos, 2018); further, if $d_m = d_\theta$, then $\mathcal{T}(\mathrm{Q}_0) = L^2_0(\mathrm{Q}_0)$, and thus Q_0 is *locally just identified* by $\mathcal{Q}$.

condition satisfied. The general relation in (78) summarizes how the stable local misspecification of structural models is translated into the stable local moment misspecification.

Moreover, the characterization (77) implies that $\mathcal{T}(\mathrm{Q}_0)$ is a linear space. Taken together, whenever $f^*_{n,t} \in L^2_0(\mathrm{Q}_0) \setminus \mathcal{T}(\mathrm{Q}_0)$, the structural model $\mathcal{Q}$ is *locally misspecified* with respect to the true local DGP characterized by $\mathrm{Q}^*_n \equiv \mathrm{Q}_{1/\sqrt{n}, f^*_{n,t}}$, which is defined in (73).

Similar to (77), the tangent set of the baseline structural model $\mathcal{Q}^{(1)}$ at Q_0 is characterized by

$$\mathcal{T}^{(1)}(\mathrm{Q}_0) \equiv \left\{ f \in L^2_0 : \lambda^{(1)}(f) \in \mathsf{lin}(D_{11}) \right\}, \text{ with } \lambda^{(1)}(f) \equiv \mathrm{E}^{\mathrm{Q}_0}\left[m^{(1)}(\cdot, \theta^{(1)}_0) f \right], \tag{79}$$

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

Instability of DGPs. To formalize the analysis on out-of-sample fit (i.e. external validity), we need to consider DGPs that allow for structural breaks in a non-stationary manner. Evidence abounds on structural changes and nonstationarity in asset pricing (for example, see the references in footnote 23). Econometric theory has largely focused on testing whether or not the model is stable.[29] However, little research has explored the implications of instability. 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, in the presence of possible instability, a model tends to have a poor out-of-sample fit if its dark matter measure is excessively large.

We consider the set $\mathcal{M}(\mathrm{Q}_0)$ consisting of all probability measures, each of which is a joint distribution for a Markov process $\mathbf{y}^n$ with local instability around the Markov process characterized by Q_0. We formalize the definition of $\mathcal{M}(\mathrm{Q}_0)$ as follows.

Definition 4. *The collection $\mathcal{M}(\mathrm{Q}_0)$ contains all joint distributions $\mathbb{P}_{1/\sqrt{n}, g, b}$, for local Markov DGPs, 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*

$$f_{n,t} = \begin{bmatrix} 1 \\ b(t/n) \end{bmatrix}^T g(\mathbf{y}_{t-1}, \mathbf{y}_t) \text{ with } g \in \mathcal{G}(\mathrm{Q}_0) \text{ and } b \in \mathcal{B}. \tag{80}$$

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

$$\theta_{n,t} \equiv \vartheta(\mathrm{Q}_{1/\sqrt{n}, f_{n,t}}), \quad \text{for any } f_{n,t} \in \mathcal{T}(\mathrm{Q}_0) \text{ with } 1 \le t \le n \text{ and sufficiently large } n. \tag{81}$$

Definition 4 says that all the local DGPs in $\mathcal{M}(\mathrm{Q}_0)$ are characterized by the pair $(g, b) \in$

[29] For example, see Nyblom (1989), Andrews (1993), and more recent contributions by Hansen (2000), Andrews (2003), and Elliott and Müller (2006).

$\mathcal{G}(\mathrm{Q}_0) \times \mathcal{B}$ and sample size n. The DGP 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 (81).

Next, we show how instability of the DGP is translated into the local moment misspecifications. Under the local DGP $\mathbb{P}_{1/\sqrt{n},g,b}$ 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 Online Appendix 4.2.

Proposition 2 (Local Biases of Moment Restrictions)**.** *Suppose Assumptions 1 – 5 hold. Under 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:*

$$\mathrm{E}^{\mathrm{Q}_{1/\sqrt{n},f_{n,t}}}\left[m_t(\theta_0)\right] = \left[\lambda(g_1) + \lambda(g_2)b(t/n)\right]/\sqrt{n} + o\left(1/\sqrt{n}\right), \tag{82}$$

where linear operator $\lambda(\cdot)$ *is defined in (77).*

Baseline Models. The baseline model provides a benchmark in the test power and out-of-sample fit analyses, characterizing the correct baseline parameter values $\theta_{n,t}^{(1)}$ and disciplining the asset pricing cross-equation restrictions. Throughout our analysis, we assume that the baseline model is correctly specified; that is, $g^* \in \mathcal{G}_{\mathrm{B}}(\mathrm{Q}_0)$, where

$$\mathcal{G}_{\mathrm{B}}(\mathrm{Q}_0) \equiv \left\{g \in \mathcal{G}(\mathrm{Q}_0) : \lambda^{(1)}(g_1) = 0, \text{ and } \lambda^{(1)}(g_2) \in \mathsf{lin}(D_{11})\right\}. \tag{83}$$

We can always shift θ_0 locally to make sure $\lambda^{(1)}(g_1) = 0$ and $\lambda^{(1)}(g_2) \in \mathsf{lin}(D_{11})$ if $\lambda^{(1)}(g_1), \lambda^{(1)}(g_2) \in \mathsf{lin}(D_{11})$. Thus, without loss of generality, we can just focus on $\mathcal{G}_{\mathrm{B}}(\mathrm{Q}_0)$.

The following corollary shows that the correct baseline parameters are time-invariant when the baseline model is correctly specified. The proof can be found in Online Appendix 4.3.

Corollary 1 (Correct Baseline Parameters)**.** *Suppose Assumptions 1 – 6 hold. Then, the correct baseline parameters* $\theta_{n,t}^{(1)} \equiv \vartheta^{(1)}(\mathrm{Q}_{1/\sqrt{n},f_{n,t}})$ *exist 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*

$$\theta_{n,t}^{(1)} - \theta_0^{(1)} = -(D_{11}^T D_{11})^{-1} D_{11}^T \lambda^{(1)}(f_{n,t})/\sqrt{n} + o\left(1/\sqrt{n}\right). \tag{84}$$

Here, linear operator $\lambda^{(1)}(\cdot)$ *is defined in (79).*

5.2. *Dark Matter and Lack of Refutability*

The local asymptotic power function in (33) can be rewritten as follows under the general semi-parametric framework:

$$q(g, \check{\varphi}) \equiv \lim_{n\to\infty} \int \check{\varphi}_n \mathrm{d}\mathbb{P}_{1/\sqrt{n},g,0}, \quad \forall\, g \in \mathcal{G}(\mathrm{Q}_0) \text{ such that } g_1 \in \mathcal{T}^{(1)}(\mathrm{Q}_0), \tag{85}$$

where $\check{\varphi} \equiv \{\check{\varphi}_n\}_{n\geq 1}$ is the sequence of test statistics and $q(\cdot, \check{\varphi})$ is a mapping from $\mathcal{G}(\mathrm{Q}_0)$ to $[0, 1]$. The test $\check{\varphi}_n$ has a local asymptotic level α if $q(g, \check{\varphi}) \leq \alpha$ for any $g_1 \in \mathcal{T}(\mathrm{Q}_0)$. A specification test $\check{\varphi} = \{\check{\varphi}_n\}_{n\geq 1}$ with a local asymptotic power function $q(\cdot, \check{\varphi})$ 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 $\Phi_\alpha(\mathrm{Q}_0)$.

The guaranteed local asymptotic power of tests, over all feasible local DGPs, 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).[30] 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 Online Appendix 4.1.

Before introducing the theorem, we define the set of alternatives for the maximin local power of GMM specification tests.

Definition 5 (Set of DGPs for Worst-Case Power)**.** *The set of alternatives for the maximin local power of GMM specification tests is defined as follows:*

$$\mathcal{A}_\kappa(\mathrm{Q}_0) \equiv \left\{g \in \mathcal{G}_B(\mathrm{Q}_0) : |\lambda^{(2)}(g_1)| \geq \kappa \text{ and } \lambda^{(2)}(g_1) \perp \mathsf{lin}(D_{21}, D_{22})\right\}, \tag{86}$$

where $\lambda^{(2)}(g_1) \equiv \mathrm{E}^{\mathrm{Q}_0}\left[m^{(2)}(\cdot, \theta_0)g_1\right]$ *includes the bottom* $d_m - d_{m,1}$ *elements of* $\lambda(g_1)$ *defined in (77), and* $\mathsf{lin}(D_{21}, D_{22})$ *is the linear space spanned by the column vectors of* $[D_{21}, D_{22}]$.

Because of $g \in \mathcal{G}_{\mathrm{B}}(\mathrm{Q}_0)$ and $\lambda^{(2)}(g_1) \perp \mathsf{lin}(D_{21}, D_{22})$, it must hold that $\lambda(g_1) \notin \mathsf{lin}(D)$, and thus, the full model $\mathcal{Q}$ is misspecified with respect to the local DGP with joint distribution $\mathbb{P}_{1/\sqrt{n},g,0}$ if $g \in \mathcal{A}_\kappa(\mathrm{Q}_0)$. The constant $\kappa > 0$ characterizes the minimum level of misspecification of the full model $\mathcal{Q}$ with respect to the local DGPs in $\mathcal{A}_\kappa(\mathrm{Q}_0)$.

Theorem 1 (Model Refutability)**.** *Suppose Assumptions 1 – 6 hold. The local asymptotic power*

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

of maximin tests is bounded above by

$$\sup_{\breve{\varphi}\in\Phi_\alpha(\mathrm{Q}_0)} \inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} q(g,\breve{\varphi}) \le M_{\frac{d_{m,2}-d_{\theta,2}}{2}}\left(\sqrt{\frac{\kappa^2}{1+\varrho(\theta_0)}}, \sqrt{c_{1-\alpha}}\right), \tag{87}$$

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

$$M_{\frac{d_{m,2}-d_{\theta,2}}{2}}(0,\sqrt{c_{1-\alpha}}) = \alpha. \tag{88}$$

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

$$\sup_{\breve{\varphi}\in\Phi_\alpha(\mathrm{Q}_0)} \inf_{g\in\mathcal{A}_\kappa(\mathrm{Q}_0)} q(g,\breve{\varphi}) \ \to \alpha, \ \text{ as } \varrho(\theta_0)\to\infty. \tag{89}$$

The generalized Marcum Q-function $M_\gamma(x_1,x_2)$ strictly increases in γ and x_1, and it strictly decreases in x_2 (e.g., Sun, Baricz, and Zhou, 2010, Theorem 1). An intuitive interpretation of (87) is that there exists an 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 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 are distant from the null, and thus, the upper bound for the test power (the right-hand side of (87)) naturally increases with κ.

Practitioners can use the dark matter measure, $\varrho(\theta)$, to detect whether a model is fragile. Although the relative-sample-size interpretation of the dark matter measure is intuitive, the quantity $M_{\frac{d_{m,2}-d_{\theta,2}}{2}}\left(\sqrt{\frac{\kappa^2}{1+\varrho(\theta_0)}}, \sqrt{c_{1-\alpha}}\right)$ derived in Theorem 1 provides a statistically meaningful metric because it characterizes an upper bound for the test power given a misspecification level κ. We refer to $M_{\frac{d_{m,2}-d_{\theta,2}}{2}}\left(\sqrt{\frac{\kappa^2}{1+\varrho(\theta_0)}}, \sqrt{c_{1-\alpha}}\right)$ as a "dark-matter bound" for refutability of a model. It can be viewed as an analogy to the well-known Anderson-Rubin (AR) test statistic, which is useful for detecting whether linear instrumental variables (IVs) are weak (e.g., Andrews and Stock, 2007).

5.3. *Dark Matter and Poor Out-of-Sample Fit*

The constant misspecification over time does not affect the out-of-sample fit of estimated models based on in-sample data. We focus on how model instability affects the out-of-sample fit of an estimated time-series model.

The asymptotic expected overfitting measure of the sequence of estimators $\check{\theta}^{(1)}_{\mathrm{e,n}}$ is[31]

$$\omega(g, b, \check{\theta}^{(1)}_{\mathrm{e}}) \equiv \lim_{l\to\infty} \lim_{n\to\infty} \int \mathcal{O}(\check{\theta}^{(1)}_{\mathrm{e,n}}, \mathbf{y}^n) \mathbf{1}_{\{|\mathcal{O}(\check{\theta}^{(1)}_{\mathrm{e,n}}, \mathbf{y}^n)| \le l\}} \mathrm{d}\mathbb{P}_{1/\sqrt{n}, g, b}, \tag{90}$$

where $\check{\theta}^{(1)}_{\mathrm{e}} \equiv \left\{\check{\theta}^{(1)}_{\mathrm{e,n}}\right\}_{n\ge1}$ is a sequence of GMM estimators, and the overfitting measure $\mathcal{O}(\check{\theta}^{(1)}_{\mathrm{e,n}}, \mathbf{y}^n)$ is defined in (37). The asymptotic expected overfitting measure $\omega(g, b, \check{\theta}^{(1)}_{\mathrm{e}})$ quantifies the extent to which the structural model over-fits the data when the true local DGP 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 degrees of overfitting are sizable should be penalized.

Overfitting of the Efficient GMM Estimator $\hat{\theta}_{e,n}$. Recall that $\lambda(g_2)$ captures the magnitude of instability. Before introducing the theorem on model overfitting tendency (Theorem 2), we define the uncertainty set of possible magnitudes of instability for discussing the worst-case overfitting tendency.

Definition 6 (A Set of Alternative DGPs for Worst-Case Overfitting)**.** *The set of possible magnitudes of instability for computing the worst-case overfitting tendency is defined as follows:*

$$\mathcal{U}_\kappa(\mathrm{Q}_0) \equiv \{g \in \mathcal{G}_B(\mathrm{Q}_0) : \lambda(g_1) \in \mathit{lin}(D),\ \text{and } |\lambda(g_2)| \le \kappa\}, \tag{91}$$

where $\lambda(g_2) \equiv \mathrm{E}^{\mathrm{Q}_0}[m(\cdot, \theta_0) g_2]$ is defined in (77). This sets focuses on the scores such that $\lambda(g_1) \in \mathit{lin}(D)$ to turn off the possibility of constant misspecifications.

The constant $\kappa > 0$ characterizes the maximum level of instability of the DGPs $\mathbb{P}_{1/\sqrt{n},g,b}$. A larger κ allows for a higher degree of instability in the fragility analysis.

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

$$\sup_{g\in\mathcal{U}_\kappa(\mathrm{Q}_0), b\in\mathcal{B}} \omega(g, b, \hat{\theta}^{(1)}_e) = d_{\theta,1} + c(\pi)\varrho(\theta_0)\kappa^2, \tag{92}$$

where $c(\pi) \equiv \pi\left(1 + \sqrt{\dfrac{\pi}{1-\pi}}\right)$ with $0 < \pi \le 1/2$. Therefore, the asymptotic expected overfitting of the efficient GMM estimator sequence $\hat{\theta}_e \equiv \left\{\hat{\theta}_{e,n}\right\}_{n\ge1}$ can be arbitrarily large, being linearly related to the dark matter measure.

[31]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).

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 instability. The latter component is the focus of our paper and directly depends on the dark matter measure, which vanishes if there is no local instability (i.e., $\kappa = 0$).

Overfitting of the Recursive GMM Estimator $\tilde{\theta}_{e,n}$**.**

Theorem 3 (Robust Estimation)**.** *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:*

$$\sup_{g\in\mathcal{U}_\kappa(\mathrm{Q}_0),b\in\mathcal{B}} \omega(g,b,\tilde{\theta}_e^{(1)}) = d_{\theta,1}, \tag{93}$$

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 identification assumptions imposed by $\mathrm{E}^{\mathrm{Q}_0}\left[m_t^{(2)}(\theta)\right] = 0$; thus, its overfitting is not affected by 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, is 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 can be appreciated through the sensitivity of efficient GMM estimators to local instability. 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 4 in Online Appendix 2.2,

$$\operatorname*{wlim}_{n\to\infty} \left[\begin{array}{c} \frac{1}{\sqrt{\pi n}} \sum_{t\leq \pi n} m_t(\theta_0) \\ \frac{1}{\sqrt{(1-\pi)n}} \sum_{t>\pi n} m_t(\theta_0) \end{array}\right] = \left[\begin{array}{c} m_e \\ m_o \end{array}\right], \text{ with } \mathrm{E}\left[\begin{array}{c} m_e \\ m_o \end{array}\right] = \left[\begin{array}{c} \nu_e(g,b,\pi) \\ \nu_o(g,b,\pi) \end{array}\right], \tag{94}$$

where (m_e, m_o) are independent normal variables with the identity covariance matrix and the means as follows:

$$\nu_e(g,b,\pi) \equiv \frac{\lambda(g^T)}{\sqrt{\pi}} \begin{bmatrix} \pi \\ \int_0^\pi b(u)\mathrm{d}u \end{bmatrix} \text{ and } \nu_o(g,b,\pi) \equiv \frac{\lambda(g^T)}{\sqrt{1-\pi}} \begin{bmatrix} 1-\pi \\ \int_\pi^1 b(u)\mathrm{d}u \end{bmatrix}. \tag{95}$$

Further, Proposition 7 in Online Appendix 2.2 shows that the in- and out-of-sample estimators satisfy

$$\underset{n\to\infty}{\mathrm{wlim}} \begin{bmatrix} \sqrt{\pi n}(\hat{\theta}^{(1)}_{\mathrm{e,n}} - \theta^{(1)}_{\mathrm{e,n}}) \\ \sqrt{(1-\pi)n}(\hat{\theta}^{(1)}_{\mathrm{o,n}} - \theta^{(1)}_{\mathrm{o,n}}) \end{bmatrix} = \begin{bmatrix} \hat{\theta}^{(1)}_e \\ \hat{\theta}^{(1)}_o \end{bmatrix}, \text{ with } \mathrm{E}\begin{bmatrix} \hat{\theta}^{(1)}_e \\ \hat{\theta}^{(1)}_o \end{bmatrix} = -(L_{\mathrm{F}} - L_{\mathrm{B}}) \begin{bmatrix} \nu_e(g,b,\pi) \\ \nu_o(g,b,\pi) \end{bmatrix}, \tag{96}$$

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 $(\hat{\theta}^{(1)}_e, \hat{\theta}^{(1)}_o)$ 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

$$\mathbf{I}_{\mathrm{F}}^{1/2}\mathrm{E}\left[\hat{\theta}^{(1)}_{\mathrm{e,n}} - \hat{\theta}^{(1)}_{\mathrm{o,n}}\right] = \beta \mathrm{E}\left[m_o - m_e\right], \tag{97}$$

where $\beta = -\mathbf{I}_{\mathrm{F}}^{1/2}(L_{\mathrm{F}} - L_{\mathrm{B}})$. 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 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 presented by Andrews, Gentzkow, and Shapiro (2017). Relative to their measure, we add a baseline model as a benchmark (replacing L_{F} by $L_{\mathrm{F}} - L_{\mathrm{B}}$), and we normalize the expected change in the efficient GMM estimator by its asymptotic covariance matrix in the full model (multiplying $\mathrm{E}[\hat{\theta}^{(1)}_{\mathrm{e,n}} - \hat{\theta}^{(1)}_{\mathrm{o,n}}]$ by $\mathbf{I}_{\mathrm{F}}^{1/2}$).

6. Discussion: How to Deal with Dark Matter

We discuss what to do with models that rely excessively on dark matter. When the moment restrictions that summarize a structural model have a large dark matter measure, the concern of misspecification and instability can offset the efficiency gain in estimation from imposing cross-equation restrictions. In such cases, a robust estimation procedure is particularly important. One candidate robust estimation method is the recursive GMM estimation (e.g., Hansen, 2007b, 2012), described in Section 3.4. Although the recursive GMM estimator has worse in-sample fit than does the efficient GMM estimator, it can deliver better out-of-sample fit when the dark matter measure is excessively large. While the original impetus of the recursive GMM estimation was primarily computational, we advocate it as a robust estimation procedure against model fragility. A hybrid estimator combining the efficient and recursive GMM estimators, by deviating from the optimal weighting matrix toward the weighting matrix associated with the recursive GMM estimator, is a potential way to construct estimators that optimally balance robustness and efficiency. Thus, our dark matter measure can be used when trading off robustness against statistical efficiency. We leave the systematic econometric investigation on optimal robust estimation in the presence of large dark matter measures for future research.

Because of their lack of refutability under a conventional econometric framework, a robust and powerful specification test is important when evaluating asset pricing models with excessively large dark matter measures. In a companion paper, Cheng, Dou, and Liao (2021) extend the econometric setting of this paper to a unified weak identification framework and provide a specification test robust to severe information imbalances. The proposed robust specification test is built on the C statistic of Eichenbaum, Hansen, and Singleton (1988) and the recent developments in conditional inference using a sufficient statistic, which can be interpreted as a quantity capturing information in the fundamental data decoupled from that of the asset pricing moment restrictions. Such decoupling is crucial for preserving limited information in the fundamental data and efficiently using it to evaluate the asset pricing moment restrictions.

Furthermore, our dark matter measure captures model complexity and thus can be used for calibration comparisons. As shown in Section 4, different calibrations of the same model (with the same functional form and moments) can have distinct dark matter measures, and those with smaller dark matter measures are preferable due to a lower overfitting tendency (i.e., better out-of-sample fit).

Even with the same economic specification and parameter values, two models can still differ from each other in the pre-specified moment restrictions that modelers focus on, and one can use our dark matter measure to exclude fragile moment restrictions. For example, the method of stepwise forward moment selection adds additional moment restrictions to the baseline one by one and features "pretesting" combined with a sequential search strategy. This method is analogous to a stepwise forward strategy for variable selection in regressions. In each step, we first pick out the set of additional moments that cannot be rejected by the robust specification test of Cheng, Dou, and Liao (2021), then search for the moment that leads to the smallest dark matter measure. We leave the formal investigation on how to use the dark matter measure for optimal moment selection for future research.

The dark matter measure highlights the parameter combinations in which model fragility is embedded. Such worst-case parameter combinations (or directions in the parameter space), as multivariate sensitivity diagnostics, suggest avenues for improving the robustness of a model by bring in additional data that better identify the problematic parameter combinations under the baseline model. For example, Barro and Ursúa (2012) and Nakamura, Steinsson, Barro, and Ursúa (2013) use international data to better estimate the distribution of consumption disasters. Without having collected the comprehensive international macroeconomic data and direct evidence on disasters, rare-disaster risk models (e.g., Rietz, 1988) must draw information about the distribution of disasters from asset prices.

Moreover, from a modeler's perspectives, one approach for improving model robustness is to enrich the baseline model specification by endogenizing the key dynamics of the fundamental variables and connecting the key baseline parameters to a broader set of fundamental 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, helping bring in data on the dynamics of research and development (R&D) investment to strengthen statistical identification of the persistence parameter of the aggregate consumption process. Dou, Ji, Tian, and Wang (2021) further connect the endogenous persistent consumption growth with the dynamics of capital misallocation. Another approach is to modify the belief formation mechanism so that the baseline model parameters including those which govern the evolution of agents' (potentially heterogeneous) beliefs are better identified by the baseline moment restrictions and do not rely excessively on the restrictions implied by the asset pricing moments (e.g., Barberis, Shleifer, and Vishny, 1998; Hansen and Sargent, 2010; Chen, Joslin, and Tran, 2012; Greenwood and Shleifer, 2014; Nagel and Xu, 2019). Specifically, the expectations of the agents inside the model can be better disciplined and identified by micro-evidence from institution and household portfolio choice and survey expectations data in those models, which may help reduce the amount of dark matter of the model. As pointed out by Brunnermeier, Farhi, Koijen, Krishnamurthy, Ludvigson, Lustig, Nagel, and Piazzesi (2021), by understanding investors' asset demand curves and fund flows, we can probably make the connection between asset prices and quantities more measurable and tangible, thus hopefully connect the key baseline parameters to a broader set of fundamental data and reduce the "dark matter" in asset pricing models (e.g., Koijen and Yogo, 2019; Dou, Kogan, and Wu, 2020).

Lastly, our measure of model fragility helps identify situations in which it could be particularly relevant to incorporate parameter uncertainty and agents' robustness considerations within an economic model. In the literature on structural estimation, including rational-expectations econometrics, economic assumptions (i.e., cross-equation restrictions) have been used extensively to increase efficiency of estimation of structural parameters.[32] If a model is fragile, its 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

[32] Classic examples include Saracoglu and Sargent (1978), Hansen and Sargent (1980), and Campbell and Shiller (1988).

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 inside the model, which is one of the key motivations for research on 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 explicitly incorporates robustness considerations into agents' decision problems, recognizing that the traditional assumption that agents possess precise knowledge of the relevant probability distributions may not be justifiable, and may serve as a source of economic dark matter. Therefore, our dark matter measure helps detect situations in which parameter uncertainty and agents' robustness considerations can substantially alter the model's implications.

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 — the dark matter measure. We show that our information imbalance measure captures a useful model property intrinsically connected to the model's tendency to over-fit the data in sample. Our dark matter measure should be used as a calibration selection criterion for structural economic models. When faced with a set of candidate calibrations consistent with available data, selecting the less fragile calibrated model can be an appealing criterion from the point of view of model refutability and out-of-sample performance.

The proposed dark matter measure is easy to implement. The worst-case direction provides guidance on which features of the model are most vulnerable to in-sample overfitting, suggesting which types of additional data or economic mechanisms would be needed to improve model refutability and alleviate out-of-sample fit concerns.

Robust inference and testing econometric procedures need to be seriously considered when the alert of excessively large dark matter is fired on a specific asset pricing model. Inspired by the idea of recursive GMM estimators, we show that robust point estimation procedures become very essential to ensuring low overfitting tendency (achieving reliable out-of-sample fit) in the presence of excessively large dark matter measures. Moreover, we show that robust specification testing procedures become vital to conducting a reliable model evaluation when the model under scrutiny features severe information imbalances (e.g., Cheng, Dou, and Liao, 2021).

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

Adrian, Tobias, Richard K. Crump, and Emanuel Moench, 2015, Regression-based estimation of dynamic asset pricing models, *Journal of Financial Economics* 118, 211–244.

Ai, Chunrong, and Xiaohong 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, Tomohiro, 2007, Bayesian predictive information criterion for the evaluation of hierarchical bayesian and empirical bayes models, *Biometrika* 94, 443–458.

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

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

Andrews, Donald W. K., and James H. Stock, 2007, *Inference with Weak Instruments*, vol. 3 of *Econometric Society Monographs* . pp. 122–173 (Cambridge University Press).

Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro, 2017, Measuring the sensitivity of parameter estimates to estimation moments, *The Quarterly Journal of Economics* 132, 1553–1592.

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

Athey, Susan, and Guido W. Imbens, 2019, Machine learning methods economists should know about, Papers Stanford University.

Bansal, Ravi, Dana Kiku, and Amir Yaron, 2012, An empirical evaluation of the long-run risks model for asset prices, *Critical Finance Review* 1, 183–221.

Bansal, Ravi, Dana Kiku, and Amir Yaron, 2016, Risks for the long run: Estimation with time aggregation, *Journal of Monetary Economics* 82, 52 – 69.

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

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307 – 343.

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

Barro, Robert J., and José 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, Peter J., Chris A. J. Klaassen, Ya'acov Ritov, and Jon A. Wellner, 1993, *Efficient and adaptive estimation for semiparametric models* . Johns Hopkins Series in the Mathematical Sciences (Johns Hopkins University Press, Baltimore, MD).

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

Bradley, Richard 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.

Brunnermeier, Markus, Emmanuel Farhi, Ralph S J Koijen, Arvind Krishnamurthy, Sydney C Ludvigson, Hanno Lustig, Stefan Nagel, and Monika Piazzesi, 2021, Review Article: Perspectives on the Future of Asset Pricing, *The Review of Financial Studies*.

Campbell, John, and Motohiro Yogo, 2006, Efficient tests of stock return predictability, *Journal of Financial Economics* 81, 27–60.

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

Campbell, John Y., and John H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205–251.

Campbell, John Y, and Robert J Shiller, 1988, Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661–76.

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

Chen, Hui, Scott Joslin, and Ngoc-Khanh Tran, 2012, Rare Disasters and Risk Sharing with Heterogeneous Beliefs, *Review of Financial Studies* 25, 2189–2224.

Chen, Xiaohong, and Andres Santos, 2018, Overidentification in Regular Models, *Econometrica* 86, 1771–1817.

Cheng, Xu, Winston Wei Dou, and Zhipeng Liao, 2021, Macro-Finance Decoupling: Robust Evaluations of Macro Asset Pricing Models, Working papers University of Pennsylvania.

Christiano, Lawrence J., and Martin Eichenbaum, 1992, Current real-business-cycle theories and aggregate labor-market fluctuations, *The American Economic Review* 82, 430–450.

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

Constantinides, George M., and Anisha Ghosh, 2011, Asset pricing tests with long-run risks in consumption growth, *The Review of Asset Pricing Studies* 1, 96–136.

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

Dou, Winston, Yan Ji, Di Tian, and Pengfei Wang, 2021, Asset pricing with misallocation, Working papers University of Pennsylvania.

Dou, Winston, and Adrien Verdelhan, 2017, The volatility of international capital flows and foreign assets, Unpublished working paper MIT.

Dou, Winston Wei, Leonid Kogan, and Wei Wu, 2020, Common fund flows: Flow hedging and factor pricing, Working papers University of Pennsylvania, MIT, and Texas A&M University.

Dou, Winston Wei, Lucian A. Taylor, Wei Wang, and Wenyu Wang, 2020, Dissecting bankruptcy frictions, *Journal of Financial Economics* Forthcoming.

Efron, Bradley, 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, Martin S., Lars Peter Hansen, and Kenneth 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, Graham, and Ulrich K. Müller, 2006, Efficient Tests for General Persistent Time Variation in Regression Coefficients, *The Review of Economic Studies* 73, 907–940.

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

Fisher, R. A., 1922, On the mathematical foundations of theoretical statistics, *Phil. Trans. R. Soc.*

Lond. A 222, 309–68.

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

Foster, F. Douglas, Tom Smith, and Robert 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, Xavier, 2012, Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance, *The Quarterly Journal of Economics* 127, 645–700.

Gârleanu, Nicolae, Stavros Panageas, and Jianfeng 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.*

Giglio, Stefano, and Dacheng Xiu, 2021, Asset Pricing with Omitted Factors, *Journal of Political Economy.*

Gilboa, Itzhak, and David Schmeidler, 1989, Maxmin expected utility with non-unique prior, *Journal of Mathematical Economics* 18, 141–153.

Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti, 2014, Misspecification-Robust Inference in Linear Asset-Pricing Models with Irrelevant Risk Factors, *The Review of Financial Studies* 27, 2139–2170.

Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti, 2017, Spurious inference in reduced-rank asset-pricing models, *Econometrica* 85, 1613–1628.

Gourio, Francois, 2012, Disaster risk and business cycles, *American Economic Review* 102, 2734–66.

Gourio, Franǫois, Michael Siemer, and Adrien Verdelhan, 2013, International risk cycles, *Journal of International Economics* 89, 471 – 484.

Greenwood, Robin, and Andrei Shleifer, 2014, Expectations of Returns and Expected Returns, *The Review of Financial Studies* 27, 714–746.

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

Hansen, Lars Peter, 1982, Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica* 50, 1029–54.

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

Hansen, Lars Peter, 2007a, Beliefs, doubts and learning: Valuing macroeconomic risk, *American Economic Review* 97, 1–30.

Hansen, Lars Peter, 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, Lars Peter, 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, Lars Peter, 2014, Nobel lecture: Uncertainty outside and inside economic models, *Journal of Political Economy* 122, 945 – 987.

Hansen, Lars Peter, and James J. Heckman, 1996, The empirical foundations of calibration, *Journal of Economic Perspectives* 10, 87–104.

Hansen, Lars Peter, and Thomas J. Sargent, 1980, Formulating and estimating dynamic linear rational expectations models, *Journal of Economic Dynamics and Control* 2, 7–46.

Hansen, Lars Peter, and Thomas J. Sargent, 2001, Robust control and model uncertainty, *American Economic Review* 91, 60–66.

Hansen, Lars P., and Thomas J. Sargent, 2010, Fragile beliefs and the price of uncertainty, *Quantitative Economics* 1, 129–162.

Hansen, Lars Peter, and Kenneth J Singleton, 1982, Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models, *Econometrica* 50, 1269–86.

Hansen, Lars Peter, and Kenneth J Singleton, 1983, Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns, *Journal of Political Economy* 91, 249–65.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman, 2001, *The elements of statistical learning* . Springer Series in Statistics (Springer-Verlag, New York) Data mining, inference, and prediction.

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

Julliard, Christian, and Anisha Ghosh, 2012, Can rare events explain the equity premium puzzle?, *Review of Financial Studies* 25, 3037–3076.

Kan, Raymond, and Chu Zhang, 1999, Gmm tests of stochastic discount factor models with useless factors, *Journal of Financial Economics* 54, 103 – 127.

Kitamura, Yuichi, Taisuke Otsu, and Kirill Evdokimov, 2013, Robustness, infinitesimal neighborhoods, and moment restrictions, *Econometrica* 81, 1185–1201.

Kleibergen, Frank, 2009, Tests of risk premia in linear factor models, *Journal of Econometrics* 149, 149 – 173.

Kleibergen, Frank, and Zhaoguo Zhan, 2019, Robust inference for consumption-based asset pricing, *The Journal of Finance*.

Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji, 2005, A smooth model of decision making under ambiguity, *Econometrica* 73, 1849–1892.

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

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

Koijen, Ralph S. J., and Motohiro Yogo, 2019, A demand system approach to asset pricing, *Journal of Political Economy* 127, 1475–1515.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2019, Shrinking the cross-section, *Journal of Financial Economics*.

Kung, Howard, and Lukas Schmid, 2015, Innovation, growth, and asset prices, *The Journal of Finance* 70, 1001–1037.

Le Cam, Lucien, and Grace Lo Yang, 2000, *Asymptotics in statistics* . Springer Series in Statistics (Springer-Verlag, New York) second edn. Some basic concepts.

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

Lehmann, Erich L., 1999, *Elements of Large-Sample Theory* . Springer Texts in Statistics (Springer-Verlag, New York).

Lettau, Martin, Sydney Ludvigson, and Jessica Wachter, 2008, The declining equity premium: What role does macroeconomic risk play?, *Review of Financial Studies* 21, 1653–1687.

Lettau, Martin, and Stijn Van Nieuwerburgh, 2008, Reconciling the return predictability evidence, *Review of Financial Studies* 21, 1607–1652.

Levit, B. Ya, 1976, On the efficiency of a class of non-parametric estimates, *Theory of Probability & Its Applications* 20, 723–740.

Lewellen, Jonathan, Stefan Nagel, and Jay Shanken, 2010, A skeptical appraisal of asset pricing tests, *Journal of Financial Economics* 96, 175 – 194.

Lewis, Karen K., 2008, Peso problem, in Steven N. Durlauf, and Lawrence E. Blume, ed.: *The New*

Palgrave Dictionary of Economics (Palgrave Macmillan: Basingstoke).

Li, Di, Lucian A. Taylor, and Wenyu Wang, 2018, Inefficiencies and externalities from opportunistic acquirers, *Journal of Financial Economics* 130, 265–290.

Li, Hong, and Ulrich K. Müller, 2009, Valid inference in partially unstable generalized method of moments models, *Review of Economic Studies* 76, 343–365.

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

McManus, Douglas A., 1991, Who invented local power analysis?, *Econometric Theory* 7, 265–268.

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

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

Müller, Ulrich K., and Mark W. Watson, 2016, Measuring Uncertainty about Long-Run Predictions, *The Review of Economic Studies* 83, 1711–1740.

Nagel, Stefan, and Kenneth J. Singleton, 2011, Estimation and evaluation of conditional asset pricing models, *The Journal of Finance* 66, 873–909.

Nagel, Stefan, and Zhengyang Xu, 2019, Asset Pricing with Fading Memory, Discussion paper.

Nakamura, Emi, Jón Steinsson, Robert Barro, and José 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, Whitney K., 1985a, Generalized method of moments specification testing, *Journal of Econometrics* 29, 229–256.

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

Newey, Whitney K., 1990, Efficient instrumental variables estimation of nonlinear models, *Econometrica* 58, 809–837.

Newey, Whitney 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, Whitney K., and Daniel McFadden, 1994, Chapter 36 large sample estimation and hypothesis testing, , vol. 4 of *Handbook of Econometrics* . pp. 2111 – 2245 (Elsevier).

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

Neyman, Jerzy, 1937, "smooth" tests for goodness of fit, *Scandinavian Actuarial Journal* 1937, 149–199.

Nyblom, Jukka, 1989, Testing for the constancy of parameters over time, *Journal of the American Statistical Association* 84, 223–230.

Ogaki, Masao, 1993, 17 generalized method of moments: Econometric applications, in *Econometrics*, vol. 11 of *Handbook of Statistics* . pp. 455 – 488 (Elsevier).

Pastor, Lubos, and Robert Stambaugh, 2001, The equity premium and structural breaks, *Journal of Finance* 56, 1207–1239.

Pesaran, M. Hashem, and Allan 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.

Rietz, Thomas A., 1988, The equity premium: A solution, *Journal of Monetary Economics* 22, 117–131.

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

Saracoglu, Rusdu, and Thomas J. Sargent, 1978, Seasonality and portfolio balance under rational expectations, *Journal of Monetary Economics* 4, 435–458.

Schorfheide, Frank, Dongho Song, and Amir Yaron, 2018, Identifying long-run risks: A bayesian mixed-frequency approach, *Econometrica* 86, 617–654.

Schorfheide, Frank, and Kenneth I. Wolpin, 2012, On the use of holdout samples for model selection, *The American Economic Review* 102, 477–481.

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

Seo, Sang Byung, and Jessica A. Wachter, 2018, Do rare events explain cdx tranche spreads?, *The Journal of Finance* 73, 2343–2383.

Seo, Sang Byung, and Jessica A. Wachter, 2019, Option prices in a model with stochastic disaster risk, *Management Science* 65, 3449–3469.

Severini, Thomas A., and Gautam Tripathi, 2013, Semiparametric efficiency bounds for microeconometric models: A survey, *Foundations and Trends in Econometrics* 6, 163–397.

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

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

Spiegelhalter, David J., Nicola G. Best, Bradley P. Carlin, and Angelika van der Linde, 2002, Bayesian measures of model complexity and fit, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 64, 583–639.

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

Stock, James H., and Jonathan Wright, 2000, GMM with Weak Identification, *Econometrica* 68, 1055–1096.

Stock, James H, Jonathan H Wright, and Motohiro Yogo, 2002, A survey of weak instruments and weak identification in generalized method of moments, *Journal of Business & Economic Statistics* 20, 518–529.

Sun, Yin, Árpád Baricz, and Shidong 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, Robert, and Keith Knight, 1999, The covariance inflation criterion for adaptive model selection, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61, 529–546.

Valkanov, Rossen, 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).

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

Welch, Ivo, and Amit Goyal, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

White, Halbert, and Ian Domowitz, 1984, Nonlinear Regression with Dependent Observations, *Econometrica* 52, 143–61.

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

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

Appendix: Technical and Regularity Assumptions

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

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

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

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

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

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

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

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

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

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

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

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

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

The following are regularity conditions on moments.

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

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

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

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

Conditions (i) and (ii) of Assumption 5 are needed to establish the functional central limit theorem (invariance principle) of McLeish (1975) 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 5 in Online Appendix 2.2 and Theorem 1 in Online Appendix 3).

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

$$\mathcal{G}_B(\mathrm{Q}_0) \equiv \left\{ g \in \mathcal{G}(\mathrm{Q}_0) : \lambda^{(1)}(g_1) = 0 \text{ and } \lambda^{(1)}(g_2) \in \mathsf{lin}(D_{11}) \right\}. \tag{98}$$

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

Assumption 6 ensures that the baseline structural model is correctly specified because $\lambda^{(1)}(f_{n,t}) = \lambda^{(1)}(g_2)b(t/n) \in \mathsf{lin}(D_{11})$ for every $t \in \{1, \cdots, n\}$. We can replace (98) with a seemingly weaker assumption $\lambda^{(1)}(g_1^*), \lambda^{(1)}(g_2^*) \in \mathsf{lin}(D_{11})$. However, this does not add generality because we can always replace θ_0 with a sequence of new reference points (reparametrization) to ensure that (98) is satisfied.