Measuring the “Dark Matter” in Asset Pricing Models

Hui Chen Winston Wei Dou Leonid Kogan*

April 16, 2017

Abstract

We propose a new quantitative measure of model fragility, based on the tendency of a model to over-fit the data in sample. Structural economic models are fragile when the cross-equation restrictions they impose on the baseline model appear excessively informative about model parameters that are otherwise difficult to estimate. Our measure is analytically tractable and helps identify main sources of model fragility. As an application, we diagnose fragility in asset pricing models with rare disasters and long-run consumption risk.

JEL Codes: C1, D83, E44, G12

Keywords: Cross-equation restrictions; Disaster risk; Fragile beliefs; Information theory; Long-run risk; Model fragility; Over-fitting; Robustness; Sensitivity analysis

---

*Chen: MIT Sloan and NBER (huichen@mit.edu). Dou: Wharton (wdou@wharton.upenn.edu). Kogan: MIT Sloan and NBER (lkogan@mit.edu). We thank Alon Brav, Bradyn Breon-Drish, Murray Carlson, Xu Cheng, John Cochrane, George Constantinides, Ian Dew-Becker, Frank Diebold, Lars Hansen, Karen Lewis, Christian Opp, Tom Sargent, Frank Schorfheide, Martin Schneider, Luke Taylor, Laura Veldkamp, Jessica Wachter, Tan Wang, S. Viswanathan, Fernando Zapatero, Stan Zin, and seminar participants at Chicago, CITE, CKGSB, INSEAD, ITAM Finance Conference, Macrofinance Workshop, MIT Sloan, Northwestern, Harvard, NYU, NBER Asset Pricing and Capital Markets Meetings, Red Rock Finance Conference, UBC Winter Finance Conference, UPenn Economics, Utah Winter Finance Conference, WFA, WashU, Wharton, and Yale SOM for comments, and Ye Luo for his comments on the proofs. All errors are our own.

1 Introduction

When building and evaluating a quantitative economic model, we care about how the model performs out of sample in addition to how well it fits the data in sample. Too much emphasis on in-sample fit favors models with excessive degrees of freedom. The common concern is that such models are fragile, meaning that they tend to over-fit the data in-sample, at the expense of poor out-of-sample performance. Such tendency towards in-sample over-fitting also makes it difficult to reject potentially mis-specified fragile models in standard statistical tests. In this paper, we develop a quantitative measure of model fragility.

Our fragility measure is a model property. It evaluates a structural model relative to a corresponding baseline sub-model. Specifically, we start with a baseline model describing the dynamics of variables $\mathbf{x}_t$ using the parameter vector θ. The structural model then describes the joint dynamics of variables $\mathbf{x}$ and potentially additional variables $\mathbf{y}$, and typically adds more parameters. In economic applications, the structural model would typically add cross-equation restrictions implied by economic theory.

As an example, consider a representative-agent endowment economy, with the baseline model describing the joint dynamics of the endowment and dividends of financial assets ($\mathbf{x}$), parameterized by θ. A structural model adds assumptions on investors' behavior, which put restrictions on the joint dynamics of the endowment, dividends, and also asset prices ($\mathbf{y}$).

Our fragility measure compares the asymptotic precision of inference about θ using the structural model relative to using the baseline model. This isolates the extra information provided by the structural model restrictions. In the context of our example, such restrictions connect asset prices to the dividends and the aggregate endowment. Mathematically, our fragility measure compares the inverse GMM Fisher information matrices for the baseline model and the full structural model. We therefore refer to it as the Fisher fragility measure.

Our Fisher fragility measure is connected to a standard measure of statistical over-fitting tendency, due to Spiegelhalter, Best, Carlin, and van der Linde (2002) (henceforth referred to as SBCL), which we extend to our setting. Specifically, we assume that the distributional properties of the model are summarized by a set of moments. For any value of the parameter vector θ, we use the GMM J-distance, $\widehat{J}(\theta; \mathbf{x}, \mathbf{y})$, as the measure of model's in-sample fit given the observations $(\mathbf{x}, \mathbf{y})$ for certain parameter value θ. Denoting $\hat{\theta}(\mathbf{x}, \mathbf{y})$ as the GMM estimate of the parameter vector and θ_0 as the true parameter value, we then extend the definition of SBCL and use the gap between the J-distances of the true model and the fitted model, $\widehat{J}(\theta_0; \mathbf{x}, \mathbf{y}) - \widehat{J}(\hat{\theta}(\mathbf{x}, \mathbf{y}); \mathbf{x}, \mathbf{y})$, as a measure of the degree of over-fitting by the estimated model. $\widehat{J}(\theta_0; \mathbf{x}, \mathbf{y})$ cannot be lower than the fitted value $\widehat{J}(\hat{\theta}(\mathbf{x}, \mathbf{y}); \mathbf{x}, \mathbf{y})$, which is chosen to minimize the J-distance in sample. In comparison to the original formulation in SBCL, our over-fitting measure focuses on a particular set of moment conditions, without requiring a full specification of the likelihood function of the structural model.

Feasible implementations of the over-fitting measure prescribe averaging the degree of over-fitting over a set of possible true models. The exact specification of the alternative model weighting must be chosen depending on the context, and determines the interpretation and the scope of usefulness of the over-fitting tendency measure. In our analysis, we use the posterior distribution for θ implied by the baseline model, $\pi(\theta|\mathbf{x})$, as the distribution over the alternative models. With this choice, we measure the over-fitting tendency of the cross-equation restrictions added by the structural model, treating the baseline model as reliable.

As we show in Theorem 1 below, the Fisher fragility measure is equal to the mean of the asymptotic distribution of the over-fitting tendency measure. We show how to decompose our Fisher fragility measure as the sum of one-dimensional fragility measures, each defined along one of the mutually orthogonal one-dimensional linear subspaces of the parameter space. Each one-dimensional subspace corresponds to a particular linear combination of parameters, $v\theta$. When restricted to such a subspace, the fragility measure corresponds to the amount of extra data needed to lower the asymptotic variance of the estimator of $v\theta$ under the baseline model to the level of asymptotic variance under the full structural model (without extra data). Moreover, we show that the value of the highest one-dimensional fragility measure captures the likelihood of instances of extreme over-fitting by the structural model.

Our assessment of model fragility can be viewed as a formal version of sensitivity analysis. Intuitively, a model is considered fragile if its key implications are excessively sensitive to small perturbations of model parameters. In practical applications, one must specify the relevant perturbations and define "excessive sensitivity." These problems are exacerbated in multivariate settings, where to assess the full extent of model fragility one must consider simultaneous perturbations of multiple parameters. Our Fisher fragility measure eliminates the need for ad hoc choices by using the precision of inference for θ from the baseline model to effectively define the relevant parameter perturbations, and using the relative precision of inference for θ under the structural model to define excessive sensitivity.

If a model is correctly specified, high sensitivity of model implications to parameters implies high precision of parameter estimation. However, fragile models tend to generate excessively high quality of in-sample fit, and hence are difficult to reject even when they are mis-specified and expected to perform poorly out of sample. The main purpose of our fragility measure is to facilitate structural model selection when there are multiple candidate models that fit a common set of observations well in sample. Our measure is thus related to the penalties for model complexity used by the statistical model selection procedures, e.g., AIC, BIC, and LASSO. As we discuss in the next section, our measure differs from standard measures of model complexity because of its focus on model sensitivity to parameter perturbations.

We evaluate the Fisher fragility measure for two models from the asset pricing literature. The first example is a rare-disaster model. In this model, parameters describing the likelihood and the magnitude of economic disasters are relatively difficult to estimate from the data unless one uses information in asset prices.[1] We derive

[1]In his 2008 Princeton Finance Lectures, John Campbell suggests that variable risk of rare

the Fisher fragility measure in this example analytically. We also illustrate how to incorporate uncertainty about the structural parameters (preference parameters in this context) when computing model fragility. The second example is a long-run risk model with a high-dimensional parameter space. We use this example to illustrate how to systematically diagnose the sources of fragility in a complex model.

1.1 Related Literature

The idea that model fragility is connected to model's degrees of freedom dates back at least to Fisher (1922). Traditionally, effective degrees of freedom of a model are measured by the number of parameters, because of the coincidence of the two in Gaussian-linear models (see, e.g. Ye, 1998; Efron, 2004). Numerous statistical model selection procedures are based on this idea.[2]

The limitations of using the number of parameters to measure model's degrees of freedom are well known. Extant literature covers several alternative approaches to measuring the "implicit degrees of freedom." Ye (1998), Shen and Ye (2002), and Efron (2004) propose to measure the so-called "generalized degrees of freedom" for Gaussian-linear models using the sensitivity of fitted values with respect to the observed data. Gentzkow and Shapiro (2013) apply a similar idea to examine identification issues in complex structural models. Spiegelhalter, Best, Carlin, and van der Linde (2002), Ando (2007) and Gelman, Hwang, and Vehtari (2013), among others, propose a Bayesian complexity measure they call "the effective number of parameters," which is based on out-of-sample model performance. These methods measure the sensitivity of model implications to parameter perturbations. The important common feature of these proposals is that they rely on the same model being evaluated to determine the magnitude of necessary parameter perturbations. This is potentially problematic when evaluating economic models that are fragile according to our definition. For

disasters might be the "dark matter for economists."

[2]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).

such models, the posterior distribution over the parameters is highly concentrated as a result of excessive model sensitivity to its parameters. Relying on this posterior to generate parameter perturbations can under-represent the true extent of model fragility. In contrast, we propose to use a baseline model to assign the weights to potential alternative models.

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

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

Our analysis of the disaster-risk model relates to several papers that have pointed out the challenges in testing such models. One implication of the low probability of disasters is the so-called "Peso problem" (see Lewis, 2008, for an overview): if observations of disasters in a particular sample under-represent their population distribution, standard inference procedures may lead to distorted conclusions. Thus, the peso problem is a particular case of the weak identification problem. Our analysis highlights that in applications subject to a 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 empirical evidence while being difficult to reject in the data. Our analysis of model fragility encapsulates such considerations in a general quantitative measure.

2 Measuring Model Fragility

In this section, we first introduce our measure of model fragility, and then derive its properties.

2.1 A Generic Model Structure

Consider a baseline model $\mathcal{P}$, which is a part of the full structural model $\mathcal{Q}$. The baseline model $\mathcal{P}$ specifies the dynamics of a vector of variables $\mathbf{x_t}$ with the underlying distribution $\mathbb{P}$. The full structural model $\mathcal{Q}$ aims to capture certain features of the distribution $\mathbb{Q}$ that govern the dynamics of $\mathbf{x_t}$ and potentially additional variables $\mathbf{y_t}$.

More precisely, the baseline model $\mathcal{P}$ of $\mathbf{x^n} \equiv (\mathbf{x_1}, \cdots, \mathbf{x_n})$ can be specified by a $D_\Theta \times 1$ parameter vector θ, while the full structural model $\mathcal{Q}$ may incorporate additional parameters ψ and describe the properties of extra data $\mathbf{y^n} \equiv (\mathbf{y_1}, \cdots, \mathbf{y_n})$. The distribution of $\mathbf{y^n}$ conditional on $\mathbf{x^n}$ depends on not only the baseline parameter vector θ, but also potentially the $D_\Psi \times 1$ nuisance parameter vector ψ. We assume the baseline parameters θ are identified by $\mathcal{P}$. The nuisance parameters ψ by definition are not part of the baseline model but should be accounted for when assessing the full model. We assume that the true parameter values θ_0 and ψ_0 are contained in the interiors of the feasible sets Θ and Ψ, respectively.

We assume that the stochastic process $\{\mathbf{x_t}\}$ is strictly stationary and ergodic with a stationary distribution $\mathbb{P}$. The true joint distribution for $\mathbf{x^n}$ is $\mathbb{P}_n$. Similarly, we assume that the joint stochastic process $\{\mathbf{x_t}, \mathbf{y_t}\}$ is strictly stationary and ergodic with a stationary distribution $\mathbb{Q}$. The econometrician does not need to specify the full functional form of the joint distribution of $(\mathbf{x^n}, \mathbf{y^n})$, denoted by $\mathbb{Q}_n$. The unknown joint density is $q(\mathbf{x^n}, \mathbf{y^n})$.

We evaluate the performance of a structural model under the Generalized Method of Moments (GMM) framework. Specifically, we assume that the model builder is concerned with the model's in-sample and out-of-sample performance as represented by a set of moment conditions,[3] based on a $D_{\mathcal{Q}} \times 1$ vector of functions $g_{\mathcal{Q}}(\theta, \psi; \mathbf{x}_t, \mathbf{y}_t)$ of the data observations $(\mathbf{x}_t, \mathbf{y}_t)$ and the parameters θ and ψ satisfying the following restrictions:

$$\mathbb{E}\left[g_{\mathcal{Q}}(\theta_0, \psi_0; \mathbf{x_t}, \mathbf{y_t})\right] = 0, \tag{1}$$

where the expectation is taken under the true probability measure $\mathbb{Q}$.

The baseline moment functions $g_{\mathcal{P}}(\theta; \mathbf{x_t})$ characterize the moment conditions of the baseline model. They constitute the first $D_{\mathcal{P}}$ elements of the larger vector of moment functions $g_{\mathcal{Q}}(\theta, \psi; \mathbf{x_t}, \mathbf{y_t})$. Thus, the baseline moments can be represented by

$$g_{\mathcal{P}}(\theta; \mathbf{x_t}) = \Gamma_{\mathcal{P}} g_{\mathcal{Q}}(\theta, \psi; \mathbf{x_t}, \mathbf{y_t}), \quad \text{where} \quad \Gamma_{\mathcal{P}} \equiv \left[I_{D_{\mathcal{P}}}, O_{D_{\mathcal{P}} \times (D_{\mathcal{Q}} - D_{\mathcal{P}})}\right]. \tag{2}$$

The moment functions $g_{\mathcal{P}}(\theta; \mathbf{x_t})$ depend only on parameters θ, since all parameters of the baseline model are included in θ. Accordingly, the moment conditions for the baseline model are

$$\mathbb{E}\left[g_{\mathcal{P}}(\theta_0; \mathbf{x_t})\right] = 0. \tag{3}$$

We denote the empirical moment conditions for the full model and the baseline model

[3] We can also adopt the CUE method of Hansen, Heaton, and Yaron (1996) or its modification Hausman, Lewis, Menzel, and Newey (2011)'s RCUE method, or some other extensions of GMM with the same first-order efficiency and possibly superior higher-order asymptotic properties. This will lead to alternative but conceptually similar measures of over-fitting. To simplify the comparison with the Fisher fragility measure, we choose to use the original GMM framework.

by

$$\widehat{g}_{\mathcal{Q},n}(\theta,\psi) \equiv \frac{1}{n}\sum_{t=1}^{n} g_{\mathcal{Q}}(\theta,\psi;\mathbf{x_t},\mathbf{y_t}) \text{ and } \widehat{g}_{\mathcal{P},n}(\theta) \equiv \frac{1}{n}\sum_{t=1}^{n} g_{\mathcal{P}}(\theta;\mathbf{x_t}),$$

respectively. Then, the optimal GMM estimator $(\hat{\theta}^{\mathcal{Q}}, \hat{\psi}^{\mathcal{Q}})$ of the full model and that of the baseline model $\hat{\theta}^{\mathcal{P}}$ minimize, respectively,

$$\widehat{J}_{n,S_{\mathcal{Q}}}(\theta,\psi) \equiv n\widehat{g}_{\mathcal{Q},n}(\theta,\psi)^T S_{\mathcal{Q}}^{-1}\widehat{g}_{\mathcal{Q},n}(\theta,\psi) \quad \text{and} \quad \widehat{J}_{n,S_{\mathcal{P}}}(\theta) \equiv n\widehat{g}_{\mathcal{P},n}(\theta)^T S_{\mathcal{P}}^{-1}\widehat{g}_{\mathcal{P},n}(\theta). \tag{4}$$

Here, $\widehat{J}_{n,S_{\mathcal{Q}}}(\theta,\psi)$ and $\widehat{J}_{n,S_{\mathcal{P}}}(\theta)$ are J-distances for gauging in-sample fit of certain parameter values;[4] $S_{\mathcal{Q}}$ and $S_{\mathcal{P}}$ have the following formulae (see Hansen, 1982),

$$S_{\mathcal{Q}} \equiv \sum_{\ell=-\infty}^{+\infty} \mathbb{E}\left[g_{\mathcal{Q}}(\theta_0,\psi_0;\mathbf{x_t},\mathbf{y_t})g_{\mathcal{Q}}(\theta_0,\psi_0;\mathbf{x_{t-\ell}},\mathbf{y_{t-\ell}})^T\right], \quad \text{and} \tag{5}$$

$$S_{\mathcal{P}} \equiv \sum_{\ell=-\infty}^{+\infty} \mathbb{E}\left[g_{\mathcal{P}}(\theta_0;\mathbf{x_t})g_{\mathcal{P}}(\theta_0;\mathbf{x_{t-\ell}})^T\right]. \tag{6}$$

That is, $S_{\mathcal{Q}}$ and $S_{\mathcal{P}}$ are the covariance matrices of the moment conditions at the true parameter values. In practice, when $S_{\mathcal{Q}}$ and $S_{\mathcal{P}}$ are unknown, we can replace them with consistent estimators $\widehat{S}_{\mathcal{Q},n}$ and $\widehat{S}_{\mathcal{P},n}$, respectively. The consistent estimators of the covariance matrices are provided by Newey and West (1987a), Andrews (1991), and Andrews and Monahan (1992).

We use GMM instead of likelihood to evaluate model performance due to the concern of likelihood mis-specification. The GMM approach gives model builder the flexibility to choose which aspects of the model to emphasize when estimating parameters and evaluating model specifications. This is in contrast to the likelihood approach, which relies on the full probability distribution implied by the structural model.

Finally, we introduce GMM Fisher information matrices. We denote the GMM Fisher information matrix for the baseline model as $\mathbf{I}_{\mathcal{P}}(\theta)$ (see Hansen, 1982; Hahn,

[4] When efficient GMM estimators are plugged in, $\widehat{J}_{n,S_{\mathcal{Q}}}(\hat{\theta}^{\mathcal{Q}},\hat{\psi}^{\mathcal{Q}})$ and $\widehat{J}_{n,S_{\mathcal{P}}}(\hat{\theta}^{\mathcal{P}})$ are commonly referred to as the Hansen's J-statistics in the GMM literature.

Newey, and Smith, 2011),

$$\mathbf{I}_{\mathcal{P}}(\theta) \equiv G_{\mathcal{P}}(\theta)^T S_{\mathcal{P}}^{-1} G_{\mathcal{P}}(\theta), \tag{7}$$

where $G_{\mathcal{P}}(\theta) \equiv \mathbb{E}\left[\nabla g_{\mathcal{P}}(\theta; \mathbf{x_t})\right]$. Below we denote $G_{\mathcal{P}} \equiv G_{\mathcal{P}}(\theta_0)$ for brevity. Similarly, we denote the GMM Fisher information matrix for the structural model as $\mathbf{I}_{\mathcal{Q}}(\theta, \psi)$,

$$\mathbf{I}_{\mathcal{Q}}(\theta, \psi) \equiv G_{\mathcal{Q}}(\theta, \psi)^T S_{\mathcal{Q}}^{-1} G_{\mathcal{Q}}(\theta, \psi), \tag{8}$$

where $G_{\mathcal{Q}}(\theta, \psi) \equiv \mathbb{E}\left[\nabla g_{\mathcal{Q}}(\theta, \psi; \mathbf{x_t}, \mathbf{y_t})\right]$, and we denote $G_{\mathcal{Q}} \equiv G_{\mathcal{Q}}(\theta_0, \psi_0)$. Computing the expectations $G_{\mathcal{P}}(\theta)$ and $G_{\mathcal{Q}}(\theta, \psi)$ requires knowing the distribution $\mathbb{Q}$. When $\mathbb{Q}$ is unknown, $G_{\mathcal{P}}(\theta)$ and $G_{\mathcal{Q}}(\theta, \psi)$ can be replaced by their consistent estimators.

For the full model $\mathcal{Q}$, we will focus on the (marginal) GMM Fisher information matrix for θ, $\mathbf{I}_{\mathcal{Q}}(\theta|\psi)$, with uncertainty in the nuisance parameters ψ explicitly accounted for. It is defined as follows:[5]

$$\mathbf{I}_{\mathcal{Q}}(\theta|\psi) \equiv \left[\Gamma_\Theta \mathbf{I}_{\mathcal{Q}}(\theta, \psi)^{-1} \Gamma_\Theta^T\right]^{-1}, \quad \text{where } \Gamma_\Theta \equiv \left[I_{D_\Theta}, O_{D_\Theta \times D_\Psi}\right]. \tag{9}$$

In the definition of the GMM Fisher information matrices, the optimal weighting matrices $S_{\mathcal{P}}^{-1}$ and $S_{\mathcal{Q}}^{-1}$ are used. The objective of our analysis is to measure the

[5] More precisely, the joint Fisher information matrix $\mathbf{I}_{\mathcal{Q}}(\theta, \psi)$ can be partitioned into a two-by-two block matrix according to θ and ψ:

$$\mathbf{I}_{\mathcal{Q}}(\theta, \psi) = \begin{bmatrix} \mathbf{I}_{\mathcal{Q}}^{(1,1)}(\theta, \psi), & \mathbf{I}_{\mathcal{Q}}^{(1,2)}(\theta, \psi) \\ \mathbf{I}_{\mathcal{Q}}^{(2,1)}(\theta, \psi), & \mathbf{I}_{\mathcal{Q}}^{(2,2)}(\theta, \psi) \end{bmatrix},$$

where $\mathbf{I}_{\mathcal{Q}}^{(1,1)}(\theta, \psi)$ is the $D_\Theta \times D_\Theta$ conditional information matrix corresponding to the baseline parameters θ with ψ fixed, $\mathbf{I}_{\mathcal{Q}}^{(2,2)}(\theta, \psi)$ is the $D_\Psi \times D_\Psi$ conditional information matrix corresponding to the nuisance parameters ψ with θ fixed, and $\mathbf{I}_{\mathcal{Q}}^{(1,2)}(\theta, \psi) = \mathbf{I}_{\mathcal{Q}}^{(2,1)}(\theta, \psi)^T$ is the $D_\Theta \times D_\Psi$ cross-information matrix corresponding to θ and ψ. Then $\mathbf{I}_{\mathcal{Q}}(\theta|\psi)$ can be written as

$$\mathbf{I}_{\mathcal{Q}}(\theta|\psi) = \mathbf{I}_{\mathcal{Q}}^{(1,1)}(\theta, \psi) - \mathbf{I}_{\mathcal{Q}}^{(1,2)}(\theta, \psi) \mathbf{I}_{\mathcal{Q}}^{(2,2)}(\theta, \psi)^{-1} \mathbf{I}_{\mathcal{Q}}^{(1,2)}(\theta, \psi)^T,$$

which generally is not equal to the Fisher information sub-matrix $\mathbf{I}_{\mathcal{Q}}^{(1,1)}(\theta, \psi)$ for baseline parameters θ, except in the special case in which $\mathbf{I}_{\mathcal{Q}}^{(1,2)}(\theta, \psi) = 0$. In general, $\mathbf{I}_{\mathcal{Q}}^{(1,2)}(\theta, \psi) \mathbf{I}_{\mathcal{Q}}^{(2,2)}(\theta, \psi)^{-1} \mathbf{I}_{\mathcal{Q}}^{(1,2)}(\theta, \psi)^T$ is the information loss about θ due to the uncertainty in the nuisance parameters ψ. We assume that the information matrices are nonsingular in this paper (Assumption A4 in Appendix A).

informativeness of the cross-equation restrictions from the structural model, which is a property of the model, and as we will show later, is connected to the model's over-fitting tendency. The optimal weighting matrices are the natural choice for this purpose.[6] We should note that in other contexts, e.g., when measuring model misspecification as in Hansen and Jagannathan (1997), it is appropriate to use alternative weighting matrices.

2.2 Model Fragility

We now introduce our measure of model fragility.

Definition 1 (Fisher Fragility Measure)**.** *The Fisher fragility measure corresponding to a full-rank $D_{\mathbf{v}} \times D_{\Theta}$ matrix $\mathbf{v}$ is defined as*

$$\varrho^{\mathbf{v}}(\theta_0|\psi_0) \equiv \mathbf{tr}\left[\left(\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right)^{-1}\left(\mathbf{v}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}^T\right)\right], \tag{10}$$

where $\mathbf{I}_{\mathcal{P}}(\theta_0)$ and $\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)$ are the Fisher information matrices defined in (7) and (9).

The Fisher fragility measure effectively compares the asymptotic covariance matrices of the two estimators of θ: the one based on the baseline model, and the one based on the structural model. This isolates the information provided by the structural model restrictions (relative to the baseline model).

In the special case where $\mathbf{v}$ is a full-rank $D_{\Theta} \times D_{\Theta}$ matrix, $\varrho^{\mathbf{v}}(\theta_0|\psi_0)$ is independent of the choice of $\mathbf{v}$. In that case we denote the Fisher fragility measure as $\varrho(\theta_0|\psi_0)$, which is the overall Fisher fragility measure. In another special case where $D_{\mathbf{v}} = 1$, our measure equals to the ratio of asymptotic variances of the two GMM estimators $\mathbf{v}\hat{\theta}^{\mathcal{Q}}$ and $\mathbf{v}\hat{\theta}^{\mathcal{P}}$. From Definition 1, it immediately follows that the Fisher fragility measure can be characterized by the solution of an eigenvalue problem. The result is summarized by Proposition 1, with the proof given in Appendix C.1.

[6] Following Hansen (1982), further analysis justifies the unique role of the optimal weighting matrix in formulating the J-test statistic. As emphasized by Newey and West (1987b), one must define the GMM likelihood ratio test statistics based on the optimal GMM estimator to guarantee the standard asymptotic properties.

Proposition 1. *For a full-rank $D_{\mathbf{v}} \times D_\Theta$ matrix $\mathbf{v}$, let $\lambda_1(\mathbf{v}) \geq \lambda_2(\mathbf{v}) \geq \cdots \geq \lambda_{D_{\mathbf{v}}}(\mathbf{v})$ be the eigenvalues of a Fisher information ratio matrix $\Pi_0(\mathbf{v})$ defined as follows*

$$\Pi_0(\mathbf{v}) \equiv \left(\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right)^{-1/2} \left(\mathbf{v}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}^T\right) \left(\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right)^{-1/2}. \tag{11}$$

Then the Fisher fragility measure is the sum of the eigenvalues,

$$\varrho^{\mathbf{v}}(\theta_0|\psi_0) = \lambda_1(\mathbf{v}) + \lambda_2(\mathbf{v}) + \cdots + \lambda_{D_{\mathbf{v}}}(\mathbf{v}). \tag{12}$$

Furthermore, $\varrho^{\mathbf{v}}(\theta_0|\psi_0)$ is no less than $D_{\mathbf{v}}$ because the smallest eigenvalue $\lambda_{D_{\mathbf{v}}}(\mathbf{v})$ is no less than one.

For notational simplicity, we suppress the dependence of the eigenvalues $\lambda_i(\mathbf{v})$ on the parameters (θ_0, ψ_0). That the Fisher fragility measure $\varrho^{\mathbf{v}}(\theta_0|\psi_0)$ is no less than $D_{\mathbf{v}}$ reflects the fact that it captures the implicit degrees of freedom for a model, beyond the number of parameters (the dimension of the parameter subspace) under consideration.

The measure $\varrho^{\mathbf{v}}(\theta_0|\psi_0)$ is defined for specific feature directions $\mathbf{v}$ in the space of baseline parameters. In applications, one might be interested in searching among a class of feature directions to find the directions with the largest discrepancy between the two GMM Fisher information matrices. This leads us to define the following worst-case Fisher fragility measure.

Definition 2. *The worst-case Fisher fragility measure for the class of D-dimensional feature functions ($D \leq D_\Theta$) is defined as:*

$$\varrho^{D}(\theta_0|\psi_0) = \max_{\mathbf{v} \in \mathbb{R}^{D \times D_\Theta}, \mathbf{Rank}(\mathbf{v}) = D} \mathbf{tr}\left[\left(\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right)^{-1} \left(\mathbf{v}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}^T\right)\right]. \tag{13}$$

The problem in (13) is a generalized eigenvalue problem. The following proposition summarizes its solution. The proof of Proposition 2 is in Appendix C.2.

Proposition 2. *Let $\lambda_i \equiv \lambda_i(I_{D_\Theta})$, with $1 \leq i \leq D_\Theta$, be the eigenvalues of Fisher* *information ratio matrix* $\Pi_0(I_{D_\Theta}) = \mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{\frac{1}{2}}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{\frac{1}{2}}$*, with the corresponding* $D_\Theta \times 1$ *eigenvectors* $e_1, e_2, \cdots, e_{D_\Theta}$*. Suppose* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{D_\Theta}$*, then the* D*-dimensional worst-case Fisher fragility measure is equal to*

$$\varrho^D(\theta_0|\psi_0) = \lambda_1 + \lambda_2 + \cdots + \lambda_D, \tag{14}$$

with the worst-case D*-dimensional linear subspace of the parameter space characterized by the matrix* $\mathbf{v}_D^* = [v_1^* \ v_2^* \ \cdots \ v_D^*]^T$*,*

$$v_i^* = \mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{\frac{1}{2}} e_i / \left| \mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{\frac{1}{2}} e_i \right|. \tag{15}$$

As a special case, the overall Fisher fragility measure is given by

$$\varrho(\theta_0|\psi_0) = \lambda_1 + \lambda_2 + \cdots + \lambda_{D_\Theta}. \tag{16}$$

From Proposition 2, it is easy to see that the worst-case Fisher fragility measure $\varrho^D(\theta_0|\psi_0)$ is monotonically increasing in the dimension of the subspace D. The monotonicity property is summarized in the following proposition, whose proof can be found in Appendix C.3.

Proposition 3. (Monotonicity) *For* $D_1 \leq D_2 \leq D_\Theta$*,*

$$\varrho^{D_1}(\theta_0|\psi_0) \leq \varrho^{D_2}(\theta_0|\psi_0). \tag{17}$$

We can view Proposition 2 as a decomposition of the overall fragility of a model into D_Θ linear subspaces of 1 dimension. The i-th largest eigenvalue λ_i $(1 \leq i \leq D_\Theta)$ of $\Pi_0(I_{D_\Theta})$ gives the marginal contribution of the 1-dimensional linear subspace associated with v_i^* to the overall fragility measure. In the language of sensitivity analysis, such a decomposition reveals the directions along which small perturbations of parameters can have the largest impact on the model output. Moreover, the decomposition also helps capture the tendency of receiving extreme over-fitting outcomes for a structural model (see Corollary 2 below).

The Fisher fragility measure has a natural "effective-sample-size" interpretation. Consider the case of $D = 1$. In this case, we ask what is the minimum sample size required for the estimator of the baseline model to match or exceed the asymptotic precision (the inverse of the variance) of the estimator for the full structural model in all one-dimensional linear subspaces of the parameter space. Because asymptotic variance scales inversely with the sample size, the required effective sample size is $\varrho^1(\theta_0|\psi_0)\, n$.[7] The effective sample size concept highlights, in the one-dimensional case, a direct connection between our measure of model fragility and the degree of informativeness of cross-equation restrictions.[8]

Our Fisher fragility measure isolates the information provided by the structural model relative to the baseline model. Thus, for different choices of the baseline model, the Fisher fragility measures for the same structural model will be interpreted differently. So far, we have been silent about how the baseline model should be chosen in relation to the full structural model. In general, there is no hard rule for this choice, other than the technical requirement that the associated baseline parameters θ be identified by the baseline model. Desirable choices of the baseline model will depend on which aspects of the model one intends to capture in the fragility analysis.

3 Model Over-Fitting Tendency

The Fisher fragility measure defined in Section 2 is an information-theoretic concept. In this section, we establish a formal connection between this measure and a model's

[7] We formalize this idea in finite samples in an information-theoretic framework in the online appendix Chen, Dou, and Kogan (2017). Both the paper (with the accompanying appendices) and the online appendix are self-contained.

[8] When evaluating competing models that are not rejected in statistical tests, our fragility measure highlights which models are more prone to in-sample over-fitting, so that statistical tests have low power against such models. These are the structural models in which the additional restrictions, relative to the baseline model, are relatively informative about model parameters. Note that the additional moment conditions that are relatively uninformative about θ (or even fully redundant, as in Breusch, Qian, Schmidt, and Wyhowski, 1999) may still help reject the model in statistical tests. In such cases, the concerns about lower power of tests and in-sample over-fitting do not apply.

over-fitting tendency, thus justifying the fragility interpretation. We first introduce an econometric measure of over-fitting, which extends a popular existing statistical over-fitting tendency measure to our structural setting (see SBCL). We then show that our Fisher fragility measure captures the average asymptotic over-fitting tendency of certain structural components of model.

3.1 Econometric Measure of Over-Fitting Tendency

Following SBCL, we introduce the measure of a model's over-fitting tendency in a Bayesian framework. Let $\pi(\theta)$ be a prior distribution on θ, and let $\pi_{\mathcal{P}}(\theta|\mathbf{x^n})$ be the posterior distribution of θ based on the likelihood of the baseline model $\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)$, which we assume to be fully specified.[9]

Definition 3 (Over-fitting Tendency Measure)**.** *We define the over-fitting tendency measure for the structural model $\mathcal{Q}$ relative to the baseline model $\mathcal{P}$ as*

$$\varrho_o(\theta_0|\psi_0, \mathbf{x^n}, \mathbf{y^n}) \equiv \int d_{S_{\mathcal{Q}}}\{\theta; \mathbf{x^n}, \mathbf{y^n}\}\pi_{\mathcal{P}}(\theta|\mathbf{x^n})d\theta, \tag{18}$$

$$\textit{where} \quad d_{S_{\mathcal{Q}}}\{\theta; \mathbf{x^n}, \mathbf{y^n}\} = \widehat{J}_{n,S_{\mathcal{Q}}}(\theta, \check{\psi}^{\mathcal{Q}}) - \widehat{J}_{n,S_{\mathcal{Q}}}(\hat{\theta}^{\mathcal{Q}}, \hat{\psi}^{\mathcal{Q}}). \tag{19}$$

Here, $\widehat{J}_{n,S_{\mathcal{Q}}}(\theta, \psi)$ is the J-distance defined in (4), $(\hat{\theta}^{\mathcal{Q}}, \hat{\psi}^{\mathcal{Q}})$ is the GMM estimator, and $(\theta, \check{\psi}^{\mathcal{Q}})$ is the constrained GMM estimator with fixed θ.

The idea of our fragility measure is to quantify the in-sample over-fitting of a structural model. In Equation (19), $d_{S_{\mathcal{Q}}}\{\theta; \mathbf{x^n}, \mathbf{y^n}\}$ is the GMM analog of the log likelihood ratio of the model with jointly-fitted baseline parameters $\hat{\theta}^{\mathcal{Q}}$ and nuisance parameters $\hat{\psi}^{\mathcal{Q}}$, which provide the best in-sample fit of the data based on the GMM criterion, against an alternative model with baseline parameters θ and the fitted nuisance parameters $\check{\psi}^{\mathcal{Q}}$. Assuming true parameter is θ instead of $\hat{\theta}^{\mathcal{Q}}$, the fact that the J-distance based on $(\hat{\theta}^{\mathcal{Q}}, \hat{\psi}^{\mathcal{Q}})$ is smaller is a symptom of over-fitting. Notice that

[9] The theoretical result in this section is the only place we assume full knowledge of the likelihood function for the baseline model.

we use $d_{S_{\mathcal{Q}}}\{\theta; \mathbf{x^n}, \mathbf{y^n}\}$ as a loss function to quantify the degree of over-fitting of the model, not as a statistic for hypothesis testing.

The weights attached to alternative models are essential for our definition of the over-fitting tendency. We consider alternative values of θ, while tuning the nuisance parameters ψ to fit the data as well as possible under the same criterion, i.e., we choose $\psi = \check{\psi}^{\mathcal{Q}}$, the constrained GMM estimator. Starting with a prior $\pi(\theta)$, we weigh the various alternative models using $\pi_{\mathcal{P}}(\theta|\mathbf{x^n})$ – the posterior for θ based on baseline model's moment conditions and data $\mathbf{x^n}$. The weighted average of $d_{S_{\mathcal{Q}}}\{\theta; \mathbf{x^n}, \mathbf{y^n}\}$ over the entire set of alternative models represents the tendency of a model to over-fit the data.

The weights on various alternative models depend on $\mathcal{P}$, therefore $\varrho_o(\theta_0|\psi_0, \mathbf{x^n}, \mathbf{y^n})$ is a measure of over-fitting tendency of $\mathcal{Q}$ *relative to the baseline model* $\mathcal{P}$. That is, the measure $\varrho_o(\theta_0|\psi_0, \mathbf{x^n}, \mathbf{y^n})$ quantifies the effective degrees of freedom exploited by the structural model by choosing θ to achieve an accurate in-sample fit under the additional structural restrictions imposed by $\mathcal{Q}$. The measure of model's over-fitting tendency depends on the choice of the baseline model. For example, consider a consumption-based asset pricing model aiming to describe simultaneously the pricing of equity and equity options. One natural choice would be to treat the statistical description of cash flows and consumption as the baseline model, with the structural model adding economic restrictions on prices of equity and options. Alternatively, if equity prices are covered by the baseline model, then the over-fitting tendency captures model's flexibility in fitting option prices, in addition to equity prices and fundamentals. In general, the choice of baseline model depends on which aspects of the model are intended to be covered by the fragility analysis.

The distribution over the alternative models also depends on the choice of the prior $\pi(\theta)$. If the econometrician does not have any information about θ beyond the baseline model and the data $\mathbf{x^n}$, an "uninformative" prior would be a desirable choice, one candidate being the Jeffrey's prior. If the econometrician has additional information about θ (e.g., from additional data), such information can be incorporated through an informative prior.

Our measure differs from the measure of SBCL in two respects. First, to give the econometrician flexibility to focus on specific features of a model and to reflect typical lack of knowledge of the full likelihood function, we adopt the GMM framework as opposed to the likelihood framework. Second, SBCL do not explicitly specify the weighting of alternative models. We operationalize the over-fitting measure by proposing a particular weighting of alternative models, tied to the baseline model.

Next, we generalize the fragility measure $\varrho_o(\theta_0|\psi_0, \mathbf{x^n}, \mathbf{y^n})$ to allow for transformations of parameters θ.

Definition 4 (Over-fitting Tendency Measure with Feature Functions)**.** *Let f be a $\mathbb{R}^{D_\Theta} \to \mathbb{R}^{D_f}$ continuously differentiable mapping with $1 \leq D_f \leq D_\Theta$. Then, we define*

$$\varrho_o^f(\theta_0|\psi_0, \mathbf{x^n}, \mathbf{y^n}) \equiv \int d_{S_\Omega}\{f(\theta); \mathbf{x^n}, \mathbf{y^n}\} \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) d\theta, \tag{20}$$

$$\text{where} \quad d_{S_\Omega}\{f(\theta); \mathbf{x^n}, \mathbf{y^n}\} = \inf_{(\tilde{\theta}, \tilde{\psi}): f(\tilde{\theta}) = f(\theta)} \widehat{J}_{n,S_\Omega}(\tilde{\theta}, \tilde{\psi}) - \widehat{J}_{n,S_\Omega}(\hat{\theta}^\Omega, \hat{\psi}^\Omega). \tag{21}$$

Here $\widehat{J}_{n,S_\Omega}(\theta, \psi)$ is defined in (4), and $(\hat{\theta}^\Omega, \hat{\psi}^\Omega)$ is the GMM estimator.

Transforming the original parameter vector is useful, for example, if one wants to measure model's robustness with respect to a low-dimensional subset in the parameter space. For instance, to measure model robustness with respect to the first D_f elements of θ ($D_f < D_\Theta$), we set $f(\theta) = \mathbf{F}\theta$, where $\mathbf{F} = \nabla f(\theta) \equiv \left[I_{D_f}, O_{D_f \times (D_\Theta - D_f)}\right]$. In the special case of $f(\theta) = \mathbf{F}\theta$, with $\mathbf{F}$ being an arbitrary full-rank $D_\Theta \times D_\Theta$ matrix, we recover the original overall fragility measure, $\varrho_o^f(\theta_0|\psi_0, \mathbf{x^n}, \mathbf{y^n}) = \varrho_o(\theta_0|\psi_0, \mathbf{x^n}, \mathbf{y^n})$.[10]

[10] A similar monotonicity property to Proposition 17 applies to $\varrho_o^f(\theta_0, \psi_0; \mathbf{x^n}, \mathbf{y^n})$. Let $\widetilde{f} = [f, f_1]'$, where f and f_1 are continuously differentiable and $D_{\widetilde{f}} \leq D_\Theta$. Then $\varrho_o^f(\theta_0|\psi_0, \mathbf{x^n}, \mathbf{y^n}) \leq \varrho_o^{\widetilde{f}}(\theta_0|\psi_0, \mathbf{x^n}, \mathbf{y^n})$.

3.2 Fisher Fragility Measures and the Over-Fitting Tendency

We now show that our Fisher fragility measure captures in a certain sense the over-fitting tendency of the structural model. We first introduce the regularity conditions sufficient for our theoretical results. We do so briefly, as these are standard (for more details see the Appendix A). We assume the process $\{\mathbf{x_t}, \mathbf{y_t}\}$ to be a strictly stationary Markov process (Assumption A1). Moreover, we assume that the time series satisfy the uniform mixing condition as in Newey (1985a) and White and Domowitz (1984). The mixing condition in Assumption A2 and the dominance condition in Assumption A3 are needed for the uniform law of large numbers as in White and Domowitz (1984). They also imply the moment continuity of stochastic functions $g_{\mathbb{Q}}$, as well as their derivatives, as in Hansen (1982). Lastly, we need identification conditions to guarantee that the minimization problem in (5) has a unique solution asymptotically (Assumptions A4 and A5). A standard sufficient condition for GMM identification is that the covariance matrix $S_{\mathbb{Q}}$ is positive definite, the moment conditions (1) and (3) are satisfied only at θ_0, and the Fisher information matrix $\mathbf{I}_{\mathbb{Q}}(\theta_0, \psi_0)$ is non-singular (see e.g., Hansen, 1982, Assumptions 3.4 - 3.6). Next, we assume the prior $\pi(\theta)$ is twice continuously differentiable and positive (Assumption A6). Our last assumption A7 imposes a restriction on feature functions, requiring them to be representable as a projection of a smooth, globally invertible mapping on a lower-dimensional linear subspace. Under the above assumptions, we establish the following theorem and its corollaries (the proofs are in Appendix B).

Theorem 1. *Consider a feature function $f : \mathbb{R}^{D_\Theta} \to \mathbb{R}^{D_f}$ with $\mathbf{v} = \nabla f(\theta_0)$ being the $D_f \times D_\Theta$ Jacobian matrix. Suppose the regularity conditions above (Assumptions A1 - A7 in Appendix A) hold. Then $\varrho_o^f(\theta_0|\psi_0, \mathbf{x^n}, \mathbf{y^n})$ converges in distribution to*

$$\underset{n\to\infty}{\text{wlim}}\, \varrho_o^f(\theta_0|\psi_0, \mathbf{x^n}, \mathbf{y^n}) = \varrho^{\mathbf{v}}(\theta_0|\psi_0) + \sum_{i=1}^{D_f} (\lambda_i(\mathbf{v}) - 1)\, \chi^2_{1,i}, \tag{22}$$

where $\chi^2_{1,i}$'s are i.i.d. chi-squared random variables with 1 *degree of freedom, and*

$\lambda_i(\mathbf{v})$'s are eigenvalues of $\Pi_0(\mathbf{v})$ defined in (11). Here the operator "wlim" denotes a variable with the limiting distribution (in the sense of weak convergence).

This result immediately leads to the following two corollaries.

Corollary 1. *Suppose the regularity conditions of Theorem 1 hold, then $\varrho^{\mathbf{v}}(\theta_0|\psi_0)$ and $\varrho_o^f(\theta_0|\psi_0, \mathbf{x^n}, \mathbf{y^n})$ are asymptotically related:*

$$\mathbb{E}\left[\operatorname*{wlim}_{n\to\infty} \varrho_o^f(\theta_0|\psi_0, \mathbf{x^n}, \mathbf{y^n})\right] = 2\varrho^{\mathbf{v}}(\theta_0|\psi_0) - D_f, \tag{23}$$

where $\mathbb{E}$ stands for the expectation under the distribution of the full model $\mathbb{Q}$.

Theorem 1 and Corollary 1 show that the Fisher fragility measure $\varrho^{\mathbf{v}}(\theta_0|\psi_0)$ captures the mean of the asymptotic distribution of the over-fitting tendency measure. Moreover, individual eigenvalues $\lambda_i(\mathbf{v})$ provide additional information about the asymptotic distribution of the over-fitting tendency. As we show in the Corollary below, the likelihood of extremely poor out-of-sample performance (i.e. an extremely large $\varrho_o^f(\theta_0|\psi_0, \mathbf{x^n}, \mathbf{y^n})$) is determined by the largest of the eigenvalues $\lambda_i(\mathbf{v})$ (recall that $\varrho^{\mathbf{v}}(\theta_0|\psi_0) = \sum_{i=1}^{D_f} \lambda_i(\mathbf{v})$).

Corollary 2. *Suppose the regularity conditions of Theorem 1 hold and the largest eigenvalue is $\lambda_1(\mathbf{v})$. Then, the tail probability of the limiting variable converges to zero at the exponential rate related negatively to $\lambda_1(\mathbf{v})$,*

$$\lim_{x\to\infty} \frac{1}{x} \ln \mathbb{Q}_n \left\{\operatorname*{wlim}_{n\to\infty} \varrho_o^f(\theta_0|\psi_0, \mathbf{x^n}, \mathbf{y^n}) > x\right\} = -\frac{1}{2\left(\lambda_1(\mathbf{v}) - 1\right)}, \tag{24}$$

where $\mathbb{Q}_n$ is the true joint probability measure for $(\mathbf{x^n}, \mathbf{y^n})$ under the full model.

Corollary 2 shows that the tail of the limiting distribution of the over-fitting tendency measure converges to zero faster when the largest eigenvalue $\lambda_1(\mathbf{v})$ is smaller. Thus, for a given value of $\varrho^{\mathbf{v}}(\theta_0|\psi_0)$, which captures the *average* tendency of over-fitting, a heavily skewed distribution of eigenvalues $\lambda_i(\mathbf{v})$ results in a larger value of $\lambda_1(\mathbf{v})$. Then, the tail of the distribution of $\varrho_o^f(\theta_0|\psi_0, \mathbf{x^n}, \mathbf{y^n})$ is heavier, and the probability of over-fitting the data to an extremely large degree is higher.

4 Applications

In this section we implement the Fisher fragility measure in the context of two widely used asset pricing models. The first example is a rare disaster model, for which we compute the fragility measure analytically. The second example is a long-run risk model. We use this example to demonstrate how one can diagnose the sources of fragility in a more complex model and deal with nuisance parameters in measuring model fragility (or "dark matter").

4.1 Disaster Risk Model

Rare economic disasters are a natural source of "dark matter" in asset pricing models. It is difficult to evaluate the likelihood and the magnitude of rare disasters statistically. Yet, agents' aversion to large disasters can have an economically large effect on asset prices.[11]

We consider a disaster risk model similar to Barro (2006). The structural model describes the log growth rate of aggregate consumption g_t and the excess log return on the market portfolio r_t. There are two regimes characterized by state variable z_t: the disaster regime ($z_t = 1$) and the normal regime ($z_t = 0$). The variable z_t has an i.i.d. Bernoulli distribution, being equal to one with probability p, independently of the other variables. The realizations of z_t are observable. Briefly, the regimes z_t are i.i.d. distributed as

$$z_t = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } 1-p. \end{cases} \tag{25}$$

In a normal regime ($z_t = 0$), $g_t = u_t$, where the log consumption growth in normal times u_t follows a normal distribution $N(\mu, \sigma^2)$. In the disaster state ($z_t = 1$), $g_t = -v_t$,

[11] See the early work by Rietz (1988), and recent developments by Barro (2006), Gabaix (2012), Martin (2012), Wachter (2013), and Collin-Dufresne, Johannes, and Lochstoer (2016), among others. Welch (2016) shows how rare-disaster models of equity returns may be rejected with the addition of derivative prices to the analysis. We limit our analysis to a simple disaster-risk model of stock returns to illustrate how our measure of model fragility helps compare alternative models that are consistent with available data.

where the log of decline in consumption v_t follows a truncated exponential distribution with density:

$$\nu_t \overset{\text{i.i.d.}}{\sim} \mathbf{1}\{v_t > \underline{v}\}\xi e^{-\xi(v_t-\underline{v})}. \tag{26}$$

Thus, the lower bound for disaster size equals to $\underline{v}$ and the average disaster size equals to $\underline{v} + 1/\xi$.

The excess log return on the stock market portfolio r_t is correlated with the consumption growth g_t. Specifically, the joint distribution of (g_t, r_t) is time-varying and depends on the underlying disaster state z_t. When the economy is in a normal regime ($z_t = 0$), g_t and r_t are jointly normal:

$$r_t = \eta + \rho\frac{\tau}{\sigma}(g_t - \mu) + \sqrt{1-\rho^2}\tau\varepsilon_{0,t}, \tag{27}$$

where $\varepsilon_{0,t}$ are independent standard normal shocks. The coefficient $\rho\frac{\tau}{\sigma}$ is the "leverage factor" in the normal regime. This factor controls the sensitivity of stock returns to consumption shocks. When the economy is in a disaster state ($z_t = 1$), the excess log return is linked to the decline in consumption with a different "leverage factor" $b > 0$:

$$r_t = bg_t + \varsigma\varepsilon_{1,t} \tag{28}$$

where $\varepsilon_{1,t}$ are independent standard normal shocks. Thus, stock returns and consumption growth are imperfectly correlated in the disaster state.

Next, we assume that the representative agent has a separable, constant relative risk aversion utility function $\sum_{t=0}^{\infty}\delta_{\text{D}}^t c_t^{1-\gamma_{\text{D}}}/(1-\gamma_{\text{D}})$, where $\gamma_{\text{D}} > 0$ is the coefficient of relative risk aversion. The log equity premium is then given by

$$\mathbb{E}[r_t] = (1-p)\eta - pb\left(\underline{v} + 1/\xi\right), \tag{29}$$

where the Euler equation implies that η is a function of the other parameters (see

Appendix D for details):

$$\eta = \gamma_{\rm D}\rho\sigma\tau - \frac{\tau^2}{2} + \ln\left[1 + e^{\gamma_{\rm D}\mu - \frac{\gamma_{\rm D}^2\sigma^2}{2}}\Delta(\xi)\frac{p}{1-p}\right] \tag{30}$$

$$\approx \gamma_{\rm D}\rho\sigma\tau - \frac{\tau^2}{2} + e^{\gamma_{\rm D}\mu - \frac{\gamma_{\rm D}^2\sigma^2}{2}}\Delta(\xi)\frac{p}{1-p}, \tag{31}$$

where

$$\Delta(\xi) = \xi\left(\frac{e^{\gamma_{\rm D}\underline{v}}}{\xi - \gamma_{\rm D}} - \frac{e^{\frac{\varsigma^2}{2} + (\gamma_{\rm D} - b)\underline{v}}}{\xi + b - \gamma_{\rm D}}\right). \tag{32}$$

Equations (29–32) provide a cross-equation restriction on the processes for consumption growth and stock market returns. The first two terms on the right hand side of (30) describe the market risk premium due to Gaussian consumption shocks. The third term is due to the disaster risk premium. The model requires $\xi > \gamma_{\rm D}$ for the risk premium to be finite, which sets an upper bound on the average disaster size and dictates how heavy the tail of the disaster size distribution can be.

The fact that the log equity premium $\mathbb{E}[r_t]$ explodes as ξ approaches the value of $\gamma_{\rm D}$ is a crucial feature for the model in this section. Even extremely rare disasters (very small p) can generate an arbitrarily large risk premium $\mathbb{E}[r_t]$ as long as the average disaster size is sufficiently large, which corresponds to low values of ξ. Extremely rare and large disasters are difficult to rule out based on standard statistical tests. Below we illustrate how our fragility measure can detect fragility in models with such features.

Fisher fragility measure. Equations (25–32) specify the full structural model. We set the baseline model to be the statistical model for rare disasters $\mathbf{x_t} = (z_t, v_t)$, described by (25) and (26). Thus, the baseline parameters are $\theta = (p, \xi)$. In this example, we leave the stock return process out of the baseline model and set $\mathbf{y_t} = (g_t, r_t)$. This choice of how to partition the model into the baseline and the structural model sets the focus of the analysis on the parameters of the disaster process. Alternative specifications could broaden the scope of the fragility analysis to include, for instance, the parameters describing the joint distribution of shocks to stock returns and consumption growth. In that case, $\mathbf{x_t}$ would include the stock return process r_t. Finally, to simplify analytical derivations, we treat the other parameters $\phi = (\gamma_{\rm D}, \mu, \sigma, \underline{v}, \tau, \rho, b, \varsigma)$ as auxiliary parameters fixed at known values, making them a part of the functional-form specification; their values are chosen to fit the data in our quantitative analysis. In this example, the nuisance parameter vector ψ is empty.

Based on the approximation (31), the Fisher fragility measure is approximately equal to (see Appendix D for details):

$$\varrho(p,\xi) \approx 2 + \frac{p\Delta\left(\xi\right)^2 + p\left(1-p\right)\xi^2\dot{\Delta}\left(\xi\right)^2}{\left(1-\rho^2\right)\tau^2\left(1-p\right)^2} e^{2\gamma_{\rm D}\mu - \gamma_{\rm D}^2\sigma^2}, \tag{33}$$

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

$$\dot{\Delta}(\xi) = -\frac{e^{\gamma_{\rm D}\underline{v}}\gamma_{\rm D}}{(\xi-\gamma_{\rm D})^2} + \frac{e^{(\gamma_{\rm D}-b)\underline{v}}(\gamma_{\rm D}-b)}{(\xi-\gamma_{\rm D}+b)^2} e^{\varsigma^2/2}. \tag{34}$$

The one-dimensional worst-case asymptotic fragility measure is $\varrho^1(p,\xi) = \varrho(p,\xi) - 1$.

As Equation (30) shows, $\Delta(\xi)$ and $\dot{\Delta}(\xi)$ are related to the sensitivity of η to the disaster probability p and disaster size parameter ξ, respectively. When ξ approaches the value of $\gamma_{\rm D}$, both $\Delta(\xi)$ and $\dot{\Delta}(\xi)$ approach infinity. Thus, disaster risk models with high average disaster size are fragile according to our measure.

Quantitative analysis. In our quantitative analysis, we use annual real per-capita consumption growth (nondurables and services) from the NIPA and returns on the CRSP value-weighted market portfolio for the period of 1929 to 2011. We fix the auxiliary parameters $\mu, \sigma, \varsigma, \tau$ and ρ 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 $\rho = 59.36\%$. The lower bound for disaster size is $\underline{v} = 7\%$. The leverage parameter b is 3. In Figure 1, we plot the 95% and 99% confidence regions for (p,ξ) based on the baseline model.

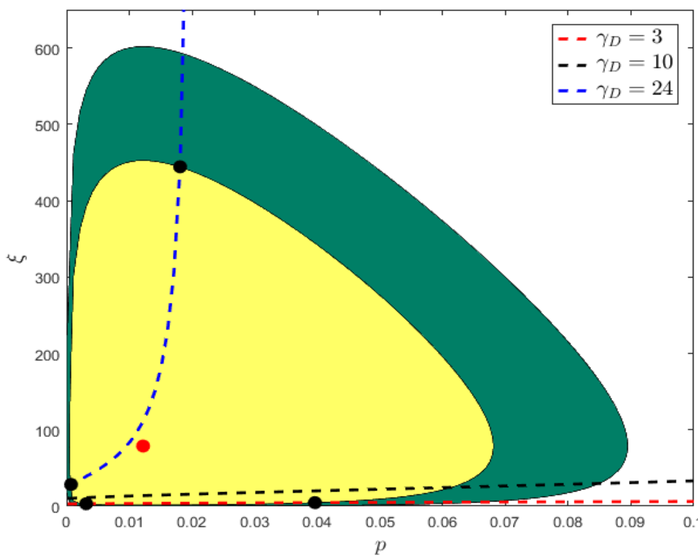

Figure 1: The 95% and 99% confidence regions of (p, ξ) for the unconstrained model and the equity premium isoquants implied by the asset pricing constraint (30) for $\gamma_{\mathrm{D}} = 3, 10, 24$. p is disaster probability, and ξ characterizes the inverse of average disaster size. The maximum likelihood estimates are $(\widehat{p}_{\mathrm{ML}}, \widehat{\xi}_{\mathrm{ML}}) = (0.0122, 78.7922)$.

The 95% confidence region for (p, ξ) is quite large. For low values of the disaster probability p, the baseline model has little power to reject models with a wide range of average disaster size values (ξ). Figure 1 also shows the equity premium isoquants for different levels of relative risk aversion: lines with the combinations of p and ξ required to match the unconditional equity premium of 5.09% for a given value of γ_{D}. The fact that these isoquants all intersect with the 95% confidence region implies that even for low risk aversion ($\gamma_{\mathrm{D}} = 3$), there exist combinations of (p, ξ) that not only match the observed equity premium, but also are "consistent with the macro data" in a sense that they cannot be rejected by the macro data based on standard statistical tests. In the remainder of this section, we refer to a calibration of (p, ξ) that is within the 95% confidence region as an "acceptable calibration."[12]

[12] 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.

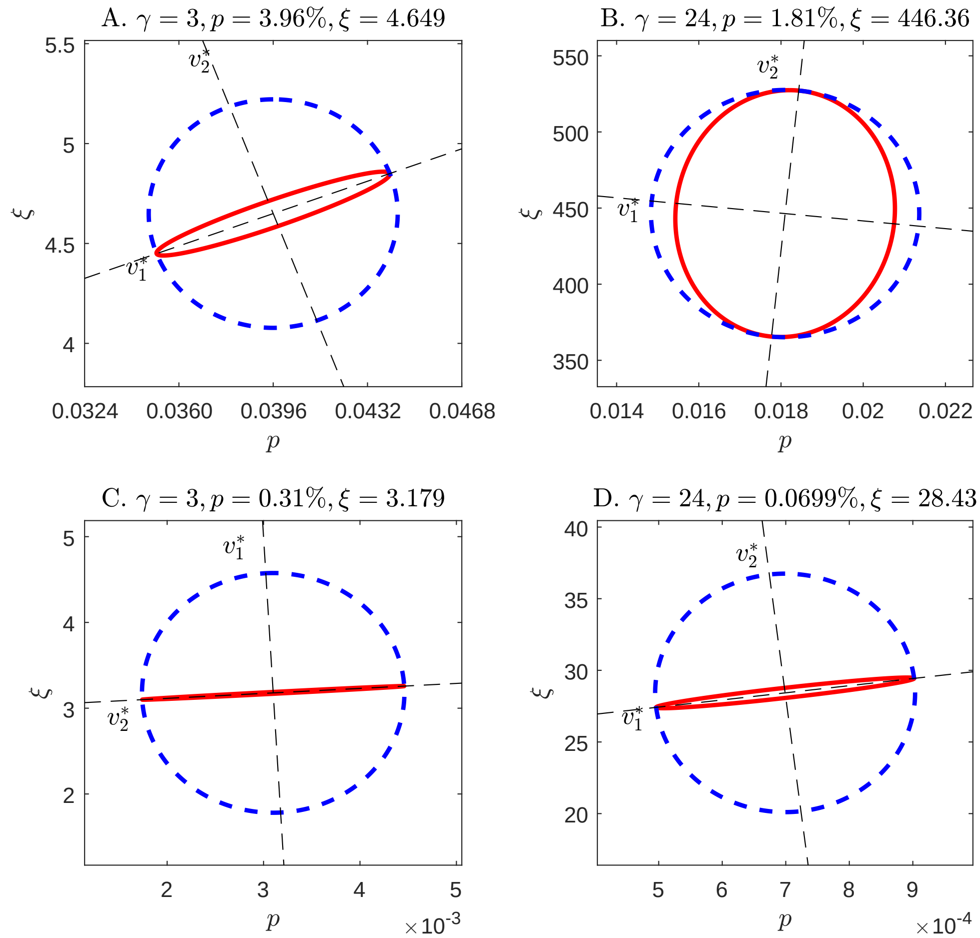

Figure 2: 95% confidence regions for the asymptotic distribution of the MLEs for four "acceptable calibrations." In Panels A through D, the Fisher fragility measures are $\varrho(p, \xi) = 75.03$, 2.49, 1.78×10^4, and 5.61×10^2 respectively. p is disaster probability, and ξ characterizes the inverse of average disaster size.

While it is difficult to distinguish among a wide range of calibrations using standard statistical tools based on the macro data, these calibrated models differ significantly based on our fragility measures. As an illustration, we focus on four alternative calibrations, as denoted by the four points located at the intersections of the equity premium isoquants ($\gamma_{\text{D}} = 3$ and 24) and the boundary of the 95% confidence region in Figure 1. For $\gamma_{\text{D}} = 3$, the two points are ($p = 3.96\%, \xi = 4.649$) and ($p = 0.31\%, \xi = 3.179$). For $\gamma_{\text{D}} = 24$, the two points are ($p = 1.81\%, \xi = 446.36$) and ($p = 0.0699\%, \xi = 28.43$).

With only two parameters in θ, we can illustrate the worst-case asymptotic fragility measure by plotting the asymptotic confidence regions for (p, ξ) in the baseline model and the structural model, as determined by the respective information matrices $\mathbf{I}_{\mathcal{P}}(\theta)$ and $\mathbf{I}_{\mathcal{Q}}(\theta)$.[13] In each panel of Figure 2, the largest dash-line circle is the 95% confidence region for (p, ξ) under the baseline model. The smaller solid-line ellipse is the 95% confidence region for (p, ξ) under the structural model. The reason that the confidence region under the structural model is smaller than that under the baseline model is that the GMM moments in the structural model contain both the moments for the baseline model and the moments (cross-equation restrictions) imposed by the structural component under fragility assessment. In this example, the two confidence regions coincide[14] in the direction of v_2^* and differ the most in the direction of v_1^*. Moreover, with enough extra data, the confidence region for the unconstrained estimator can be made small enough to reside within the confidence region of the constrained estimator.

In Panel A of Figure 2, with $\gamma_{\text{D}} = 3, p = 3.96\%, \xi = 4.649$, $\varrho(p, \xi) = 75.07$ and $\varrho^1(p, \xi) = 74.07$. This means that under the baseline model, we need to increase the amount of consumption data by a factor of 74.07 to match or exceed the precision in estimation of any linear combination of p and ξ afforded by the equity premium constraint. Panels C and D of Figure 2 correspond to the calibrations with "extra rare and large disasters." For $\gamma_{\text{D}} = 3$ and 24, $\varrho^1(p, \xi)$ rises to 1.78×10^4 and 5.60×10^2, respectively. If, in Panel B of Figure 2, we raise γ_{D} to 24 while changing the annual disaster probability to 1.81% and lowering the average disaster size to 7.002% ($\xi = 446.36$). As a result, $\varrho^1(p, \xi)$ declines to 1.49. The reason for reduced fragility in this calibration is the combination of higher disaster probability and a lower average disaster size.

So far, we have been examining the fragility of a specific calibrated structural model. We can also assess the fragility of a general class of models, relative to the baseline

[13] In fact, we use all the score functions of likelihoods to construct the moments, so the optimal GMM estimation is asymptotically equivalent to the MLE in our analysis of this disaster risk model.

[14] This is not true in general. When localized, the deterministic cross-equation restriction from the equity premium in this model is a linear constraint. Thus, the parameter estimates are not affected along the direction of the constraint.

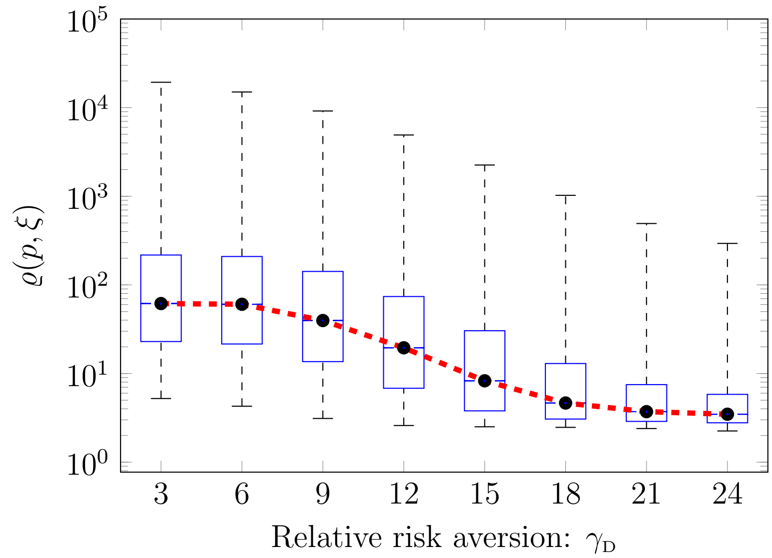

Figure 3: Distribution of the Fisher fragility measure $\varrho(p,\xi)$ for different levels of risk aversion. For each γ_{D}, the boxplot shows the $1, 25, 50, 75$, and 99-th percentile of the distribution of $\varrho(p,\xi)$ based on the constrained posterior for (p,ξ).

model of rare disasters, by plotting the distribution of $\varrho(\theta)$ based on a particular distribution of θ. For example, if econometricians are interested in fragility of a class of disaster risk models where the auxiliary parameters ϕ are fixed at given levels ϕ_0 and the uncertainty of baseline parameters θ is explicitly taken into account, we propose to use the posterior for (p,ξ) under the structural model (i.e., constrained posterior distribution) denoted by $\pi_{\mathbb{Q}}(\theta|\mathbf{x^n},\mathbf{y^n})$ as the distribution of θ.[15] Since the constrained posterior incorporates the information from the data and the asset pricing constraint, it can be viewed as summarizing full knowledge of the distribution of θ under the assumption that the model constraint is valid.

For each value of γ_{D}, Figure 3 shows the $1, 25, 50, 75$, and 99-th percentile of the distribution of $\varrho(\theta)$ based on $\pi_{\mathbb{Q}}(\theta|\mathbf{x^n},\mathbf{y^n})$. The Fisher fragility measures are higher when the levels of risk aversion are low. For example, for $\gamma_{\text{D}} = 3$, the $25, 50$, and 75-th percentile of the distribution of $\varrho(p,\xi)$ are 23.0, 61.6, and 217.4, respectively. Smaller values of γ_{D} force the constrained posterior for θ to place more weight on rare and

[15] We generate Figure 3 based on a simulated sample from the constrained posterior $\pi_{\mathbb{Q}}(\theta|\mathbf{x^n},\mathbf{y^n})$. For that, we use the Approximate Bayesian Computation (ABC) method, described in the online Appendix.

large disasters, which imposes particularly strong restrictions on the parameters (p, ξ). As γ_{D} rises, the mass of the constrained posterior shifts towards smaller disasters, which imply lower information ratios. For $\gamma_{\text{D}} = 24$, the 25, 50, and 75-th percentiles of the distribution of $\varrho(p, \xi)$ decline to 2.8, 3.5, and 5.8, respectively.

4.2 Long-run risk model

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

$$V_t = \left[(1 - \delta_{\text{L}}) C_t^{1-1/\psi_{\text{L}}} + \delta_{\text{L}} \left(\mathbb{E}_t \left[V_{t+1}^{1-\gamma_{\text{L}}} \right] \right)^{\frac{1-1/\psi_{\text{L}}}{1-\gamma_{\text{L}}}} \right]^{\frac{1}{1-1/\psi_{\text{L}}}}, \tag{35}$$

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

$$\Delta c_{t+1} = \mu_c + x_t + \sigma_t \epsilon_{c,t+1} \tag{36a}$$

$$x_{t+1} = \rho x_t + \varphi_x \sigma_t \epsilon_{x,t+1} \tag{36b}$$

$$\widetilde{\sigma}_{t+1}^2 = \overline{\sigma}^2 + \nu(\widetilde{\sigma}_t^2 - \overline{\sigma}^2) + \sigma_w \epsilon_{\sigma,t+1} \tag{36c}$$

$$\sigma_{t+1}^2 = \max\left(\underline{\sigma}^2, \widetilde{\sigma}_{t+1}^2\right) \tag{36d}$$

where the shocks $\epsilon_{c,t}, \epsilon_{x,t}$, and $\epsilon_{\sigma,t}$ are *i.i.d.* $N(0,1)$ and mutually independent. The volatility process (36c) 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.[16]

Next, the log dividend growth Δd_t follows

$$\Delta d_{t+1} = \mu_d + \phi_d x_t + \varphi_{d,c}\sigma_t\epsilon_{c,t+1} + \varphi_{d,d}\sigma_t\epsilon_{d,t+1}, \tag{37}$$

where the shocks $\epsilon_{d,t}$ are $i.i.d.$ $N(0,1)$ and mutually independent with the other shocks in (36a–36c).

From the consumption Euler equation, we derive a linear approximation of the stochastic discount factor,

$$m_{t+1} = \Gamma_0 + \Gamma_1 x_t + \Gamma_2\sigma_t^2 - \lambda_c\sigma_t\epsilon_{c,t+1} - \lambda_x\varphi_x\sigma_t\epsilon_{x,t+1} - \lambda_\sigma\sigma_w\epsilon_{\sigma,t+1}. \tag{38}$$

The formulae for the coefficients Γ_0, Γ_1, Γ_2, λ_c, λ_x, and λ_σ are standard in the long-run risk model literature (see the online Appendix). Moreover, the equilibrium excess (log) return follows

$$r^e_{m,t+1} = \mu^e_{r,t} + \beta_c\sigma_t\epsilon_{c,t+1} + \beta_x\sigma_t\epsilon_{x,t+1} + \beta_\sigma\sigma_w\epsilon_{\sigma,t+1} + \varphi_{d,d}\sigma_t\epsilon_{d,t+1}, \tag{39}$$

where the conditional average (log) excess return is

$$\mu^e_{r,t} = \lambda_c\beta_c\sigma_t^2 + \lambda_x\beta_x\varphi_x\sigma_t^2 + \lambda_\sigma\beta_\sigma\sigma_w^2 - \frac{1}{2}\sigma^2_{r_m,t}, \tag{40}$$

$$\text{where } \sigma^2_{r_m,t} = \beta_c^2\sigma_t^2 + \beta_x^2\sigma_t^2 + \beta_\sigma^2\sigma_w^2 + \varphi_{d,d}^2\sigma_t^2. \tag{41}$$

The expressions for β_c, β_x, and β_σ are also in the online Appendix.

There are stochastic singularities in the model. For instance, the excess log market

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

Table 1: Parameters of the Benchmark Long-Run Risk Model (Model 1)

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

Note: Model 2 has $\nu = 0.98$ and $\gamma_{\mathrm{L}} = 27$, while the other parameters are the same as in Model 1.

return $r^e_{m,t+1}$ 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 with other variables. This is a standard approach in DSGE literature for dealing with stochastic singularity. In our computation of the fragility measure, we consider the moments based on the joint dynamics of time series $(\Delta c_{t+1}, x_t, \sigma^2_t, \Delta d_{t+1}, r^e_{m,t+1})$. Details can be found in the online Appendix.

Quantitative analysis. We choose the model of consumption (36a–36d) as the baseline model $\mathcal{P}$. We assume that the econometrician observes the process for consumption, the latent variables x_t and σ^2_t, and the process for asset returns. We make the latent variables observable to be consistent with the postulated process for asset returns, which is derived assuming that these variables are observable.

Accordingly, the baseline parameters are $\theta = (\mu_c, \rho, \varphi_x, \overline{\sigma}^2, \nu, \sigma_w, \mu_d, \phi_d, \varphi_{d,c}, \varphi_{d,d})$ with $D_\Theta = 10$ and $\mathbf{x_t} = (\Delta c_{t+1}, x_t, \sigma^2_t, \Delta d_{t+1})$. By measuring the fragility of the long-run risk model relative to this particular baseline, we can interpret the fragility measure as quantifying the additional information that asset pricing restrictions provide for the consumption and dividend dynamics (in particular, the long-run risk components) relative to the information contained in consumption and dividend data. We explicitly account for uncertainty about preference parameters γ_{L} and ψ_{L} by including them into the nuisance parameter vector ψ. Thus, $\psi = (\gamma_{\text{L}}, \psi_{\text{L}})$. The extra data investigated by the full structural model $\mathfrak{Q}$ are $\mathbf{y_t} = r^e_{m,t+1}$. Other parameters are included in the auxiliary parameter vector $\phi = (\delta_{\text{L}}, \varphi_r)$ which are fixed at known values, and these values are part of the imposed functional-form specification of the structural component that is under the fragility assessment. Note that the baseline model covers the joint dynamics of consumption growth and dividend growth. The structural model adds the description of the distribution of stock returns in relation to the consumption and dividend growth processes.

The parameter values of Model 1 follows Bansal, Kiku, and Yaron (2012) and is summarized in Table 1. As Bansal, Kiku, and Yaron (2012) (Table 2, p. 194) show, the simulated first and second moments, based on the parametrization of Model 1, match the set of key asset pricing moments in the data reasonably well. The same is true for Model 2 whose parameter values are also reported in Table 2 (see the online Appendix).

First, consider Panel I of Table 2. This panel contains fragility measures computed under the specification that treats preference parameters as nuisance parameters. The row (M1) reports fragility measures for Model 1. The Fisher fragility measure is $\varrho = 276.3$, while the worst-case 1-dimensional Fisher fragility measure is also high, $\varrho^1 = 196.3$. This implies that to match the precision of the estimator for the full structural model in all 1-D directions, the estimator based on the baseline model would require a time-series sample that is 196.3 times as long.

The values of ϱ and ϱ^1 suggest that the asset pricing implications of the structural model are highly sensitive to plausible perturbations in parameter values. Note that to quantify the full extent of model fragility, it is not sufficient to consider perturbations of parameters one at a time. We compute the fragility measure for each individual parameter $\varrho^{\mathbf{v}}$, where $\mathbf{v}$ is the appropriate standard basis vector $\mathbf{e}_i$. All of the univariate measures are much lower than the full fragility measure ϱ or the

Table 2: Fragility Measures for the Long-Run Risk Models

Model	ϱ	ϱ^1	$\varrho^{\mathbf{v}}$									
			μ_c	ρ	φ_x	$\overline{\sigma}^2$	ν	σ_w	μ_d	ϕ_d	$\varphi_{d,c}$	$\varphi_{d,d}$
	I. Nuisance parameter vector ψ: $(\gamma_{\mathrm{L}}, \psi_{\mathrm{L}})$											
(M1)	276.3	196.3	1.0	1.1	1.0	48.9	97.8	1.0	1.0	3.4	1.0	1.0
(M2)	34.0	21.1	1.0	1.1	1.0	1.0	3.4	1.0	1.4	4.2	1.0	1.0
	II. Nuisance parameter vector ψ: empty											
(M1)	$3.58 \cdot 10^5$	$3.57 \cdot 10^5$	1.0	2.1	1.1	115.6	117.5	1.3	1.1	7.1	1.0	1.0
(M2)	323.3	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-D fragility measure ϱ^1 for Model 1 (M1) is given by $v_1^* = [0.000, 0.000, -0.000, 0.020, -0.001, 0.999, -0.001, 0.000, -0.000, 0.000]$. Model 2 (M2) has $\nu = 0.98$ and $\gamma_{\mathrm{L}} = 27$ with other parameters unchanged. In panel I, the uncertainty of preference parameters $(\gamma_{\mathrm{L}}, \psi_{\mathrm{L}})$ are accounted for. In panel II, these parameters are fixed as auxiliary parameters ϕ with nuisance parameter vector ψ empty.

worst-case one-dimensional fragility measure ϱ^1, with the the fragility measure being higher 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.

In comparison, in Panel II of Table 2 we show fragility measures under the specification that ignores the uncertainty about preference parameters. This type of analysis is natural if the model is not fully estimated, but rather the preference parameters are fixed at certain values. For instance, one may specifically design a model to capture the moments of asset returns with a low value of risk aversion. In that case, the choice of the preference parameters is effectively subsumed by the specification of the functional form of the model, and treating them as auxiliary parameters is in line with the logic of the model construction. The fragility measures in Panel II are higher. In particular, the overall fragility ϱ and the worst-case 1-D fragility ϱ^1 increase dramatically from 276.3 to 3.58×10^5 and from 196.3 to 3.57×10^5, respectively.

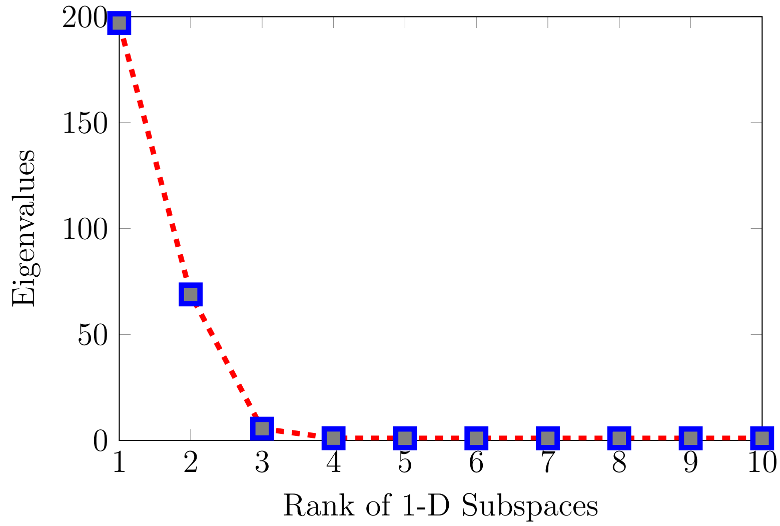

Figure 4: Eigenvalues for $\mathbf{I}_{\mathbb{Q}}(\theta_0|\psi_0)^{\frac{1}{2}}\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{I}_{\mathbb{Q}}(\theta_0|\psi_0)^{\frac{1}{2}}$ of the benchmark parametrization (Model 1).

Diagnosing the sources of fragility. Besides measuring the fragility of the model, the Fisher fragility measures can be used to diagnose the sources of fragility. The ranking of the eigenvalues of $\mathbf{I}_{\mathbb{Q}}(\theta_0|\psi_0)^{\frac{1}{2}}\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{I}_{\mathbb{Q}}(\theta_0|\psi_0)^{\frac{1}{2}}$ is informative. Each eigenvalue denotes the marginal contribution of a one-dimensional subspace to the overall fragility measure (see Definition 2 and Proposition 2). As Figure 4 shows, there are large differences between the eigenvalues. Model fragility along the worst direction in one-dimensional subspaces, as captured by the leading eigenvalue, is 196.3, which accounts for over 71% of the total fragility. This result means that one can effectively reduce the dimensionality (from 10 to 1) when analyzing the fragility of the model.

Having determined the worst direction (more precisely, a one-dimensional subspace) in the parameter space, v_1^*, we now investigate which properties of the model are most sensitive to the change in θ along that direction. For illustration, we focus on four such properties, the risk loading and price of risk for volatility shocks $(\beta_\sigma\sigma_w, \lambda_\sigma\sigma_w)$, and for growth shocks $(\beta_x\sigma_t, \lambda_x\varphi_x\sigma_t)$. The conditional market excess return depends on these parameter combinations (see Equation (40)).

In Figure 5, we plot the sensitivities of $\beta_\sigma, \lambda_\sigma, \beta_x$ and λ_x with respect to perturbations of θ along the worst direction v_1^* (solid line). We measure the size of a

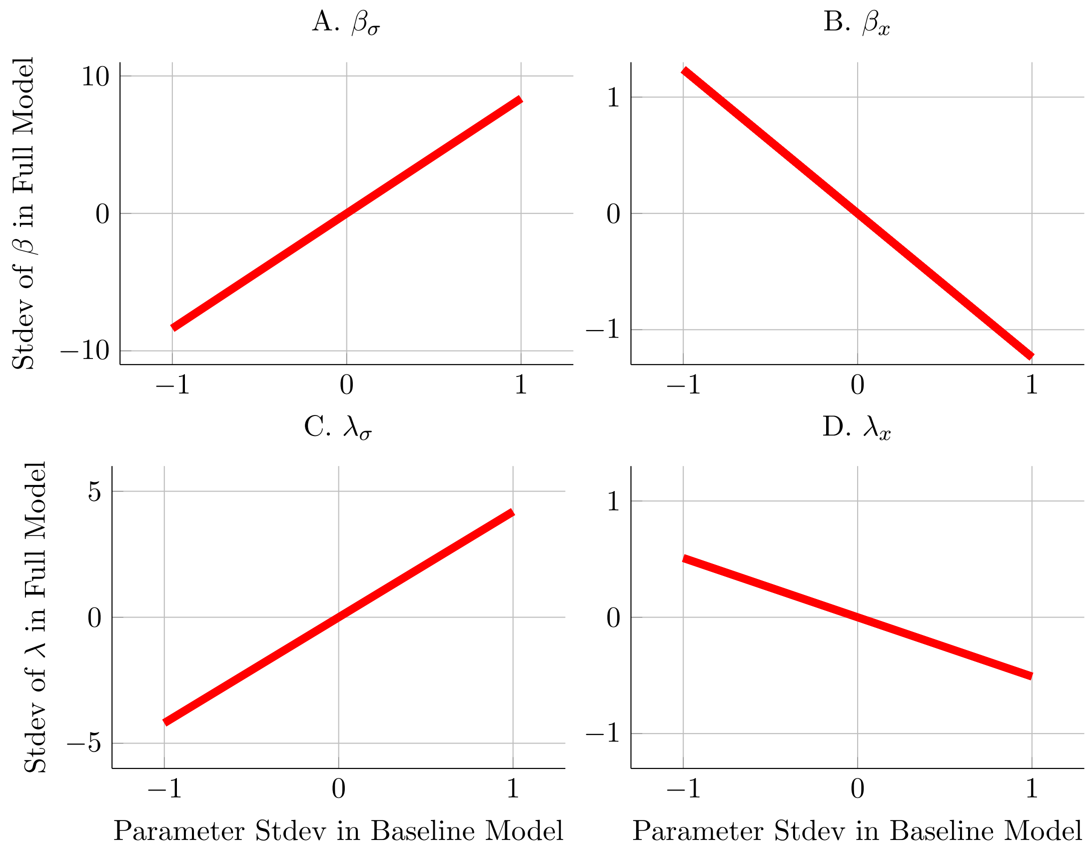

Figure 5: Sensitivity of return betas and risk prices with respect to the perturbation along the worst-case direction in Model 1 with nuisance parameters $\psi = (\gamma_{\rm L}, \psi_{\rm L})$.

perturbation of θ relative to the standard deviation of θ in the baseline model $\mathcal{P}$. We normalize changes in the objects of interest (e.g., β_σ in panel A) by their own standard deviations in the full structural model $\mathcal{Q}$.

The risk loading and the price of risk for volatility shocks are both highly sensitive to changes in θ along the direction of v_1^*. For example, a one-standard deviation change in θ along the direction of v_1^* leads to a nine-standard deviation change in β_σ under the full structural model $\mathcal{Q}$. Thus, an important source of fragility of the long-run risk model based on the benchmark parametrization (Model 1 in Table 1) is in the model implications for the risk exposure of stock returns to volatility shocks. In comparison, panels B and D show that model implications for the stock return exposure to growth rate shocks, and the price of risk of such shocks, are far less sensitive to the parameter values along the worst-case direction.

Finally, we can further trace the sources of fragility by examining how λ_σ and β_σ are determined. For example, the fact that the persistence parameter for the conditional variance of consumption growth, ν, is close to 1, makes both β_σ and λ_σ sensitive to changes in θ. This motivates us to consider an alternative parametrization (Model 2 in Table 1) with a lower value of ν. Specifically, we change ν from 0.999 to 0.98, and simultaneously raise the coefficient of relative risk aversion γ from 10 to 27 in order to match the unconditional equity premium as in the benchmark parametrization (Model 1). The rest of the parameters are unchanged. Model 2 produces asset pricing moments largely similar to those in Model 1. However, based on our measures, Model 2 is much less fragile. As Panel I of Table 2 shows, under Model 2, ϱ declines from 276.3 to 34, and ϱ^1 declines from 196.3 to 21.1.

Fragility of a model with latent state variables depends critically on whether the agents in the model are subject to the same limitations as the econometrician. A model in which agents observe the latent variables but the econometrician does not (see e.g., Schorfheide, Song, and Yaron, 2014) is potentially even more fragile than its fully observable counterpart, since lack of observability adds effective degrees of freedom. In contrast, models where the agents themselves have to learn about the data generating process, (see e.g., Collin-Dufresne, Johannes, and Lochstoer, 2016), the cross-equation restrictions implied by asset prices differ, and it is therefore difficult to anticipate the effect on model fragility. Furthermore, models in which agents effectively act on distorted beliefs, (e.g., Hansen and Sargent, 2010) may exhibit less fragility because asset prices may reflect agents' beliefs, rather than supplying excessive information about the consumption and cash flow processes.

5 Conclusion

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

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

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

References

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

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

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

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

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

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

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

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

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

Breusch, T., H. Qian, P. Schmidt, and D. Wyhowski, 1999, "Redundancy of moment conditions," *Journal of Econometrics*, 91, 89–111.

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

Chen, H., W. W. Dou, and L. Kogan, 2017, "Supplement to "Measuring the 'Dark Matter' in Asset Pricing Models"," Discussion paper.

Chernozhukov, V., and H. Hong, 2003, "An MCMC approach to classical estimation," *J. Econometrics*, 115, 293–346.

Clarke, B. S., 1999, "Asymptotic normality of the posterior in relative entropy," *IEEE Trans. Inform. Theory*, 45, 165–176.

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

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

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

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

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

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

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

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

Gentzkow, M., and J. M. Shapiro, 2013, "Measuring the Sensitivity of Parameter Estimates to Sample Statistics," Chicago booth working papers.

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

Hahn, J., W. Newey, and R. Smith, 2011, "Tests for neglected heterogeneity in moment condition models," CeMMAP working papers CWP26/11, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

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

Hansen, L., and R. Jagannathan, 1997, "Assessing Specification Errors in Stochastic Discount Factor Models," *Journal of Finance*, 52, 557–90.

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

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

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

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

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

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

Hausman, J., R. Lewis, K. Menzel, and W. Newey, 2011, "Properties of the CUE estimator and a modification with moments," *Journal of Econometrics*, 165, 45 – 57.

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

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

Krantz, S. G., and H. R. Parks, 2013, *The implicit function theorem* . Modern Birkhäuser Classics, Birkhäuser/Springer, New York, History, theory, and applications, Reprint of the 2003 edition.

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

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

Martin, I., 2012, "Consumption-Based Asset Pricing with Higher Cumulants," Forthcoming, Review of Economic Studies.

Nakagawa, K., 2005, "Tail probability of random variable and Laplace transform," *Appl. Anal.*, 84, 499–522.

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

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

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

Newey, W. K., and K. D. West, 1987b, "Hypothesis testing with efficient method of moments estimation," *Internat. Econom. Rev.*, 28, 777–787.

Pinsker, M. S., 1964, *Information and Information Stability of Random Variables and Processes* . originally published in Russian in 1960, San Francisco: Holden-Day, San Francisco.

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

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

Schorfheide, F., D. Song, and A. Yaron, 2014, "Identifying Long-Run Risks: A Bayesian Mixed-Frequency Approach," Nber working papers, National Bureau of Economic Research, Inc.

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

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

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

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

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

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

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

Welch, I., 2016, "The (time-varying) importance of disaster risk," *Financial Analyst Journal*, 72, 14–30.

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

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

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

Appendix

A Regularity Conditions for Theoretical Results

The regularity conditions we choose to impose on the behavior of the data are influenced by two major considerations. First, our assumptions are chosen to allow for processes with sequential dependence. In particular, the processes should be relevant for intertemporal asset pricing models. Second, our assumptions are sufficient conditions. We are not attempting to establish the weakest conditions that guarantee the results; instead, we impose regularity conditions that are relatively straightforward to verify in practice.

Assumption A1 (Stationarity Condition)
The underlying time series (x_t, y_t) with $t = 1, \cdots, n$ follow an m_S-order strictly stationary Markov process.

Remark. *This assumption implies that the marginal conditional density for x_t can be specified as $\pi_{\mathcal{P}}(x_t|\theta, x_{t-1}, \cdots, x_{t-m_S})$. Define the stacked vectors $\mathbf{x_t} = (x_t, \cdots, x_{t-m_S+1})^T$ and $\mathbf{y_t} = (y_t, \cdots, y_{t-m_S+1})^T$. Then the marginal conditional density from the parametric family specified for the baseline model can be rewritten as $\pi_{\mathcal{P}}(\mathbf{x_t}; \theta)$, and the stacked vectors $(\mathbf{x_t}, \mathbf{y_t})$ follow a first-order Markov process.*

Assumption A2 (Mixing Condition)
Assume that there exists a constant λ_D satisfying $\lambda_D \geq 2d_D/(d_D - 1)$, where d_D is the constant in Assumption A3 below, such that $(\mathbf{x_t}, \mathbf{y_t})$ for $t = 1, 2, \cdots, n$ is uniform mixing and there exists a constant $\bar{\phi}$ such that the uniform mixing coefficients satisfy

$$\phi(m) \leq \bar{\phi} m^{-\lambda_D} \quad \text{for all possible probabilistic models,}$$

where $\phi(m)$ is the uniform mixing coefficient. The definition of $\phi(m)$ is standard and can be found, for example, in White and Domowitz (1984) or Bradley (2005).

Remark. *Following the literature (see, e.g., White and Domowitz, 1984; Newey, 1985b; Newey and West, 1987a), we adopt the mixing conditions as a convenient way of describing economic and financial data which allows time dependence and heteroskedasticity. In particular, we employ the uniform mixing which is discussed in White and Domowitz (1984).*

Assumption A3 (Dominance Condition)
The function $g_{\mathcal{Q}}(\theta, \psi; \mathbf{x}, \mathbf{y})$ is twice continuously differentiable in (θ, ψ) almost surely. There exist dominating measurable functions $a_1(\mathbf{x}, \mathbf{y})$ and $a_2(\mathbf{x}, \mathbf{y})$, and constant $d_D > 1$, such that almost everywhere

$$\begin{aligned}
&|g_{\mathcal{Q}}(\theta, \psi; \mathbf{x}, \mathbf{y})|^2 \leq a_1(\mathbf{x}, \mathbf{y}), \quad ||\nabla g_{\mathcal{Q}}(\theta, \psi; \mathbf{x}, \mathbf{y})||_{\mathbb{S}}^2 \leq a_1(\mathbf{x}, \mathbf{y}), \\
&||\nabla^2 g_{\mathcal{Q},(i)}(\theta, \psi; \mathbf{x}, \mathbf{y})||_{\mathbb{S}}^2 \leq a_1(\mathbf{x}, \mathbf{y}), \quad \text{for } i = 1, \cdots, D_g, \\
&|q(\mathbf{x}, \mathbf{y})| \leq a_2(\mathbf{x}, \mathbf{y}), \quad |q(\mathbf{x_1}, \mathbf{y_1}, \mathbf{x_t}, \mathbf{y_t})| \leq a_2(\mathbf{x_1}, \mathbf{y_1}) a_2(\mathbf{x_t}, \mathbf{y_t}), \quad \text{for } t \geq 2, \\
&\int [a_1(\mathbf{x}, \mathbf{y})]^{d_D} a_2(\mathbf{x}, \mathbf{y}) \mathrm{d}x\mathrm{d}y < +\infty, \quad \int a_2(\mathbf{x}, \mathbf{y}) \mathrm{d}x\mathrm{d}y < +\infty,
\end{aligned}$$

where $|| \cdot ||_{\mathbb{S}}$ is the spectral norm of matrices.

Remark. *The dominating function assumption is widely adopted in the literature (Newey, 1985a,b; Newey and West, 1987a).*

Consider the assumption in Section 3 where $g_{\mathcal{P}}(\theta; \mathbf{x})$ contains all the score functions of the well specified baseline likelihood. For each pair of j and k, it holds that for some constant $\zeta > 0$ and a *sufficiently large constant $C > 0$,*

$$\mathbb{E} \sup_{\theta \in \Theta} \left| \frac{\partial^2}{\partial \theta_j \partial \theta_k} \ln \pi_{\mathcal{P}}(\mathbf{x}; \theta) \right|^{2+\zeta} < C, \quad \mathbb{E} \sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta_j} \ln \pi_{\mathcal{P}}(\mathbf{x}; \theta) \right|^{2+\zeta} < C, \quad \text{and} \tag{42}$$

the dominance condition, together with the uniform mixing assumption and stationarity assumption, implies the stochastic equicontinuity condition (i) in Proposition 3 of Chernozhukov and Hong (2003), which is later used in the proof of Proposition 4.

Assumption A4 (Nonsingularity Condition)
The Fisher information matrices $\mathbf{I}_{\mathcal{P}}(\theta)$ and $\mathbf{I}_{\mathcal{Q}}(\theta, \psi)$ are positive definite for all θ, ψ.

Remark. *Assumption A4 implies that the covariance matrices $S_{\mathcal{P}}$ and $S_{\mathcal{Q}}$ are positive definite, and the expected moment function gradients $G_{\mathcal{P}}(\theta)$ and $G_{\mathcal{Q}}(\theta, \psi)$ have full rank for all θ and ψ.*

Assumption A5 (Identification Condition)
The true baseline parameter vector θ_0 is identified by the baseline moment conditions in the sense that $\mathbb{E}[g_{\mathcal{P}}(\theta; \mathbf{x})] = 0$ only if $\theta = \theta_0$. Similarly, the true parameters (θ_0, ψ_0) of the full model are identified by the moment conditions in the sense that $\mathbb{E}[g_{\mathcal{Q}}(\theta, \psi; \mathbf{x}, \mathbf{y})] = 0$ only if $\theta = \theta_0$ and $\psi = \psi_0$.

Remark. *Consider the assumption of well-specified likelihood in Section 3. The continuous differentiability of moment functions, together with the identification condition, imply that the parametric family of distributions $\mathbb{P}_\theta$, as well as the moment conditions, are sound: the convergence of a sequence of parameter values is equivalent to the weak convergence of the distributions:*

$$\theta \to \theta_0 \quad \Leftrightarrow \quad \mathbb{P}_\theta \to \mathbb{P}_{\theta_0} \quad \Leftrightarrow \quad \mathbb{E}[\ln(d\mathbb{P}_\theta / d\mathbb{P})] \to \mathbb{E}[\ln(d\mathbb{P}_{\theta_0} / d\mathbb{P})] = 0. \tag{43}$$

The convergence of a sequence of parameter values is equivalent to the convergence of the moment conditions:

$$(\theta, \psi) \to (\theta_0, \psi_0) \quad \Leftrightarrow \quad \mathbb{E}[g_{\mathcal{Q}}(\theta, \psi; \mathbf{x}, \mathbf{y})] \to \mathbb{E}[g_{\mathcal{Q}}(\theta_0, \psi_0; \mathbf{x}, \mathbf{y})] = 0. \tag{44}$$

Assumption A6 (Regular Bayesian Condition)
Suppose the parameter set is $\Theta \times \Psi \subset \mathbb{R}^{D_\Theta + D_\Psi}$ with Θ and Ψ being compact. The prior is absolutely continuous with respect to the Lebesgue measure with Radon-Nikodym density $\pi(\theta, \psi)$, which is twice continuously differentiable and positive. Denote $\overline{\pi} \equiv \max_{\theta \in \Theta, \psi \in \Psi} \pi(\theta, \psi)$ and $\underline{\pi} \equiv \min_{\theta \in \Theta, \psi \in \Psi} \pi(\theta, \psi)$.

Remark. *Compactness implies total boundedness. In those cases where the parameter set for the prior is not compact, e.g., due to the adoption of an uninformative prior, one can truncate the parameter set at sufficiently large values so that the main results are not affected.*

Assumption A7 (Regular Feature Function Condition)
The feature function $f = (f_1, \cdots, f_{D_f}) : \Theta \to \mathbb{R}^{D_f}$ is a twice continuously differentiable vector-valued function. We assume that there exist $D_\Theta - D_f$ twice continuously differentiable functions $f_{D_f+1}, \cdots, f_{D_\Theta}$ on Θ such that $F = (f_1, f_2, \cdots, f_{D_\Theta}) : \Theta \to \mathbb{R}^{D_\Theta}$ is a one-to-one mapping (i.e. an injection) and $F(\Theta)$ is a connected and compact D_Θ-dimensional subset of $\mathbb{R}^{D_\Theta}$.

Remark. *A simple sufficient condition for the regular feature function condition to hold is that each function f_i $(i = 1, \cdots, D_f)$ is a proper and twice continuously differentiable function on $\mathbb{R}^{D_\Theta}$ and $\frac{\partial f(\theta)}{\partial(\theta_{(1)}, \cdots, \theta_{(D_f)})} > 0$ at each $\theta \in \mathbb{R}^{D_\Theta}$. In this case, we can simply choose $f_k(\theta) \equiv \theta_{(k)}$ for $k = D_f + 1, \cdots, D_\Theta$. Then, the Jacobian determinant of F is nonzero at each $\theta \in \mathbb{R}^{D_\Theta}$ and F is a proper and twice differentiable mapping $\mathbb{R}^{D_\Theta} \to \mathbb{R}^{D_\Theta}$. According to the Hadamard's Global Inverse Function Theorem (e.g. Krantz and Parks, 2013), F is a one-to-one mapping and $F(\Theta)$ is a connected and compact D_Θ-dimensional subset of $\mathbb{R}^{D_\Theta}$.*

B Proof of Theorem 1 And Its Corollaries

B.1 Asymptotic Normality of Posteriors

Proposition 4. *Under Assumptions A1 - A6 in Appendix A, it holds that*

$$\mathbf{D}_{KL}(\pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}})||N(\hat{\theta}^{\mathcal{P}}_{ML}, n^{-1}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1})) \to 0 \quad in \ \ \mathbb{P}_n.$$

Proof. We extend the proof of Theorem 2.1 in Clarke (1999) which is under the i.i.d. condition. First, we show that $\sup_{\theta\in\Theta}|\widehat{H}_{\mathcal{P},n}(\theta)| = O_p(1)$ where

$$\widehat{H}_{\mathcal{P},n}(\theta) \equiv -\frac{1}{n}\sum_{t=1}^{n}\ln \pi_{\mathcal{P}}(\mathbf{x_t};\theta). \tag{45}$$

When n is large enough, we have

$$\sup_{\theta\in\Theta}|\widehat{H}_{\mathcal{P},n}(\theta)| \le 1 + \frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}|\ln \pi_{\mathcal{P}}(\mathbf{x_t};\theta)|.$$

Based on the mixing condition and the dominance condition, it follows from Theorem 2.3 of White and Domowitz (1984) that

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}|\ln \pi_{\mathcal{P}}(\mathbf{x_t};\theta)| \to \mathbb{E}\sup_{\theta\in\Theta}|\ln \pi_{\mathcal{P}}(\mathbf{x_t};\theta)| \quad \text{a.s.}$$

which further implies that $\sup_{\theta\in\Theta}|\widehat{H}_{\mathcal{P},n}(\theta)| = O_p(1)$. Second, we show that

$$\int u^T u \left|\pi_{\mathcal{P}}(\hat{\theta}^{\mathcal{P}}_{\text{ML}} + u/\sqrt{n}|\mathbf{x}^{\mathbf{n}}) - \varphi_{\mathcal{P}}(u)\right| \mathrm{d}u \to 0 \quad \text{in} \ \ \mathbb{P}_n \tag{46}$$

where $\varphi_{\mathcal{P}}(u) = \sqrt{\det(\mathbf{I}_{\mathcal{P}}(\theta_0))/(2\pi)^{D_\Theta}}\exp\left[-\frac{1}{2}u^T\mathbf{I}_{\mathcal{P}}(\theta_0)u\right]$. Clarke (1999) shows that when $\mathbf{x_t}$ is i.i.d., the limit result (46) holds under the regularity conditions in Assumptions A3 - A6. To extend this limit result to allow for weak dependence, we appeal to Theorem 1 and Proposition 3 of Chernozhukov and Hong (2003), whose conditions are implied by Assumptions A1 - A6 in Appendix A. ☐

B.2 Proof of Theorem 1

In this section, we prove the result of Theorem 1. One major additional technical challenge, compared with standard large-sample inferences of GMM (see Hansen, 1982), is that we need to establish uniform convergence and bounds for constrained GMM over a set of locally mis-specified moment constraints. Thus we need special treatments in our proof in establishing the uniform convergence and bounds.

Because of Assumption A7, as well as the fact that the definition of our over-fitting tendency measure (Definition 4) and Assumptions A1 - A6 are invariant under invertible and second-order smooth transformations of parameters, we can assume that $f(\theta) = \theta_1 \equiv (\theta_{(1)}, \cdots, \theta_{(D_f)})^T$ and hence $\nabla f(\theta) \equiv \mathbf{v} = \left[I_{D_f}, O_{D_f\times(D_\Theta - D_f)}\right]$, without loss of generality. That is, $f(\theta) = \theta_1 = \mathbf{v}\theta$. The remaining baseline parameters are summarized in $\theta_2 \equiv (\theta_{(D_f+1)}, \cdots, \theta_{(D_\Theta)})^T$. Then, the over-fitting tendency measure with feature functions can be written as

$$\varrho^f_o(\theta_0|\psi_0, \mathbf{x}^{\mathbf{n}}, \mathbf{y}^{\mathbf{n}}) \equiv \int d_{S_\mathcal{Q}}\{\mathbf{v}\theta; \mathbf{x}^{\mathbf{n}}, \mathbf{y}^{\mathbf{n}}\}\pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}})\mathrm{d}\theta, \tag{47}$$

$$\text{where} \quad d_{S_\mathcal{Q}}\{\mathbf{v}\theta; \mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n}\} = \inf_{(\tilde{\theta},\tilde{\psi}):\mathbf{v}\tilde{\theta}=\mathbf{v}\theta} \widehat{J}_{n,S_\mathcal{Q}}(\tilde{\theta},\tilde{\psi}) - \widehat{J}_{n,S_\mathcal{Q}}(\hat{\theta}^\mathcal{Q},\hat{\psi}^\mathcal{Q}). \tag{48}$$

We present the proof as a sequence of steps.

(1) A local reparametrization. To establish the asymptotic equivalence result in a parametric setting, we follow the convention of asymptotic statistics (see, e.g. van der Vaart, 1998) to consider the local reparametrization:

$$(\theta,\psi) = (\hat{\theta}^\mathcal{P},\hat{\psi}^\mathcal{P}) + \frac{1}{\sqrt{n}}(u,h). \tag{49}$$

Thus, the over-fitting tendency measure defined in (47) and (48) can be rewritten as

$$\varrho_o^f(\theta_0|\psi_0,\mathbf{x}^\mathbf{n},\mathbf{y}^\mathbf{n}) \equiv \int d_{S_\mathcal{Q}}\{\mathbf{v}u; \mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n}\}\pi_\mathcal{P}(\hat{\theta}^\mathcal{P} + u/\sqrt{n})\mathrm{d}u, \tag{50}$$

$$d_{S_\mathcal{Q}}\{\mathbf{v}u; \mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n}\} = \inf_{(\tilde{u},\tilde{h}):\mathbf{v}\tilde{u}=\mathbf{v}u} \widehat{J}_{n,S_\mathcal{Q}}\left(\hat{\theta}^\mathcal{P} + \tilde{u}/\sqrt{n},\ \hat{\psi}^\mathcal{P} + \tilde{h}/\sqrt{n}\right) - \widehat{J}_{n,S_\mathcal{Q}}\left(\hat{\theta}^\mathcal{P} + \hat{u}^\mathcal{Q}/\sqrt{n},\ \hat{\psi}^\mathcal{P} + \hat{h}^\mathcal{Q}/\sqrt{n}\right). \tag{51}$$

The transformed parameter $u_1 = \mathbf{v}u$ is defined by $\mathbf{v}u = \sqrt{n}\left(\mathbf{v}\theta - \mathbf{v}\hat{\theta}^\mathcal{P}\right)$, and other transformed parameters are defined analogously. The GMM estimator of the transformed parameters is $(\hat{u}^\mathcal{Q},\hat{h}^\mathcal{Q})$, such that $(\hat{\theta}^\mathcal{Q},\hat{\psi}^\mathcal{Q}) = \left(\hat{\theta}^\mathcal{P} + \hat{u}^\mathcal{Q}/\sqrt{n}, \hat{\psi}^\mathcal{P} + \hat{h}^\mathcal{Q}/\sqrt{n}\right)$.

(2) Uniform quadratic bounds for the over-fitting gap $d_{S_\mathcal{Q}}\{\mathbf{v}u; \mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n}\}$. Assumption A3 (dominance condition) implies that the distance $d_{S_\mathcal{Q}}\{\mathbf{v}u; \mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n}\}$ is bounded by

$$0 \le d_{S_\mathcal{Q}}\{\mathbf{v}u; \mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n}\} \le n\underline{\lambda}_{S_\mathcal{Q}}^{-1} \sup_{\theta\in\Theta,\psi\in\Psi} |\widehat{g}_{\mathcal{Q},n}(\theta,\psi)|^2 \le \underline{\lambda}_{S_\mathcal{Q}}^{-1}\sum_{t=1}^n a_1(\mathbf{x_t},\mathbf{y_t}),$$

where $\underline{\lambda}_{S_\mathcal{Q}}$ is the smallest eigenvalue of $S_\mathcal{Q}$. The uniform upper bound is crude because it does not take advantage of $\widehat{g}_{\mathcal{Q},n}(\theta,\psi)$ being close to zero when (θ,ψ) is in the local neighborhood of (θ_0,ψ_0).

Now, we provide much sharper quadratic bounds for $d_{S_\mathcal{Q}}\{\mathbf{v}u; \mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n}\}$. First, it holds that

$$d_{S_\mathcal{Q}}\{\mathbf{v}u; \mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n}\} \le \widehat{J}_{n,S_\mathcal{Q}}\left(\hat{\theta}^\mathcal{P} + \frac{u}{\sqrt{n}}, \hat{\psi}^\mathcal{Q}\right) - \widehat{J}_{n,S_\mathcal{Q}}\left(\hat{\theta}^\mathcal{P} + \frac{\hat{u}^\mathcal{Q}}{\sqrt{n}}, \hat{\psi}^\mathcal{Q}\right). \tag{52}$$

The inequality above is point by point for all samples $(\mathbf{x}^\mathbf{n},\mathbf{y}^\mathbf{n})$ and all $u \in \mathbb{R}^{D_\Theta}$.

The second-order Taylor expansion of $\widehat{J}_{n,S_\mathcal{Q}}\left(\hat{\theta}^\mathcal{P} + u/\sqrt{n}, \hat{\psi}^\mathcal{Q}\right)$ around $(\hat{\theta}^\mathcal{Q},\hat{\psi}^\mathcal{Q})$ implies

$$\begin{aligned} &\widehat{J}_{n,S_\mathcal{Q}}\left(\hat{\theta}^\mathcal{P} + u/\sqrt{n}, \hat{\psi}^\mathcal{Q}\right) - \widehat{J}_{n,S_\mathcal{Q}}\left(\hat{\theta}^\mathcal{P} + \hat{u}^\mathcal{Q}/\sqrt{n}, \hat{\psi}^\mathcal{Q}\right) \\ &\qquad = \frac{1}{2}(u-\hat{u}^\mathcal{Q})^T\left\{\Gamma_\Theta\left[n^{-1}\nabla^2\widehat{J}_{n,S_\mathcal{Q}}(\theta_u,\hat{\psi}^\mathcal{Q})\right]\Gamma_\Theta^T\right\}(u-\hat{u}^\mathcal{Q}), \end{aligned} \tag{53}$$

where Γ_Θ is the projection matrix defined in (9), and θ_u lies on the segment between $\hat{\theta}^\mathcal{P} + u/\sqrt{n}$ and $\hat{\theta}^\mathcal{Q} = \hat{\theta}^\mathcal{P} + \hat{u}^\mathcal{Q}/\sqrt{n}$. The Hessian matrix is

$$\begin{aligned} n^{-1}\nabla^2\widehat{J}_{n,S_\mathcal{Q}}(\theta_u,\hat{\psi}^\mathcal{Q}) &= 2\left[\nabla\widehat{g}_{\mathcal{Q},n}(\theta_u,\hat{\psi}^\mathcal{Q})\right]^T S_\mathcal{Q}^{-1}\left[\nabla\widehat{g}_{\mathcal{Q},n}(\theta_u,\hat{\psi}^\mathcal{Q})\right] \\ &\quad + 2\sum_{i=1}^{D_g}\nabla^2\hat{g}_{\mathcal{Q},n,(i)}(\theta_u,\hat{\psi}^\mathcal{Q})\left[\mathbf{e}_i^T S_\mathcal{Q}^{-1}\hat{g}_{\mathcal{Q},n}(\theta_u,\hat{\psi}^\mathcal{Q})\right], \end{aligned} \tag{54}$$

where $\mathbf{e}_i$ is the basis vector with its i-th element equal to 1 and others equal to 0, and $\hat{g}_{\mathcal{Q},n,(i)}(\theta_u, \hat{\psi}^{\mathcal{Q}})$ is the i-th element of $\hat{g}_{\mathcal{Q},n}(\theta_u, \hat{\psi}^{\mathcal{Q}})$.

According to the expression above and Assumption A3, there exists a sequence of integrable nonnegative variables $D_{\mathcal{Q},n}(\mathbf{x^n}, \mathbf{y^n})$ such that $D_{\mathcal{Q},n}(\mathbf{x^n}, \mathbf{y^n})$'s first moments are uniformly bounded over n and

$$\left|\left|\Gamma_\Theta \left[n^{-1}\nabla^2 J_{n,S_{\mathcal{Q}}}(\theta_u, \hat{\psi}^{\mathcal{Q}}; \mathbf{x^n}, \mathbf{y^n})\right] \Gamma_\Theta^T\right|\right|_{\mathcal{S}} \leq D_{\mathcal{Q},n}(\mathbf{x^n}, \mathbf{y^n}), \quad \text{for all } u. \tag{55}$$

From (52), (53), and (55), the over-fitting gap $d_{S_{\mathcal{Q}}}\{\mathbf{v}u; \mathbf{x^n}, \mathbf{y^n}\}$ has the upper bound

$$d_{S_{\mathcal{Q}}}\{\mathbf{v}u; \mathbf{x^n}, \mathbf{y^n}\} \leq D_{\mathcal{Q},n}(\mathbf{x^n}, \mathbf{y^n})|u - \hat{u}^{\mathcal{Q}}|^2, \quad \text{for all } u. \tag{56}$$

(3) Uniform lower bounds for the Hessian matrix of J-distance $n^{-1}\nabla^2 \widehat{J}_{n,S_{\mathcal{Q}}}(\theta, \psi)$. Now we show that there exist constants $\underline{\lambda}_J > 0$ and $\delta > 0$ such that

$$\mathbb{Q}_n \left\{\text{All eigenvalues of } n^{-1}\nabla^2 \widehat{J}_{n,S_{\mathcal{Q}}}(\theta, \psi) > \underline{\lambda}_J \text{ for all } (\theta, \psi) \in \mathcal{N}_0(\delta)\right\} \to 1, \tag{57}$$

where $\mathcal{N}_0(\delta) \equiv \left\{(\theta, \psi): \ |\theta - \theta_0|^2 + |\psi - \psi_0|^2 \leq \delta^2\right\}$. To prove the convergence result of (57), we first appeal to the Uniform Law of Large Numbers (ULLN) in White and Domowitz (1984): the sample averages $g_{\mathcal{Q},n}(\theta, \psi; \mathbf{x^n}, \mathbf{y^n})$, $\nabla g_{\mathcal{Q},n}(\theta, \psi; \mathbf{x^n}, \mathbf{y^n})$, and $\nabla^2 g_{\mathcal{Q},n}(\theta, \psi; \mathbf{x^n}, \mathbf{y^n})$ converge to their population means in probability uniformly over $\Theta \times \Psi$ due to the dominance condition and the mixing condition. Thus,

$$n^{-1}\nabla^2 \widehat{J}_{n,S_{\mathcal{Q}}}(\theta, \psi) \to 2G_{\mathcal{Q}}(\theta, \psi)^T S_{\mathcal{Q}}^{-1} G_{\mathcal{Q}}(\theta, \psi) \ \text{ in } \mathbb{Q}_n \text{ uniformly over } (\theta, \psi). \tag{58}$$

Because of the second-order continuous differentiability of the moment functions and the dominance condition, the limit on the right-hand side of (58) is a continuous function of (θ, ψ) due to the Dominance Convergence Theorem. Moreover, at (θ_0, ψ_0), the limiting function in (58) is equal to $G_{\mathcal{Q}}^T S_{\mathcal{Q}}^{-1} G_{\mathcal{Q}}$ which is positive definite. Thus, the uniform convergence and the continuity of the limiting function directly imply the result in (57).

Furthermore, under the reparametrization in (49), the result above implies that for an arbitrary constant $K > 0$, the probability that the smallest eigenvalue of $\nabla^2 \widehat{J}(\hat{\theta}^{\mathcal{P}} + u/\sqrt{n}, \hat{\psi}^{\mathcal{P}} + h/\sqrt{n})$ is bigger than $\underline{\lambda}_J$ converges to one uniformly over $|u| \leq K$ and $|h| \leq K$.

For the remainder of the proof, it suffices to focus on the set defined in (57), which has probability approaching one asymptotically. Following the convention, we refer to this set as the "big probability set."

(4) First-order conditions of GMM estimators. The GMM estimator $(\hat{\theta}^{\mathcal{Q}}, \hat{\psi}^{\mathcal{Q}}) = (\hat{\theta}^{\mathcal{P}}, \hat{\psi}^{\mathcal{P}}) + \frac{1}{\sqrt{n}}(\hat{u}^{\mathcal{Q}}, \hat{h}^{\mathcal{Q}})$ satisfies the first-order condition

$$0 = \nabla \widehat{J}_{n,S_{\mathcal{Q}}}(\hat{\theta}^{\mathcal{Q}}, \hat{\psi}^{\mathcal{Q}}). \tag{59}$$

The constrained GMM estimator $(\check{\theta}^{\mathcal{Q}}, \check{\psi}^{\mathcal{Q}})$ with restriction $\check{\theta}^{\mathcal{Q}} = \theta$ is the minimizer of $\widehat{J}_{n,S_{\mathcal{Q}}}(\tilde{\theta}, \tilde{\psi})$ subject to the constraint $\mathcal{R}(\tilde{\theta}, \tilde{\psi}; \theta, \psi) = 0$ where

$$\mathcal{R}(\tilde{\theta}, \tilde{\psi}; \theta, \psi) \equiv \mathbf{v}\Gamma_\Theta \begin{pmatrix} \tilde{\theta} - \theta \\ \tilde{\psi} - \psi \end{pmatrix} = \frac{1}{\sqrt{n}} \mathbf{v}\Gamma_\Theta \begin{pmatrix} \tilde{u} - u \\ \tilde{h} - h \end{pmatrix} = \frac{1}{\sqrt{n}} \mathcal{R}(\tilde{u}, \tilde{h}; u, h).$$

The constraint is linear and effectively restricts the first D_f elements of the baseline parameter vector to be $\theta_1 = \mathbf{v}\theta$. Recall that θ_1 is the first D_f elements of θ. The gradient of the constraint is $\nabla \mathcal{R}(\tilde{\theta}, \tilde{\psi}) \equiv \mathbf{v}\Gamma_\Theta$. The first-order condition and the complementarity condition (equality constraint) imply

$$\nabla \widehat{J}_{n,S_{\mathcal{Q}}}(\check{\theta}^{\mathcal{Q}}, \check{\psi}^{\mathcal{Q}}) = (\mathbf{v}\Gamma_{\Theta})^T \Lambda_n, \tag{60}$$

and

$$\mathcal{R}(\check{\theta}^{\mathcal{Q}}, \check{\psi}^{\mathcal{Q}}; \theta, \psi) = 0, \tag{61}$$

where Λ_n is a $D_f \times 1$ vector containing the Lagrange multipliers of constraints. The constrained GMM estimator $(\check{\theta}^{\mathcal{Q}}, \check{\psi}^{\mathcal{Q}})$ depends on the parameters $u_1 = \mathbf{v}u$ or $\theta_1 = \mathbf{v}\theta$.

(5) Uniform bounds and convergence for constrained GMM estimators. Under Assumptions A1-A5, one can show that the (reparametrized) GMM estimators $(\hat{u}^{\mathcal{Q}}, \hat{h}^{\mathcal{Q}})$ are $O_p(1)$ variables (see Hansen, 1982). However, the (reparametrized) constrained GMM estimators $(\check{u}^{\mathcal{Q}}, \check{h}^{\mathcal{Q}})$ depend on the restriction parameter $u_1 = \mathbf{v}u$, and thus it is unclear whether the constrained GMM estimators are uniformly $O_p(1)$ variables over all u, even though $(\check{u}^{\mathcal{Q}}, \check{h}^{\mathcal{Q}})$ are $O_p(1)$ for each fixed u.

The second-order Taylor expansion of $J_{n,S_{\mathcal{Q}}}(\check{\theta}^{\mathcal{Q}}, \check{\psi}^{\mathcal{Q}})$ around $(\hat{\theta}^{\mathcal{Q}}, \hat{\psi}^{\mathcal{Q}})$ implies that

$$d_{S_{\mathcal{Q}}}\{\mathbf{v}u; \mathbf{x^n}, \mathbf{y^n}\} = \frac{1}{2} \begin{pmatrix} \check{u}^{\mathcal{Q}} - \hat{u}^{\mathcal{Q}} \\ \check{h}^{\mathcal{Q}} - \hat{h}^{\mathcal{Q}} \end{pmatrix}^T \left[n^{-1} \nabla^2 \widehat{J}_{n,S_{\mathcal{Q}}}(\theta_u, \psi_u) \right] \begin{pmatrix} \check{u}^{\mathcal{Q}} - \hat{u}^{\mathcal{Q}} \\ \check{h}^{\mathcal{Q}} - \hat{h}^{\mathcal{Q}} \end{pmatrix}, \tag{62}$$

where (θ_u, ψ_u) lies on the segment between $(\hat{\theta}^{\mathcal{P}} + \check{u}^{\mathcal{Q}}/\sqrt{n},\ \hat{\psi}^{\mathcal{P}} + \check{h}^{\mathcal{Q}}/\sqrt{n})$ and $(\hat{\theta}^{\mathcal{P}} + \hat{u}^{\mathcal{Q}}/\sqrt{n},\ \hat{\psi}^{\mathcal{P}} + \hat{h}^{\mathcal{Q}}/\sqrt{n})$.

On the big probability set (57), it holds that

$$d_{S_{\mathcal{Q}}}\{\mathbf{v}u; \mathbf{x^n}, \mathbf{y^n}\} \geq \underline{\lambda}_J \left(|\check{u}^{\mathcal{Q}} - \hat{u}^{\mathcal{Q}}|^2 + |\check{h}^{\mathcal{Q}} - \hat{h}^{\mathcal{Q}}|^2 \right). \tag{63}$$

Combining (56) and (63), it holds that for any $\epsilon > 0$, there exist a large enough constant $K > 0$ and a small enough $\underline{\lambda}_J > 0$ such that

$$\limsup_{n \to \infty} \mathbb{Q}_n \left\{ |\check{u}^{\mathcal{Q}} - \hat{u}^{\mathcal{Q}}|^2 + |\check{h}^{\mathcal{Q}} - \hat{h}^{\mathcal{Q}}|^2 \leq \underline{\lambda}_J^{-1} D_{\mathcal{Q},n}(\mathbf{x^n}, \mathbf{y^n})(|\hat{u}^{\mathcal{Q}}|^2 + K^2),\ \forall\ |u| \leq K \right\} < \epsilon$$

where $D_{\mathcal{Q},n}(\mathbf{x^n}, \mathbf{y^n})$ is defined in (55). This result is crucial since it implies that the constrained GMM estimators $(\check{\theta}^{\mathcal{Q}}, \check{\psi}^{\mathcal{Q}})$ converges to (θ_0, ψ_0) at the rate of $\sqrt{n}$ in probability uniformly over $|u| \leq K$ for large enough constant K. Therefore, according to the ULLN result of (58), it follows that for large enough K

$$n^{-1} \nabla^2 \widehat{J}_{n,S_{\mathcal{Q}}}(\theta_u, \psi_u) \to 2 G_{\mathcal{Q}}^T S_{\mathcal{Q}}^{-1} G_{\mathcal{Q}},\ \text{uniformly over } |u| \leq K \text{ in } \mathbb{Q}_n. \tag{64}$$

where $G_{\mathcal{Q}} \equiv G_{\mathcal{Q}}(\theta_0, \psi_0)$. Combining the second-order Taylor expansion result (62) and the asymptotic result of (64), we get

$$d_{S_{\mathcal{Q}}}\{\mathbf{v}u; \mathbf{x^n}, \mathbf{y^n}\} = \begin{pmatrix} \check{u}^{\mathcal{Q}} - \hat{u}^{\mathcal{Q}} \\ \check{h}^{\mathcal{Q}} - \hat{h}^{\mathcal{Q}} \end{pmatrix}^T G_{\mathcal{Q}}^T S_{\mathcal{Q}}^{-1} G_{\mathcal{Q}} \begin{pmatrix} \check{u}^{\mathcal{Q}} - \hat{u}^{\mathcal{Q}} \\ \check{h}^{\mathcal{Q}} - \hat{h}^{\mathcal{Q}} \end{pmatrix} + o_{p,K}(1), \tag{65}$$

where the term $o_{p,K}(1)$ converges to zero uniformly over $|u| \leq K$ in probability.

(6) A normal approximation of weighting posterior distributions in the integral. As a result of the uniform bound on the over-fitting gap $d_{S_{\mathcal{Q}}}\{\mathbf{v}\theta; \mathbf{x^n}, \mathbf{y^n}\}$, the approximation error of replacing

$\pi_{\mathcal{P}}(\theta|\mathbf{x^n})$ in (47) by the normal density for $N(\hat{\theta}^{\mathcal{P}}, n^{-1}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1})$, denoted by $\varphi_{\mathcal{P},n}(\theta)$, is also bounded:

$$\left|\varrho_o^f(\theta_0|\psi_0,\mathbf{x^n},\mathbf{y^n}) - \int d_{S_{\mathcal{Q}}}\{\mathbf{v}\theta;\mathbf{x^n},\mathbf{y^n}\}\varphi_{\mathcal{P},n}(\theta)\mathrm{d}\theta\right| \\ \leq \int d_{S_{\mathcal{Q}}}\{\mathbf{v}\theta;\mathbf{x^n},\mathbf{y^n}\}|\pi_{\mathcal{P}}(\theta|\mathbf{x^n}) - \varphi_{\mathcal{P},n}(\theta)|\mathrm{d}\theta. \tag{66}$$

After changing variables, $\theta = \hat{\theta}^{\mathcal{P}} + \frac{u}{\sqrt{n}}$, the inequality (66) can be rewritten as

$$\left|\varrho_o^f(\theta_0|\psi_0,\mathbf{x^n},\mathbf{y^n}) - \int d_{S_{\mathcal{Q}}}\{\mathbf{v}u;\mathbf{x^n},\mathbf{y^n}\}\varphi_{\mathcal{P}}(u)\mathrm{d}u\right| \\ \leq \int d_{S_{\mathcal{Q}}}\{\mathbf{v}u;\mathbf{x^n},\mathbf{y^n}\}|\pi_{\mathcal{P}}(\hat{\theta}^{\mathcal{P}} + u/\sqrt{n}|\mathbf{x^n}) - \varphi_{\mathcal{P}}(u)|\mathrm{d}u, \tag{67}$$

where $\varphi_{\mathcal{P}}(u) = \sqrt{\det(\mathbf{I}_{\mathcal{P}}(\theta_0))/(2\pi)^{D_\Theta}}\exp\left[-\frac{1}{2}u^T\mathbf{I}_{\mathcal{P}}(\theta_0)u\right]$. From the inequalities (56) and (67), it follows that

$$\left|\varrho_o^f(\theta_0|\psi_0,\mathbf{x^n},\mathbf{y^n}) - \int d_{S_{\mathcal{Q}}}\{\mathbf{v}u;\mathbf{x^n},\mathbf{y^n}\}\varphi_{\mathcal{P}}(u)\mathrm{d}u\right| \\ \leq D_{\mathcal{Q},n}(\mathbf{x^n},\mathbf{y^n})\int |u-\hat{u}^{\mathcal{Q}}|^2|\pi_{\mathcal{P}}(\hat{\theta}^{\mathcal{P}} + u/\sqrt{n}|\mathbf{x^n}) - \varphi_{\mathcal{P}}(u)|\mathrm{d}u. \tag{68}$$

Convergence in relative entropy implies convergence in total variation distance due to Pinsker's inequality (see, e.g. Pinsker, 1964). Thus, Proposition 4 and the intermediate convergence result (46) imply that

$$\int |u-\hat{u}^{\mathcal{Q}}|^2|\pi_{\mathcal{P}}(\hat{\theta}^{\mathcal{P}} + u/\sqrt{n}|\mathbf{x^n}) - \varphi_{\mathcal{P}}(u)|\mathrm{d}u = o_p(1). \tag{69}$$

Therefore, because $D_{\mathcal{Q},n}\{\mathbf{x^n},\mathbf{y^n}\}$ defined in (56) is $O_p(1)$, it holds that

$$\varrho_o^f(\theta_0|\psi_0,\mathbf{x^n},\mathbf{y^n}) = \int d_{S_{\mathcal{Q}}}\{\mathbf{v}u;\mathbf{x^n},\mathbf{y^n}\}\varphi_{\mathcal{P}}(u)\mathrm{d}u + o_p(1). \tag{70}$$

Thus, as far as the asymptotic properties of $\varrho_o^f(\theta_0|\psi_0,\mathbf{x^n},\mathbf{y^n})$ are concerned, we need to only focus on the integral in (70).

(7) A Wald-type approximation for the over-fitting gap $d_{S_{\mathcal{Q}}}\{\mathbf{v}u;\mathbf{x^n},\mathbf{y^n}\}$. We define a Wald-type distance between u_1 and $\hat{u}_1^{\mathcal{Q}}$,

$$w_{S_{\mathcal{Q}}}\{\mathbf{v}u;\mathbf{x^n},\mathbf{y^n}\} = \begin{pmatrix}\hat{u}^{\mathcal{Q}}-u\\ \hat{h}^{\mathcal{Q}}-h\end{pmatrix}^T\Gamma_\Theta^T\mathbf{v}^T\left\{\mathbf{v}\Gamma_\Theta\left[G_{\mathcal{Q}}^TS_{\mathcal{Q}}^{-1}G_{\mathcal{Q}}\right]^{-1}\Gamma_\Theta^T\mathbf{v}^T\right\}^{-1}\mathbf{v}\Gamma_\Theta\begin{pmatrix}\hat{u}^{\mathcal{Q}}-u\\ \hat{h}^{\mathcal{Q}}-h\end{pmatrix}$$

Based on the definition of $\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)$ in (9), the Wald-type distance above can be rewritten as

$$w_{S_{\mathcal{Q}}}\{\mathbf{v}u;\mathbf{x^n},\mathbf{y^n}\} = \left(\hat{u}^{\mathcal{Q}}-u\right)^T\mathbf{v}^T\left[\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right]^{-1}\mathbf{v}\left(\hat{u}^{\mathcal{Q}}-u\right) \\ = \mathcal{R}(\hat{u}^{\mathcal{Q}},\hat{h}^{\mathcal{Q}};u,h)^T\Gamma_\Theta^T\mathbf{v}^T\left[\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right]^{-1}\mathbf{v}\Gamma_\Theta\mathcal{R}(\hat{u}^{\mathcal{Q}},\hat{h}^{\mathcal{Q}};u,h). \tag{71}$$

Our idea is to approximate $d_{S_{\mathcal{Q}}}\{\mathbf{v}u;\mathbf{x^n},\mathbf{y^n}\}$ by the Wald-type distance $w_{S_{\mathcal{Q}}}\{\mathbf{v}u;\mathbf{x^n},\mathbf{y^n}\}$. More precisely, we shall show that

$$\int d_{S_{\mathcal{Q}}}\{\mathbf{v}u;\mathbf{x^n},\mathbf{y^n}\}\varphi_{\mathcal{P}}(u)\mathrm{d}u = \int w_{S_{\mathcal{Q}}}\{\mathbf{v}u;\mathbf{x^n},\mathbf{y^n}\}\varphi_{\mathcal{P}}(u)\mathrm{d}u + o_p(1). \tag{72}$$

Because of the Gaussian tail of $\varphi_{\mathcal{P}}(u)$ and the quadratic bounds of $d_{S_{\mathcal{Q}}}$ and $w_{S_{\mathcal{Q}}}$, it suffices to show that for any large enough constant K,

$$\int_{|u|\leq K} d_{S_{\mathcal{Q}}}\{\mathbf{v}u;\mathbf{x^n},\mathbf{y^n}\}\varphi_{\mathcal{P}}(u)\mathrm{d}u = \int_{|u|\leq K} w_{S_{\mathcal{Q}}}\{\mathbf{v}u;\mathbf{x^n},\mathbf{y^n}\}\varphi_{\mathcal{P}}(u)\mathrm{d}u + o_p(1). \tag{73}$$

(8) Proof of the Wald-type approximation in (73). The starting point of the proof of Wald-type approximations goes back to the second-order Taylor expansion of $J_{n,S_{\mathcal{Q}}}(\check{\theta}^{\mathcal{Q}},\check{\psi}^{\mathcal{Q}})$ around $(\hat{\theta}^{\mathcal{Q}},\hat{\psi}^{\mathcal{Q}})$ and its large-sample approximation in (65). Because the constraint for $(\check{u}^{\mathcal{Q}},\check{h}^{\mathcal{Q}})$ is $\mathbf{v}u=\mathbf{v}\check{u}^{\mathcal{Q}}$, the following equality holds for any (u,h):

$$\mathcal{R}(\hat{u}^{\mathcal{Q}},\hat{h}^{\mathcal{Q}};u,h) = -\mathbf{v}\Gamma_{\Theta}\begin{pmatrix}\check{u}^{\mathcal{Q}}-\hat{u}^{\mathcal{Q}}\\ \check{h}^{\mathcal{Q}}-\hat{h}^{\mathcal{Q}}\end{pmatrix}. \tag{74}$$

Notice that $\mathbf{v}\Gamma_{\Theta}$ is not invertible. Thus, we cannot directly obtain an expression for $\begin{pmatrix}\check{u}^{\mathcal{Q}}-\hat{u}^{\mathcal{Q}}\\ \check{h}^{\mathcal{Q}}-\hat{h}^{\mathcal{Q}}\end{pmatrix}$ through (74). Instead, consider the first-order Taylor expansion of $\nabla\widehat{J}_{n,S_{\mathcal{Q}}}(\check{\theta}^{\mathcal{Q}},\check{\psi}^{\mathcal{Q}})$ around $(\hat{\theta}^{\mathcal{Q}},\hat{\psi}^{\mathcal{Q}})$,

$$\nabla\widehat{J}_{n,S_{\mathcal{Q}}}(\check{\theta}^{\mathcal{Q}},\check{\psi}^{\mathcal{Q}}) = n^{-1}\nabla^2\widehat{J}_{n,S_{\mathcal{Q}}}(\theta'_u,\psi'_u)\begin{pmatrix}\check{u}^{\mathcal{Q}}-\hat{u}^{\mathcal{Q}}\\ \check{h}^{\mathcal{Q}}-\hat{h}^{\mathcal{Q}}\end{pmatrix}, \tag{75}$$

where (θ'_u,ψ'_u) lies on the segment between $(\hat{\theta}^{\mathcal{P}}+\check{u}^{\mathcal{Q}}/\sqrt{n},\ \hat{\psi}^{\mathcal{P}}+\check{h}^{\mathcal{Q}}/\sqrt{n})$ and $(\hat{\theta}^{\mathcal{P}}+\hat{u}^{\mathcal{Q}}/\sqrt{n},\ \hat{\psi}^{\mathcal{P}}+\hat{h}^{\mathcal{Q}}/\sqrt{n})$. Then, it follows from (64) that

$$\nabla\widehat{J}_{n,S_{\mathcal{Q}}}(\check{\theta}^{\mathcal{Q}},\check{\psi}^{\mathcal{Q}}) = 2G_{\mathcal{Q}}^TS_{\mathcal{Q}}^{-1}G_{\mathcal{Q}}\begin{pmatrix}\check{u}^{\mathcal{Q}}-\hat{u}^{\mathcal{Q}}\\ \check{h}^{\mathcal{Q}}-\hat{h}^{\mathcal{Q}}\end{pmatrix} + o_{p,K}(1), \tag{76}$$

where the term $o_{p,K}(1)$ converges to zero uniformly over $|u|\leq K$ in probability. Thus,

$$\begin{pmatrix}\check{u}^{\mathcal{Q}}-\hat{u}^{\mathcal{Q}}\\ \check{h}^{\mathcal{Q}}-\hat{h}^{\mathcal{Q}}\end{pmatrix} = \left(2G_{\mathcal{Q}}^TS_{\mathcal{Q}}^{-1}G_{\mathcal{Q}}\right)^{-1}\nabla\widehat{J}_{n,S_{\mathcal{Q}}}(\check{\theta}^{\mathcal{Q}},\check{\psi}^{\mathcal{Q}}) + o_{p,K}(1). \tag{77}$$

We now represent $\nabla\widehat{J}_{n,S_{\mathcal{Q}}}(\check{\theta}^{\mathcal{Q}},\check{\psi}^{\mathcal{Q}})$ in (77) by the Lagrange multiplier Λ_n using the first-order condition of constrained GMM estimators as in (60), and plug the resulting expression for $\begin{pmatrix}\check{u}^{\mathcal{Q}}-\hat{u}^{\mathcal{Q}}\\ \check{h}^{\mathcal{Q}}-\hat{h}^{\mathcal{Q}}\end{pmatrix}$ back into (74) to get

$$\mathcal{R}(\hat{u}^{\mathcal{Q}},\hat{h}^{\mathcal{Q}};u,h) = -\mathbf{v}\Gamma_{\Theta}\left(2G_{\mathcal{Q}}^TS_{\mathcal{Q}}^{-1}G_{\mathcal{Q}}\right)^{-1}(\mathbf{v}\Gamma_{\Theta})^T\Lambda_n + o_{p,K}(1). \tag{78}$$

Thus,

$$\Lambda_n = -\left\{\mathbf{v}\Gamma_{\Theta}\left(2G_{\mathcal{Q}}^TS_{\mathcal{Q}}^{-1}G_{\mathcal{Q}}\right)^{-1}(\mathbf{v}\Gamma_{\Theta})^T\right\}^{-1}\mathcal{R}(\hat{u}^{\mathcal{Q}},\hat{h}^{\mathcal{Q}};u,h) + o_{p,K}(1). \tag{79}$$

Combining (60), (77), and (79), we get

$$\begin{pmatrix}\check{u}^{\mathcal{Q}}-\hat{u}^{\mathcal{Q}}\\ \check{h}^{\mathcal{Q}}-\hat{h}^{\mathcal{Q}}\end{pmatrix} = \left(G_{\mathcal{Q}}^TS_{\mathcal{Q}}^{-1}G_{\mathcal{Q}}\right)^{-1}\Gamma_{\Theta}^T\mathbf{v}^T\left[\mathbf{v}\Gamma_{\Theta}\left(G_{\mathcal{Q}}^TS_{\mathcal{Q}}^{-1}G_{\mathcal{Q}}\right)^{-1}\Gamma_{\Theta}^T\mathbf{v}^T\right]^{-1}\mathcal{R}(\hat{u}^{\mathcal{Q}},\hat{h}^{\mathcal{Q}};u,h) + o_{p,K}(1).$$

Finally, we plug the relationship above into (65) and obtain the following approximation,

$$d_{S_{\mathcal{Q}}}\{\mathbf{v}u;\mathbf{x^n},\mathbf{y^n}\} = \mathcal{R}(\hat{u}^{\mathcal{Q}},\hat{h}^{\mathcal{Q}};u,h)^T\Gamma_{\Theta}^T\mathbf{v}^T\left[\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right]^{-1}\mathbf{v}\Gamma_{\Theta}\mathcal{R}(\hat{u}^{\mathcal{Q}},\hat{h}^{\mathcal{Q}};u,h) + o_{p,K}(1). \tag{80}$$

Comparing (71) and (80), we see that we have established the large-sample relationship in (73).

As a final step of the proof, we derive the asymptotic distribution of $\int w_{S_{\mathcal{Q}}}\{\mathbf{v}u; \mathbf{x^n}, \mathbf{y^n}\}\varphi_{\mathcal{P}}(u)\mathrm{d}u$.

(9) Asymptotic distribution of $\int w_{S_{\mathcal{Q}}}\{\mathbf{v}u; \mathbf{x^n}, \mathbf{y^n}\}\varphi_{\mathcal{P}}(u)\mathbf{d}u$. Consider the decomposition

$$w_{S_0}\{\mathbf{v}u; \mathbf{x^n}, \mathbf{y^n}\} = u^T\mathbf{v}^T\left[\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right]^{-1}\mathbf{v}u \tag{81}$$

$$+ 2(\hat{u}^{\mathcal{Q}})^T\mathbf{v}^T\left[\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right]^{-1}\mathbf{v}u \tag{82}$$

$$+ (\hat{u}^{\mathcal{Q}})^T\mathbf{v}^T\left[\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right]^{-1}\mathbf{v}\hat{u}^{\mathcal{Q}}. \tag{83}$$

The integral of term (82) over the Gaussian density $\varphi_{\mathcal{P}}(u)$ is zero, while the integral of term (83) over $\varphi_{\mathcal{P}}(u)$ is just itself. The integral of term (81) over $\varphi_{\mathcal{P}}(u)$ is

$$\begin{aligned}\int \mathbf{tr}\left\{\left[\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right]^{-1}\mathbf{v}uu^T\mathbf{v}^T\right\}\varphi_{\mathcal{P}}(u)\mathrm{d}u &= \mathbf{tr}\left[\left(\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right)^{-1}\left(\mathbf{v}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}^T\right)\right] \\ &= \varrho^{\mathbf{v}}(\theta_0|\psi_0).\end{aligned}$$

Thus,

$$\int w_{S_{\mathcal{Q}}}\{\mathbf{v}u; \mathbf{x^n}, \mathbf{y^n}\}\varphi_{\mathcal{P}}(u)\mathrm{d}u = \varrho^{\mathbf{v}}(\theta_0|\psi_0) + (\hat{u}^{\mathcal{Q}})^T\mathbf{v}^T\left[\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right]^{-1}\mathbf{v}\hat{u}^{\mathcal{Q}}. \tag{84}$$

To derive the asymptotic approximation of $(\hat{u}^{\mathcal{Q}})^T\mathbf{v}^T\left[\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right]^{-1}\mathbf{v}\hat{u}^{\mathcal{Q}}$, we appeal to the standard large-sample approximations of GMM; see, for example, Hansen (1982, Theorem 3.1). Specifically, we have

$$\mathbf{v}\hat{u}^{\mathcal{Q}} = -\mathbf{v}\left[\Gamma_{\Theta}\left(G_{\mathcal{Q}}^T S_{\mathcal{Q}}^{-1} G_{\mathcal{Q}}\right)^{-1} G_{\mathcal{Q}}^T S_{\mathcal{Q}}^{-1} - \left(G_{\mathcal{P}}^T S_{\mathcal{P}}^{-1} G_{\mathcal{P}}\right)^{-1} G_{\mathcal{P}} S_{\mathcal{P}}^{-1}\Gamma_{\mathcal{P}}\right]\sqrt{n}\hat{g}_{\mathcal{Q},n}(\theta_0, \psi_0) + o_p(1).$$

Thus, $\hat{u}^{\mathcal{Q}}$ is asymptotically normally distributed

$$\underset{n\to\infty}{\mathrm{wlim}}\, \mathbf{v}\hat{u}^{\mathcal{Q}} = \mathbf{v}\Upsilon_{\mathcal{Q}} Z$$

where $Z \sim N(0, I_{D_{\mathcal{Q}}})$ and

$$\Upsilon_{\mathcal{Q}} \equiv -\left[\Gamma_{\Theta}\left(G_{\mathcal{Q}}^T S_{\mathcal{Q}}^{-1} G_{\mathcal{Q}}\right)^{-1} G_{\mathcal{Q}}^T S_{\mathcal{Q}}^{-1} - \left(G_{\mathcal{P}}^T S_{\mathcal{P}}^{-1} G_{\mathcal{P}}\right)^{-1} G_{\mathcal{P}}^T S_{\mathcal{P}}^{-1}\Gamma_{\mathcal{P}}\right] S_{\mathcal{Q}}^{1/2}. \tag{85}$$

The Continuous Mapping Theorem implies that

$$\underset{n\to\infty}{\mathrm{wlim}} (\hat{u}^{\mathcal{Q}})^T\mathbf{v}^T\left[\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right]^{-1}\mathbf{v}\hat{u}^{\mathcal{Q}} = Z^T\Upsilon_{\mathcal{Q}}^T\mathbf{v}^T\left[\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right]^{-1}\mathbf{v}\Upsilon_{\mathcal{Q}} Z. \tag{86}$$

Consider the Singular Value Decomposition (SVD):

$$\Upsilon_{\mathcal{Q}}^T\mathbf{v}^T\left[\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right]^{-1/2} = \mathbf{U}\mathbf{\Sigma}\widetilde{\mathbf{U}}^T,$$

where $\mathbf{U}$ is a $D_{\mathcal{Q}} \times D_{\mathcal{Q}}$ orthonormal matrix, $\widetilde{\mathbf{U}}$ is a $D_f \times D_f$ orthonormal matrix, and $\mathbf{\Sigma}$ is a $D_{\mathcal{Q}} \times D_f$ diagonal matrix with singular values on the diagonal line. This allows us to rewrite (86) as

$$\underset{n\to\infty}{\mathrm{wlim}} (\hat{u}^{\mathcal{Q}})^T\mathbf{v}^T\left[\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right]^{-1}\mathbf{v}\hat{u}^{\mathcal{Q}} = \left(\mathbf{U}^T Z\right)^T\mathbf{\Sigma}\mathbf{\Sigma}^T\mathbf{U}^T Z = \sigma_1 z_1^2 + \cdots + \sigma_{D_f} z_{D_f}^2, \tag{87}$$

where $\mathbf{U}^T Z = (z_1, \cdots, z_{D_f}, z_{D_f+1}, \cdots, z_{D_\Theta})^T$ contains D_Θ i.i.d. standard normal random variables, and $\sigma_1, \cdots, \sigma_{D_f}$ are the nonzero diagonal elements of $\mathbf{\Sigma}\mathbf{\Sigma}^T$. In fact, $\sigma_1, \cdots, \sigma_{D_f}$ are the eigenvalues of the matrix

$$\widetilde{\mathbf{U}}\mathbf{\Sigma}^T\mathbf{\Sigma}\widetilde{\mathbf{U}}^T = \left[\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right]^{-1/2}\mathbf{v}\Upsilon_{\mathcal{Q}}\Upsilon_{\mathcal{Q}}^T\mathbf{v}^T\left[\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right]^{-1/2} \tag{88}$$

From the definition of $\Upsilon_{\mathcal{Q}}$ in (85), it follows that

$$\Upsilon_{\mathcal{Q}}\Upsilon_{\mathcal{Q}}^T = \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1} - \mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}. \tag{89}$$

We use the following two identities in deriving (89):

$$\Gamma_{\mathcal{P}} S_{\mathcal{Q}} \Gamma_{\mathcal{P}}^T = S_{\mathcal{P}} \quad \text{and} \quad \Gamma_{\mathcal{P}} G_{\mathcal{Q}} = [G_{\mathcal{P}}, O_{D_{\mathcal{P}}\times D_{\Psi}}] = [G_{\mathcal{P}}, O_{D_{\mathcal{P}}\times D_{\Psi}}]\,\Gamma_{\Theta}^T\Gamma_{\Theta}. \tag{90}$$

Plugging (89) back into (88) yields

$$\begin{aligned}\widetilde{\mathbf{U}}\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\widetilde{\mathbf{U}}^T &= \left[\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right]^{-1/2}\left[\mathbf{v}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}^T\right]\left[\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right]^{-1/2} - I_{D_f}\\ &= \Pi_0(\mathbf{v}) - I_{D_f},\end{aligned}$$

where $\Pi_0(\mathbf{v})$ is the Fisher-information-ratio matrix defined in (11). Thus, the eigenvalues σ_i's are

$$\sigma_i = \lambda_i(\mathbf{v}) - 1, \quad \text{for all } i = 1, \cdots, D_f, \tag{91}$$

where $\lambda_i(\mathbf{v})$'s are eigenvalues of $\Pi_0(\mathbf{v})$ as in Proposition 1. Then, according to (84), (87), and (91),

$$\operatorname*{wlim}_{n\to\infty}\int w_{S_{\mathcal{Q}}}\{\mathbf{v}u;\mathbf{x}^{\mathbf{n}},\mathbf{y}^{\mathbf{n}}\}\varphi_{\mathcal{P}}(u)\mathrm{d}u = \varrho^{\mathbf{v}}(\theta_0|\psi_0) + \sum_{i=1}^{D_f}(\lambda_i(\mathbf{v})-1)\,\chi^2_{1,i}, \tag{92}$$

where $\chi^2_{1,i}$'s are i.i.d. chi-squared random variables with 1 degree of freedom.

B.3 Proof of The Corollaries

Corollary 1 follows immediately from Theorem 1. Corollary 2 follows from Theorem 1 of Nakagawa (2005). Specifically, the Laplace-Stieltjes transform of the cumulative distribution function of $\operatorname{wlim}_{n\to\infty}\varrho_o^f(\theta_0|\psi_0,\mathbf{x}^{\mathbf{n}},\mathbf{y}^{\mathbf{n}}) = \varrho^{\mathbf{v}}(\theta_0|\psi_0) + \sum_{i=1}^{D_f}(\lambda_i(\mathbf{v})-1)\,\chi^2_{1,i}$ is

$$\mathcal{M}(z) \equiv \mathbb{E}e^{-z\left\{\varrho^{\mathbf{v}}(\theta_0|\psi_0)+\sum_{i=1}^{D_f}(\lambda_i(\mathbf{v})-1)\chi^2_{1,i}\right\}} = e^{-z\varrho^{\mathbf{v}}(\theta_0|\psi_0)}\prod_{i=1}^{D_f}\left[1+2z\,(\lambda_i(\mathbf{v})-1)\right]^{-1/2}. \tag{93}$$

Let $\Re z$ denote the real part of z. Thus, the abscissa of convergence of $\mathcal{M}(z)$ is equal to $-\frac{1}{2(\lambda_1(\mathbf{v})-1)}$ where $\lambda_1(\mathbf{v})$ is the largest eigenvalue; that is, when $\Re z > -\frac{1}{2(\lambda_1(\mathbf{v})-1)}$, the transform $\mathcal{M}(z)$ converges, and when $\Re z < -\frac{1}{2(\lambda_1(\mathbf{v})-1)}$, the transform $\mathcal{M}(z)$ diverges. Therefore, according to Theorem 1 of Nakagawa (2005), the tail probability has the convergence property stated in the corollary.

C Proof of Propositions on Fisher Fragility Measure

C.1 Proof of Proposition 1

This result follows from the fact that

$$\begin{aligned}&\mathbf{tr}\left[\left(\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right)^{-1}\left(\mathbf{v}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}^T\right)\right]\\ &\quad = \mathbf{tr}\left[\left(\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right)^{-1/2}\left(\mathbf{v}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}^T\right)\left(\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right)^{-1/2}\right],\end{aligned}$$

and the fact that the trace of a symmetric matrix is equal to the sum of all of its eigenvalues. Because $g_{\mathcal{P}}(\theta;\mathbf{x_t})$ is part of $g_{\mathcal{Q}}(\theta,\psi;\mathbf{x_t},\mathbf{y_t})$, according to Hansen (1982), the asymptotic covariance matrices satisfy $\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1} \leq \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}$. Thus, for any full-rank matrix $\mathbf{v}$, it holds that $\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T \leq \mathbf{v}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}^T$. As a result, the smallest eigenvalue is no less than one.

C.2 Proof of Proposition 2

We define $\mathbf{u} = \mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1/2}\mathbf{v}$, and rewrite $\varrho^D(\theta_0|\psi_0)$ as

$$\begin{aligned}\varrho^D(\theta_0|\psi_0) &= \max_{\mathbf{u}\in\mathbb{R}^{D_\Theta\times D},\mathbf{Rank}(\mathbf{u})=D} \mathbf{tr}\left[\left(\mathbf{u}^T\mathbf{u}\right)^{-1}\left(\mathbf{u}^T\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{1/2}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{1/2}\mathbf{u}\right)\right] \\ &= \max_{\mathbf{u}\in\mathbb{R}^{D_\Theta\times D},\mathbf{Rank}(\mathbf{u})=D} \mathbf{tr}\left[\mathbf{u}\left(\mathbf{u}^T\mathbf{u}\right)^{-1}\mathbf{u}^T\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{1/2}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{1/2}\right]\end{aligned}$$

The linear operator $\mathfrak{P}_{\mathbf{u}} \equiv \mathbf{u}\left(\mathbf{u}^T\mathbf{u}\right)^{-1}\mathbf{u}^T$ is the projection operator onto the subspace spanned by the column vectors of $\mathbf{u}$. The projection operator can be equivalently expressed in terms of the orthonormal column vectors lying in the subspace spanned by $\mathbf{u}$. Thus, without loss of generality, we can assume that the column vectors of $\mathbf{u}$ are orthonormal vectors, i.e. $\mathbf{u}^T\mathbf{u}$ is a D-dimensional identity matrix. Then, it follows that

$$\begin{aligned}\varrho^D(\theta_0|\psi_0) &= \max_{\mathbf{u}\in\mathbb{R}^{D_\theta\times D},\mathbf{Rank}(\mathbf{u})=D,\mathbf{u}^T\mathbf{u}=I} \mathbf{tr}\left[\mathbf{u}\mathbf{u}^T\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{1/2}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{1/2}\right] \\ &= \max_{\mathbf{u}\in\mathbb{R}^{D_\theta\times D},\mathbf{Rank}(\mathbf{u})=D,\mathbf{u}^T\mathbf{u}=I} \mathbf{tr}\left[\mathbf{u}^T\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{1/2}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{1/2}\mathbf{u}\right] \\ &= \max_{\mathbf{u}\in\mathbb{R}^{D_\theta\times D},\mathbf{Rank}(\mathbf{u})=D,\mathbf{u}^T\mathbf{u}=I} \sum_{i=1}^{D} u_i^T\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{1/2}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{1/2}u_i \\ &= \lambda_1 + \lambda_2 + \cdots + \lambda_D.\end{aligned}$$

The argmax matrix is $\mathbf{u}^* = [e_1^*, e_2^*, \cdots, e_D^*]$, whose column vectors are the corresponding eigenvectors. Thus, the worst-case matrix is $\mathbf{v}^* = [v_1^*, v_2^*, \cdots, v_D^*]$ with $v_i^* = \mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{1/2}e_i^*$. The eigenvalues are

$$\lambda_i = \frac{(v_i^*)^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}v_i^*}{(v_i^*)^T\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}v_i^*}, \quad i = 1, \cdots, D \tag{94}$$

Furthermore, the eigenvalue problem above is equivalent to

$$\lambda_i = \frac{(\hat{v}_i^*)^T\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)\hat{v}_i^*}{(\hat{v}_i^*)^T\mathbf{I}_{\mathcal{P}}(\theta_0)\hat{v}_i^*}, \tag{95}$$

where $\hat{v}_i^* = \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1/2}\hat{e}_i^*$, with $\hat{e}_i^*$ being the corresponding eigenvector of $\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1/2}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1/2}$.

C.3 Proof of Proposition 3

From Proposition 2, it holds that

$$\varrho^{D_1}(\theta_0|\psi_0) = \lambda_1 + \cdots + \lambda_{D_1}, \text{ and } \varrho^{D_2}(\theta_0|\psi_0) = \lambda_1 + \cdots + \lambda_{D_2}. \tag{96}$$

As the eigenvalues are nonnegative, it must be that $\varrho^{D_1}(\theta_0|\psi_0) \leq \varrho^{D_2}(\theta_0|\psi_0)$ when $D_1 \leq D_2$.

D Derivations of The Disaster Risk Model

We first show how to derive the Euler equation, and then we show how to obtain the Fisher fragility measure $\varrho(p, \xi)$.

D.1 The Euler Equation

The total return of market equity from t to $t+1$ is $e^{r_{M,t+1}}$, which is unknown at t, and the total return of the risk-free bond from t to $t+1$ is $e^{r_{f,t}}$, which is known at t. Thus, the excess log return of equity is $r_{t+1} = r_{M,t+1} - r_{f,t}$. The state-price density is $\Lambda_t = \delta_{\mathrm{D}}^t c_t^{-\gamma_{\mathrm{D}}}$, and the inter-temporal marginal rate of substitution is $\Lambda_{t+1}/\Lambda_t = \delta_{\mathrm{D}} e^{-\gamma_{\mathrm{D}} g_{t+1}}$. The Euler equations for the risk-free rate and the market equity return are

$$1 = \mathbb{E}_t \left[\frac{\Lambda_{t+1}}{\Lambda_t} e^{r_{M,t+1}} \right] \quad \text{and} \quad e^{-r_{f,t}} = \mathbb{E}_t \left[\frac{\Lambda_{t+1}}{\Lambda_t} \right]. \tag{97}$$

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

$$\mathbb{E}_t \left[\frac{\Lambda_{t+1}}{\Lambda_t} \right] = \mathbb{E}_t \left[\frac{\Lambda_{t+1}}{\Lambda_t} e^{r_{t+1}} \right]. \tag{98}$$

The left-hand side of (98) is equal to

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

and the right-hand side of (98) is equal to

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

Thus, the Euler equation (98) can be rewritten as

$$(1-p) e^{-\gamma_{\mathrm{D}} \mu + \frac{1}{2} \gamma_{\mathrm{D}}^2 \sigma^2} \left[e^{\eta + \frac{\tau^2}{2} - \gamma_{\mathrm{D}} \rho \sigma \tau} - 1 \right] = p \Delta(\xi), \tag{99}$$

where

$$\Delta(\xi) = \xi \left(\frac{e^{\gamma_{\mathrm{D}} \underline{v}}}{\xi - \gamma_{\mathrm{D}}} - \frac{e^{\frac{\varsigma^2}{2} + (\gamma_{\mathrm{D}} - b)\underline{v}}}{\xi + b - \gamma_{\mathrm{D}}} \right).$$

Rearranging terms in (99) leads to the Euler equation in (30). Using the Taylor expansion, we have the following approximation,

$$e^{\eta + \frac{\tau^2}{2} - \gamma_{\mathrm{D}} \rho \sigma \tau} - 1 \approx \eta + \frac{\tau^2}{2} - \gamma_{\mathrm{D}} \rho \sigma \tau, \tag{100}$$

which, combined with (99), gives the approximated Euler equation in (31).

D.2 Fisher fragility measure

The joint probability density for a rare disaster event z and size v in the baseline model is

$$f_{\mathcal{P}}(z, v | p, \xi) = p^z (1-p)^{1-z} \left[\mathbf{1}\{v > \underline{v}\} \xi \exp\left\{ -\xi (v - \underline{v}) \right\} \right]^z \delta(v)^{1-z}, \tag{101}$$

where $\delta(\cdot)$ is the Dirac delta function. The corresponding Fisher information matrix is

$$\mathbf{I}_{\mathcal{P}}(\theta) = \begin{bmatrix} \frac{1}{p(1-p)} & 0 \\ 0 & \frac{p}{\xi^2} \end{bmatrix}. \tag{102}$$

Next, the probability density function $f_{\mathcal{Q}}(z, v, r, u|\theta, \phi)$ for the structural model is

$$\begin{aligned} &f_{\mathcal{Q}}(z, v, r, u|\theta, \phi) = p^z(1-p)^{1-z} \\ &\times \left[\frac{1}{2\pi\sigma\tau\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(u-\mu)^2}{\sigma^2} + \frac{(r-\eta(\theta,\phi))^2}{\tau^2} - \frac{2\rho(u-\mu)(r-\eta(\theta,\phi))}{\sigma\tau}\right]\right\}\right]^{1-z} \\ &\times \left[\mathbf{1}\{v > \underline{v}\}\xi \exp\{-\xi(v-\underline{v})\} \frac{1}{\sqrt{2\pi}\varsigma} \exp\left\{-\frac{1}{2\varsigma^2}(r-bv)^2\right\}\right]^z \mathbf{1}\{\eta(\theta,\phi) > \underline{\eta}^*, \xi > \gamma_{\mathrm{D}}\}, \end{aligned}$$

where

$$\eta(\theta, \phi) \equiv \gamma_{\mathrm{D}}\rho\sigma\tau - \frac{\tau^2}{2} + \ln\left[1 + e^{\gamma_{\mathrm{D}}\mu - \frac{\gamma_{\mathrm{D}}^2\sigma^2}{2}}\xi\left(\frac{e^{\gamma_{\mathrm{D}}\underline{v}}}{\xi - \gamma_{\mathrm{D}}} - e^{\frac{1}{2}\varsigma^2}\frac{e^{(\gamma_{\mathrm{D}}-b)\underline{v}}}{\xi + b - \gamma_{\mathrm{D}}}\right)\frac{p}{1-p}\right]. \tag{103}$$

From $f_{\mathcal{Q}}(z, v, r, u|\theta, \phi)$ we can derive the Fisher information for (p, ξ) under the full structural model. The Fisher fragility measure can then be computed numerically.

We can also derive the closed-form solution for the Fisher fragility measure in (33) if we use the approximated Euler equation in (31). In this case, using the notation introduced by (32) and (34), we can express the Fisher information for (p, ξ) under the full structural model as

$$\mathbf{I}_{\mathcal{Q}}(\theta) \approx \begin{bmatrix} \frac{1}{p(1-p)} + \frac{\Delta(\xi)^2}{(1-\rho^2)\tau^2}\frac{e^{2\gamma_{\mathrm{D}}\mu - \gamma_{\mathrm{D}}^2\sigma^2}}{(1-p)^3} & \frac{p}{(1-\rho^2)\tau^2}\frac{e^{2\gamma_{\mathrm{D}}\mu - \gamma_{\mathrm{D}}^2\sigma^2}}{(1-p)^2}\Delta(\xi)\dot{\Delta}(\xi) \\ \frac{p}{(1-\rho^2)\tau^2}\frac{e^{2\gamma_{\mathrm{D}}\mu - \gamma_{\mathrm{D}}^2\sigma^2}}{(1-p)^2}\Delta(\xi)\dot{\Delta}(\xi) & \frac{p}{\xi^2} + \frac{\dot{\Delta}(\xi)^2}{(1-\rho^2)\tau^2}e^{2\gamma_{\mathrm{D}}\mu - \gamma_{\mathrm{D}}^2\sigma^2}\frac{p^2}{1-p} \end{bmatrix}. \tag{104}$$

Following Proposition 2, the worst-case Fisher fragility is the largest eigenvalue of the matrix $\Pi_0(I_{D_\Theta}) \equiv \mathbf{I}_{\mathcal{Q}}(\theta_0)^{1/2}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{I}_{\mathcal{Q}}(\theta_0)^{1/2}$, and it is also the largest eigenvalue of $\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1/2}\mathbf{I}_{\mathcal{Q}}(\theta_0)\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1/2}$. In this case, the eigenvalues and eigenvectors are available in closed form. This gives us the formula for $\varrho(p, \xi)$ and $\varrho^1(p, \xi)$ in (33). The minimum Fisher fragility in this case is 1.