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

Hui Chen   Winston Wei Dou   Leonid Kogan*

April 15, 2017

Contents

*Chen: MIT Sloan and NBER (huichen@mit.edu). Dou: Wharton (wdou@wharton.upenn.edu). Kogan: MIT Sloan and NBER (lkogan@mit.edu).

1 Information-Theoretic Interpretation for Model Fragility

In Chen, Dou, and Kogan (2017), we provide an econometric justification for our Fisher fragility measure: it essentially quantifies the over-fitting tendency of the functional-form specifications for certain structural components of a model. In this section, we formalize the intuition that excessive informativeness of cross-equation restrictions is fundamentally associated with model fragility. More precisely, we formally show that structural economic models are fragile when the cross-equation restrictions appear excessively informative about certain combinations of model parameters that are otherwise difficult to estimate (we refer to such parameter combinations as "dark matter"). To do so, we introduce a measure of informativeness of the cross-equation restrictions, and express it in terms of the effective sample size. This informational measure of cross-equation restrictions is interpretable in finite samples and provides economic intuition for the Fisher fragility measure.

We develop our analysis in a Bayesian framework. Starting with a prior on (θ, ψ), denoted by $\pi(\theta, \psi)$, we obtain posterior distributions through the baseline model and the structural model, respectively. Then, the discrepancy between the two posteriors shows how the cross-equation restrictions affect the inference about θ.

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 $\mathbf{x^n} \equiv (\mathbf{x_1}, \cdots, \mathbf{x_n})$ 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 $(\mathbf{x^n}, \mathbf{y^n}) \equiv \{(\mathbf{x_t}, \mathbf{y_t}) : t = 1, \cdots, n\}$, which we denote 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. The seminal paper by Hansen and Singleton (1982) pioneers the literature of applying GMM to evaluate rational expectation asset pricing models. Specifically, we assume that the model builder is concerned with the model's in-sample and out-of-sample performances as represented by a set of moment conditions,[1] based on a $D_{\mathbb{Q}} \times 1$ vector of functions $g_{\mathbb{Q}}(\theta, \psi; \mathbf{x}, \mathbf{y})$ of

[1]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 extension 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 overfitting. To simplify the comparison with the Fisher fragility measure, we chose to use the original GMM framework.

data observations $(\mathbf{x}_t, \mathbf{y}_t)$ and the parameter vectors θ and ψ satisfying the following conditions:

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

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 whole vector of moment functions $g_{\mathcal{Q}}(\theta, \psi; \mathbf{x_t}, \mathbf{y_t})$. Thus, the baseline moments can be represented by the full set of moments weighted by a special matrix:

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

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 is

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

Denote the empirical moment conditions for the full model and the baseline model by

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

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 often referred to as the J-distances, and $S_{\mathcal{Q}}$ and $S_{\mathcal{P}}$ have the following explicit formulae (see Hansen, 1982),

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

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

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

We use GMM to evaluate model performance because of the concern of likelihood mis-specification. The GMM approach gives the model builder flexibility to choose which aspects of the model to emphasize when estimating model 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 some further notation. We denote the GMM Fisher information matrix for the baseline model as $\mathbf{I}_{\mathcal{P}}(\theta)$ (see Hansen, 1982; Hahn, Newey, and Smith, 2011), and

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

where $G_{\mathcal{P}}(\theta) \equiv \mathbb{E}\left[\nabla g_{\mathcal{P}}(\theta; \mathbf{x_t})\right]$,and for brevity, we denote $G_{\mathcal{P}} \equiv G_{\mathcal{P}}(\theta_0)$. We denote the analog 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 for brevity, we denote $G_{\mathcal{Q}} \equiv G_{\mathcal{Q}}(\theta_0, \psi_0)$. Computing the expectation $G_{\mathcal{Q}}(\theta)$ and $G_{\mathcal{Q}}(\theta, \psi)$ requires knowing the distribution $\mathbb{Q}$. In cases when $\mathbb{Q}$ is unknown, $G_{\mathcal{P}}(\theta)$ in (7) and $G_{\mathcal{Q}}(\theta, \psi)$ in (8) can be replaced by their consistent estimators $\nabla g_{\mathcal{P}}(\theta; \mathbf{x_t})$ and $\nabla g_{\mathcal{Q}}(\theta, \psi; \mathbf{x_t}, \mathbf{y_t})$. For the full model $\mathcal{Q}$, we will focus on its implied Fisher information matrix $\mathbf{I}_{\mathcal{Q}}(\theta|\psi)$:

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

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

$$\mathbf{I}_{\mathcal{Q}}(\theta, \psi) = \begin{bmatrix} \mathbf{I}_{\mathcal{Q}}^{(1,1)}(\theta, \psi), & \mathbf{I}_{\mathcal{Q}}^{(1,2)}(\theta, \psi) \\ \mathbf{I}_{\mathcal{Q}}^{(2,1)}(\theta, \psi), & \mathbf{I}_{\mathcal{Q}}^{(2,2)}(\theta, \psi) \end{bmatrix}, \tag{10}$$

where $\mathbf{I}_{\mathcal{Q}}^{(1,1)}(\theta, \psi)$ is the $D_\Theta \times D_\Theta$ information matrix corresponding to baseline parameters θ, $\mathbf{I}_{\mathcal{Q}}^{(2,2)}(\theta, \psi)$ is the $D_\Psi \times D_\Psi$ information matrix corresponding to nuisance parameters ψ, 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

$$\mathbf{I}_{\mathcal{Q}}(\theta|\psi) = \mathbf{I}_{\mathcal{Q}}^{(1,1)}(\theta, \psi) - \mathbf{I}_{\mathcal{Q}}^{(1,2)}(\theta, \psi) \mathbf{I}_{\mathcal{Q}}^{(2,2)}(\theta, \psi)^{-1} \mathbf{I}_{\mathcal{Q}}^{(1,2)}(\theta, \psi)^T, \tag{11}$$

which generally is not equal to the Fisher information sub-matrix $\mathbf{I}_{\mathcal{Q}}^{(1,1)}(\theta, \psi)$ for baseline parameters θ, except the special case in which $\mathbf{I}_{\mathcal{Q}}^{(1,2)}(\theta, \psi) = 0$, i.e. the knowledge of θ and that of ψ are not informative about each other. We assume that the information matrices are nonsingular in this paper (Assumption A4 in Appendix 1.4).

We first introduce a moment-based Bayesian method with limited-information likelihoods in Subsection 1.1. Then, we introduce the definition of effective sample size to gauge the informativeness of the cross-equation restrictions and state the main results in Subsection 1.2. Third, we predefine the necessary special notations in Subsection 1.3. Fourth, we introduce the standard regularity conditions in Subsection 1.4. Fifth, we prove the basic lemmas in Subsection 1.5, which are themselves interesting and general. Sixth, in Subsection 1.6, we state and prove propositions which serve as intermediate steps for the proof of the main results. Finally, the main results stated in Subsection 1.2 are proved in Subsection 1.9 - 1.11.

1.1 Bayesian Analysis with Limited-Information Likelihoods

When likelihood-based methods are difficult or unreliable, one robust tool for the econometrician to derive the posteriors of baseline model $\pi_{\mathcal{P}}(\theta|\mathbf{x^n})$ and full structural model $\pi_{\mathcal{Q}}(\theta, \psi|\mathbf{x^n}, \mathbf{y^n})$ is the limited-information likelihood (LIL). It relies on certain moment conditions used for model evaluation. Of course, the full likelihood function can be used when likelihood methods are possible.[2] Here, restricted to the moment constraints, we embed a likelihood that is closest to the underlying true distribution into the Bayesian paradigm. We use the Kullback-Leibler divergence (also known as the relative entropy) to gauge the discrepancy between probability measures.[3]

We first focus on the limited-information likelihood for the full structural model. The large-sample results of Kim (2002) provide an asymptotic Bayesian interpretation of GMM, using the exponential quadratic form and ignoring the finite-sample validity and a valid parametric family for likelihoods. However, our information-theoretic rationale to model fragility indeed requires a valid finite-sample interpretation. To achieve this goal, we adopt the framework of Kitamura and Stutzer (1997), and we focus on the case that moment functions are stable autoregressive processes.[4] More precisely, there exists a set of autoregressive (AR) coefficients $\omega_0, \cdots, \omega_{m_g}$ such that the error terms of the following *stable* AR regression have zero means and zero serial correlations:

$$g^{\omega}_{\mathcal{Q}}(\theta_0, \psi_0; \mathbf{z_t}) \equiv \sum_{j=0}^{m_g} \omega_j g_{\mathcal{Q}}(\theta_0, \psi_0; \mathbf{x_{t-j}}, \mathbf{y_{t-j}}) \tag{12}$$

where $\mathbf{z_t} \equiv \{\mathbf{x_{t-j}}, \mathbf{y_{t-j}} : \ j = 0, \cdots, m_g\}$. This assumption does not offer the most general setting; however, it should provide a fair approximation to the dynamics of moment conditions in many cases studied in finance and economics, and more important, it allows a factorization of the likelihood for finite-sample interpretation while guaranteeing the first-order asymptotic efficiency of the analogous maximum likelihood estimator. Importantly, as a result of stability, the original moment condition $\mathbb{E}\left[g_{\mathcal{Q}}(\theta_0, \psi_0; \mathbf{x_t}, \mathbf{y_t})\right] = 0$ is equivalent to the moment condition:

$$\mathbb{E}\left[g^{\omega}_{\mathcal{Q}}(\gamma_0; \mathbf{z_t})\right] = 0, \tag{13}$$

[2] The idea of acknowledging that it is often very difficult to come up with precise distributions to be used as likelihood functions and thus choosing a likelihood (or a prior) among a set of sampling models (or priors) in Bayesian analysis is referred to as "robust Bayesian analysis" (see, e.g. Wasserman, 1992; Berger, 1994; Pericchi and Pérez, 1994).

[3] Alternatively, the empirical likelihood (EL) of Owen (1988, 1990, 1991) and the exponential tilted empirical likelihood (ETEL) of Kitamura and Stutzer (1997) and Schennach (2007) can also be used as the likelihood part of the Bayesian inference. In fact, they are the Bayesian empirical likelihood (BEL) of Lazar (2003) and the Bayesian exponential tilted empirical likelihood (BETEL) of Schennach (2005). One application of BEL and BETEL is the statistical analysis of disaster risk models in Julliard and Ghosh (2012).

[4] The same assumption is also adopted by Kim (2002, Remarks 1 and 3). More details can be found in Assumption A9 in Appendix 1.4. The assumption can be weakened to the case that the lag size m_g increases with sample size n.

where we define $\gamma \equiv (\theta, \psi)$ and $\gamma_0 \equiv (\theta_0, \psi_0)$ for notational simplicity.

Given each γ, we define the set of probability measures, denoted by $\mathfrak{Q}(\gamma; n)$, such that

$$\mathfrak{Q}(\gamma; n) \equiv \left\{ Q_n : \ \mathbb{E}_{Q_n}\left[g_{\mathcal{Q}}^{\omega}(\gamma; \mathbf{z_t}) \right] = 0, \quad \forall\, t = 1, \cdots, n \right\}. \tag{14}$$

The true distribution of $\mathbf{z^n} = (\mathbf{z_1}, \cdots, \mathbf{z_n})$, denoted by $\mathbb{Q}_n$, belongs to the distribution set $\mathfrak{Q}(\gamma_0; n)$. The limited-information likelihood $\mathbb{Q}_{\gamma,n}$ for Bayesian analysis of full model is chosen according to the principle of minimum Kullback-Leibler divergence:

$$\mathbb{Q}_{\gamma,n} = \underset{Q_n \in \mathfrak{Q}(\gamma;n)}{\operatorname{argmin}} \mathbf{D}_{KL}(Q_n || \mathbb{Q}_n) = \underset{Q_n \in \mathfrak{Q}(\gamma;n)}{\operatorname{argmin}} \int \ln\left(\mathrm{d}Q_n / \mathrm{d}\mathbb{Q}_n \right) \mathrm{d}Q_n \tag{15}$$

where $\mathrm{d}Q_n/\mathrm{d}\mathbb{Q}_n$ is the Radon-Nikodym derivative (or density) of Q_n with respect to the true probability measure $\mathbb{Q}_n$. It is well-known that the limited-information likelihood $\mathbb{Q}_{\gamma,n}$ has the following Gibbs canonical density (see, e.g., Csiszár, 1975; Cover and Thomas, 1991, Chapter 11):

$$\mathrm{d}\mathbb{Q}_{\gamma,n}/\mathrm{d}\mathbb{Q}_n = \exp\left\{ \eta_{\mathcal{Q}}(\gamma)^T \sum_{t=1}^{n} g_{\mathcal{Q}}^{\omega}(\gamma; \mathbf{z_t}) - n A_{\mathcal{Q}}(\gamma) \right\}, \quad \forall\, \gamma \tag{16}$$

where $A_{\mathcal{Q}}(\gamma) \equiv \ln \mathbb{E}\left[e^{\eta_{\mathcal{Q}}(\gamma)^T g_{\mathcal{Q}}^{\omega}(\gamma; \mathbf{z_t})} \right]$, and the Lagrangian multipliers $\eta_{\mathcal{Q}}(\gamma)$ are chosen to make the moment conditions satisfied:

$$0 = \mathbb{E}\left[g_{\mathcal{Q}}^{\omega}(\gamma; \mathbf{z}) e^{\eta_{\mathcal{Q}}(\gamma)^T g_{\mathcal{Q}}^{\omega}(\gamma; \mathbf{z})} \right], \quad \forall\, \gamma. \tag{17}$$

We denote $\pi_{\mathcal{Q}}(\mathbf{z_t}; \gamma) \equiv \exp\left\{ \eta_{\mathcal{Q}}(\gamma)^T g_{\mathcal{Q}}^{\omega}(\gamma; \mathbf{z_t}) - A_{\mathcal{Q}}(\gamma) \right\}$. The posterior density is

$$\pi_{\mathcal{Q}}(\gamma | \mathbf{x^n}, \mathbf{y^n}) \propto \pi(\gamma) \exp\left\{ \eta_{\mathcal{Q}}(\gamma)^T \sum_{t=1}^{n} g_{\mathcal{Q}}^{\omega}(\gamma; \mathbf{z_t}) - n A_{\mathcal{Q}}(\gamma) \right\}. \tag{18}$$

More discussions on the validity of the limited-information likelihood and Bayesian analysis can be found in Appendix 1.1. Analogously, the limited-information likelihood $\mathbb{P}_{\theta,n}$ for the baseline model is constructed based on the baseline moment conditions $\mathbb{E}\left[g_{\mathcal{P}}(\theta, \mathbf{x_t}) \right] = 0$ and the principle of minimum Kullback-Leibler divergence (15). It has the density function:

$$\mathrm{d}\mathbb{P}_{\theta,n}/\mathrm{d}\mathbb{P}_n = \exp\left\{ \eta_{\mathcal{P}}(\theta)^T \sum_{t=1}^{n} g_{\mathcal{P}}^{\omega}(\theta; \mathbf{w_t}) - n A_{\mathcal{P}}(\theta) \right\} \tag{19}$$

where $\mathbf{w_t} = \{ \mathbf{x_{t-j}} : \ j = 0, \cdots, m_g \}$ with underlying true distribution $\mathbb{P}_n$; $g_{\mathcal{P}}^{\omega}(\theta; \mathbf{w_t})$ is the corresponding sub-vector of smoothed moments $g_{\mathcal{Q}}^{\omega}(\theta; \mathbf{z_t})$ for the baseline model; the functions $\eta_{\mathcal{P}}(\theta)$ and $A_{\mathcal{P}}(\theta)$ are defined analogous to $\eta_{\mathcal{Q}}(\gamma)$ and $A_{\mathcal{Q}}(\gamma)$. We denote $\pi_{\mathcal{P}}(\mathbf{w_t}; \theta) \equiv \exp\left\{ \eta_{\mathcal{P}}(\theta)^T g_{\mathcal{P}}^{\omega}(\theta; \mathbf{w_t}) - A_{\mathcal{P}}(\theta) \right\}$. The MLE for $\mathbb{P}_{\theta,n}$, denoted by $\hat{\theta}_{\mathrm{ML}}^{\mathcal{P}}$, also satisfies asymptotic normality with variance $\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}$.

Similar to full model, the posterior of baseline model is

$$\pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \propto \pi(\theta) \exp\left\{\eta_{\mathcal{P}}(\theta)^T \sum_{t=1}^{n} g^{\omega}_{\mathcal{P}}(\theta; \mathbf{w_t}) - nA_{\mathcal{P}}(\theta)\right\}. \tag{20}$$

Without loss of generality, we assume that $g_{\mathcal{P}}(\theta; \mathbf{x_t}) = g^{\omega}_{\mathcal{P}}(\theta; \mathbf{w_t})$ and $g_{\mathcal{Q}}(\gamma; \mathbf{x_t}, \mathbf{y_t}) = g^{\omega}_{\mathcal{Q}}(\gamma; \mathbf{z_t})$. It should also be noted that the functions in the limited-information likelihoods $\eta_{\mathcal{P}}(\theta)$, $A_{\mathcal{P}}(\theta)$, $\eta_{\mathcal{Q}}(\gamma)$, and $A_{\mathcal{Q}}(\gamma)$ are not known. It is innocuous for our theoretical exercises and results. However, in practice, they need to be replaced by their approximating counterparts. For example, $\eta_{\mathcal{P}}(\theta)$ can be estimated by solving the following equation, for each θ,

$$0 = \frac{1}{n} \sum_{t=1}^{n} g_{\mathcal{P}}(\theta; \mathbf{x_t}) e^{\widehat{\eta}_{\mathcal{P}}(\theta)^T g_{\mathcal{P}}(\theta; \mathbf{x_t})} \tag{21}$$

and then we estimate $A_{\mathcal{P}}(\theta)$ by $\widehat{A}_{\mathcal{P}}(\theta)$ as follows:

$$\widehat{A}_{\mathcal{P}}(\theta) = \ln\left[\frac{1}{n} \sum_{t=1}^{n} e^{\widehat{\eta}_{\mathcal{P}}(\theta)^T g_{\mathcal{P}}(\theta; \mathbf{x_t})}\right]. \tag{22}$$

The functional estimators $\widehat{\eta}_{\mathcal{Q}}(\gamma)$ and $\widehat{A}_{\mathcal{Q}}(\gamma)$ can be constructed in the similar way. Under the regularity conditions, the functional estimators converge to the true functions uniformly in probability.

Discussion: The Validity of Limited-Information Likelihood $\mathbb{Q}_{\gamma,n}$ The GMM Bayesian methods of Kim (2002) and Chernozhukov and Hong (2003) are primarily for large sample analysis and potentially invalid for finite-sample interpretation due to the fact that the GMM framework based on a quadratic form empirical moments under general assumptions are justified asymptotically (see, Hansen, 1982). Specifically, Chernozhukov and Hong (2003) motivate the exponential of moment conditions' quadratic form as the limited-information likelihood mainly from a computational perspective, without providing a compelling theoretical rationalization. Kim (2002) justifies the particular exponential quadratic form by appealing to the principle of minimum Kullback-Leibler divergence yet based on moment conditions in an asymptotic sense. To guarantee a valid finite-sample interpretation and valid parametric family for likelihoods, we adopt the framework of Kitamura and Stutzer (1997).

Here are several reasons for the limited-information likelihood in (16) to be used as a valid parametric family for likelihood within the Bayesian paradigm. First, the set of distributions $\mathbb{Q}_{\gamma,n}$ characterize a proper parametric family of likelihoods for statistical inference, since $\mathbb{Q}_{\gamma,n}$ is the true distribution if and only if $\gamma = \gamma_0$, that is $\eta_{\mathcal{Q}}(\gamma_0) = A_{\mathcal{Q}}(\gamma_0) = 0$. Further, the MLE of the parametric family $\mathbb{Q}_{\gamma,n}$, denoted by $\hat{\gamma}^{\mathbb{Q}}_{\mathrm{ML}}$, is asymptotically first-order equivalent to the GMM estimator Hansen (1982) and the ET estimator Kitamura and Stutzer (1997) with a normal asymptotic distribution (see Section 1.6):

$$\underset{n\to\infty}{\text{wlim}}\sqrt{n}\left(\hat{\gamma}_{\text{ML}}^{\mathbb{Q}}-\gamma_0\right)=N(0,\mathbf{I}_{\mathbb{Q}}(\gamma_0)^{-1}),\quad \text{with } n\mathbf{I}_{\mathbb{Q}}(\gamma_0)=-\mathbb{E}\left[\nabla^2\ln\left(d\mathbb{Q}_{\gamma,n}/d\mathbb{Q}_n\right)\right].$$

It should be noted that Kitamura and Stutzer (1997) propose the exponential tilted estimator based on Fenchel duality:

$$\eta_{\mathbb{Q}}(\gamma)=\underset{\eta}{\operatorname{argmin}}\,\mathbb{E}\left[e^{\eta^T g_{\mathbb{Q}}(\gamma;\mathbf{z})}\right],\quad \forall\ \gamma. \tag{23}$$

Their motivation is computational. But we are not proposing new estimation method here. Thus, we directly consider the maximum likelihood estimator for the limited-information likelihood family $\mathbb{Q}_{\gamma,n}$.

Second, the limited-information likelihood (16) can be factorized properly so as to be consistent with the stationary Markovian properties of underlying time series, since the sequential dependence of $\mathbb{Q}_{\gamma,n}$ are fully captured by the true distribution $\mathbb{Q}_n$.

Third, the Bayesian analysis based on (16) is equivalent to Bayesian exponentially tilted empirical likelihood (BETEL) asymptotically (see, e.g., Julliard and Ghosh, 2012). Schennach (2005) provides a probabilistic interpretation of the exponential tilted empirical likelihood that justifies its use in Bayesian inference.

1.2 The Effective-Sample Size

We quantify the discrepancy between probability distributions using a standard statistical measure, the relative entropy (also known as the Kullback-Leibler divergence). The relative entropy between $\pi_{\mathcal{P}}(\theta|\mathbf{x^n})$ and the marginal posterior $\pi_{\mathbb{Q}}(\theta|\mathbf{x^n},\mathbf{y^n})$ is

$$\mathbf{D}_{KL}\left[\pi_{\mathbb{Q}}(\theta|\mathbf{x^n},\mathbf{y^n})||\pi_{\mathcal{P}}(\theta|\mathbf{x^n})\right]=\int\ln\left(\frac{\pi_{\mathbb{Q}}(\theta|\mathbf{x^n},\mathbf{y^n})}{\pi_{\mathcal{P}}(\theta|\mathbf{x^n})}\right)\pi_{\mathbb{Q}}(\theta|\mathbf{x^n},\mathbf{y^n})d\theta\,. \tag{24}$$

Intuitively, we can think of the log posterior ratio $\ln(\pi_{\mathbb{Q}}(\theta|\mathbf{x^n},\mathbf{y^n})/\pi_{\mathcal{P}}(\theta|\mathbf{x^n}))$ as a measure of the discrepancy between the two posteriors at a given θ. Then the relative entropy is the average discrepancy between the two posteriors over all possible θ, where the average is computed under the constrained posterior. $\mathbf{D}_{KL}(\pi_{\mathbb{Q}}(\theta|\mathbf{x^n},\mathbf{y^n})||\pi_{\mathcal{P}}(\theta|\mathbf{x^n}))$ is finite if and only if the support of the posterior $\pi_{\mathbb{Q}}(\theta|\mathbf{x^n},\mathbf{y^n})$ is a subset of the support of the posterior $\pi_{\mathcal{P}}(\theta|\mathbf{x^n})$, that is, Assumption A6 in Subsection 1.4 holds.

The magnitude of relative entropy is difficult to interpret directly, and we propose an intuitive "effective sample size" interpretation. Instead of imposing the cross-equation restrictions from the structural model, one can gain extra information about θ within the baseline model with additional data. We evaluate the amount of additional data under the baseline model needed to match the informativeness of cross-equation restrictions.

Suppose we draw additional data $\tilde{\mathbf{x}}^{\mathbf{m}}$ of sample size m from the Bayesian predictive distribution

$$\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\mathbf{x}^{\mathbf{n}}) \equiv \int \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta)\pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}})\mathrm{d}\theta, \tag{25}$$

where the effective sample $\tilde{\mathbf{x}}^{\mathbf{m}}$ is independent of observed sample $\mathbf{x}^{\mathbf{n}}$ given baseline parameters θ, that is, $\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \theta|\mathbf{x}^{\mathbf{n}}) = \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta)\pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}})$. Again, we measure the gain in information from this additional sample $\tilde{\mathbf{x}}^{\mathbf{m}}$ using relative entropy,

$$\mathbf{D}_{KL}(\pi_{\mathcal{P}}(\theta|\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}})||\pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}})) = \int \ln\left(\frac{\pi_{\mathcal{P}}(\theta|\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}})}{\pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}})}\right)\pi_{\mathcal{P}}(\theta|\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}})\mathrm{d}\theta\,. \tag{26}$$

$\mathbf{D}_{KL}(\pi_{\mathcal{P}}(\theta|\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}})||\pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}}))$ depends on the realization of the additional sample of data $\tilde{\mathbf{x}}^{\mathbf{m}}$. The average relative entropy (information gain) over possible future samples $\{\tilde{\mathbf{x}}^{\mathbf{m}}\}$ according to the Bayesian predictive distribution $\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\mathbf{x}^{\mathbf{n}})$ equals the *mutual information* between $\tilde{\mathbf{x}}^{\mathbf{m}}$ and θ given $\mathbf{x}^{\mathbf{n}}$:

$$\begin{aligned}\mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}};\theta|\mathbf{x}^{\mathbf{n}}) &\equiv \mathbb{E}^{\tilde{\mathbf{x}}^{\mathbf{m}}|\mathbf{x}^{\mathbf{n}}}\left[\mathbf{D}_{KL}(\pi_{\mathcal{P}}(\theta'|\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}})||\pi_{\mathcal{P}}(\theta'|\mathbf{x}^{\mathbf{n}}))\right] \\ &= \int\int \mathbf{D}_{KL}(\pi_{\mathcal{P}}(\theta'|\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}})||\pi_{\mathcal{P}}(\theta'|\mathbf{x}^{\mathbf{n}}))\,\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta)\,\pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}})\,\mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\,\mathrm{d}\theta. \end{aligned} \tag{27}$$

Like the relative entropy, the mutual information is always positive. It is easy to check that $\mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}};\theta|\mathbf{x}^{\mathbf{n}}) = 0$ when $m = 0$. Under the assumption that the prior distribution is nonsingular and the parameters in the likelihood function are well identified, and additional general regularity conditions, $\mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}};\theta|\mathbf{x}^{\mathbf{n}})$ is monotonically increasing in m and converges to infinity as m increases. These properties ensure that we can find an extra sample size m that equates (approximately, due to the fact that m is an integer) $\mathbf{D}_{KL}(\pi_{\mathcal{Q}}(\theta|\mathbf{x}^{\mathbf{n}}, \mathbf{y}^{\mathbf{n}})||\pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}}))$ with $\mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}};\theta|\mathbf{x}^{\mathbf{n}})$. It is only meaningful to match 1-dimensional distributions even in asymptotic senses. Thus, the results in this section are valid only for scalar feature functions.

Definition 1 (Effective-Sample Size Information Measure)**.** *For a feature function vector $f : \mathbb{R}^{D_\Theta} \to \mathbb{R}$, we define the effective-sample measure of the informativeness of the cross-equation restrictions as*

$$\varrho^{f}_{KL}(\mathbf{x}^{\mathbf{n}}, \mathbf{y}^{\mathbf{n}}) = \frac{n + m^*_f}{n}, \tag{28}$$

where m^*_f enables the matching of two information quantities

$$\mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}^*_{\mathbf{f}}}; f(\theta)|\mathbf{x}^{\mathbf{n}}) \leq \mathbf{D}_{KL}(\pi_{\mathcal{Q}}(f(\theta)|\mathbf{x}^{\mathbf{n}}, \mathbf{y}^{\mathbf{n}})||\pi_{\mathcal{P}}(f(\theta)|\mathbf{x}^{\mathbf{n}})) < \mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}^*_{\mathbf{f}}+\mathbf{1}}; f(\theta)|\mathbf{x}^{\mathbf{n}}), \tag{29}$$

with $\mathbf{D}_{KL}(\pi_{\mathcal{Q}}(f(\theta)|\mathbf{x}^{\mathbf{n}}, \mathbf{y}^{\mathbf{n}})||\pi_{\mathcal{P}}(f(\theta)|\mathbf{x}^{\mathbf{n}}))$ being the relative entropy between the constrained and unconstrained posteriors of $f(\theta)$ and $\mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}}; f(\theta)|\mathbf{x}^{\mathbf{n}})$ being the conditional mutual information between the additional sample of data $\mathbf{x}^{\mathbf{m}}$ and the transformed parameter $f(\theta)$ given the existing sample of data $\mathbf{x}^{\mathbf{n}}$.

For scalar-valued feature functions ($D_f = 1$), there exists a direct connection between our Fisher fragility measure and the relative-entropy based informativeness measure (i.e. effective-sample size information measure in Definition 1). The effective-sample size ratio $\varrho_{KL}^{f}(\mathbf{x^n}, \mathbf{y^n})$ is defined with finite-sample validity. The following theorem establishes asymptotic equivalence between the Fisher fragility measure $\varrho^{\mathbf{v}}(\theta_0|\psi_0)$ and the effective-sample size information measure $\varrho_{KL}^{f}(\mathbf{x^n}, \mathbf{y^n})$. Intuitively, the extra effective-sample size m_f^* increases proportionally in the observed sample size n where the limiting proportional growth rate $\mathrm{wlim}_{n\to\infty}\, \varrho_{KL}^{f}(\mathbf{x^n}, \mathbf{y^n})$ is stochastic.

Theorem 1. *Consider a feature function $f : \mathbb{R}^{D_\Theta} \to \mathbb{R}$ with $\mathbf{v} = \nabla f(\theta_0)$. Under the standard regularity conditions A1 - A8 stated in Subsection 1.4, it must hold that*

$$\underset{n\to\infty}{\mathrm{wlim}} \ln \varrho_{KL}^{f}(\mathbf{x^n}, \mathbf{y^n}) = \ln\left[\varrho^{\mathbf{v}}(\theta_0|\psi_0)\right] + \left[1 - \varrho^{\mathbf{v}}(\theta_0|\psi_0)^{-1}\right](\chi_1^2 - 1), \tag{30}$$

where χ_1^2 is a chi-square random variable with degrees of freedom 1. It immediately implies that

$$\mathbb{E}\left[\underset{n\to\infty}{\mathrm{wlim}} \ln \varrho_{KL}^{f}(\mathbf{x^n}, \mathbf{y^n})\right] = \ln\left[\varrho^{\mathbf{v}}(\theta_0|\psi_0)\right].$$

The result of Theorem 1 follows immediately from the following approximation results summarized in Theorem 2 and Theorem 3. Without loss of generality, we assume that $g_{\mathcal{Q}}^{\omega}(\theta, \psi; \mathbf{z_t}) \equiv g_{\mathcal{Q}}(\theta, \psi; \mathbf{x_t}, \mathbf{y_t})$ or equivalently $m_g = 0$ in Equation (12). Also, because of Assumption A7 (the regular feature function condition), as well as the fact that the definition of our "dark matter" measure in Chen, Dou, and Kogan (2017) and regularity assumptions A1 - A6 in Appendix 1.4 are invariant under invertible and second-order smooth transformations of parameters, we can assume that $D_\Theta = 1$ in the intuitive proofs. The full proofs are in the online appendix.

Theorem 2 (Kullback-Leibler Divergence)**.** *Consider a feature function $f : \mathbb{R}^{D_\Theta} \to \mathbb{R}$ with $\mathbf{v} = \nabla f(\theta_0)$. Let the MLE for the limited-information likelihood of the baseline model $\mathcal{P}$ be $\hat{\theta}_{ML}^{\mathcal{P}}$, and the analogy of the structural model $\mathcal{Q}$ be $\hat{\gamma}_{ML}^{\mathcal{Q}} = (\hat{\theta}_{ML}^{\mathcal{Q}}, \hat{\psi}_{ML}^{\mathcal{Q}})$. Under the regularity conditions stated in Appendix 1.4,*

$$\begin{aligned}\mathbf{D}_{KL}(\pi_{\mathcal{Q}}(f(\theta)|\mathbf{x^n}, \mathbf{y^n})||\pi_{\mathcal{P}}(f(\theta)|\mathbf{x^n})) - \frac{n}{2\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T}(f(\widehat{\theta}_{ML}^{\mathcal{P}}) - f(\widehat{\theta}_{ML}^{\mathcal{Q}}))^2 \\ \to \frac{1}{2}\ln\frac{\mathbf{v}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}^T}{\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T} + \frac{1}{2}\frac{\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T}{\mathbf{v}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}^T} - 1/2,\end{aligned} \tag{31}$$

where convergence is in probability under $\mathbb{Q}_n$.

This theorem generalizes the results in Lin, Pittman, and Clarke (2007, Theorem 3) in two important aspects. It focuses on the results of the moment-based LIL framework, instead of the standard likelihood-based framework. And also, it allows for general weak dependence among the observations which makes our results applicable to time series models in finance and economics.

Here we provide an intuitive proof. The detailed and rigorous technical proof can be found in Subsection 1.10. When n is large, it following approximations hold under the standard conditions:

$$\pi_{\mathcal{P}}\left(f(\theta)|\mathbf{x^n}\right) \approx N\left(f(\hat{\theta}^{\mathcal{P}}_{\mathrm{ML}}), n^{-1}\mathbf{v}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}^T\right), \quad \text{and} \tag{32}$$

$$\pi_{\mathcal{Q}}\left(f(\theta)|\mathbf{x^n},\mathbf{y^n}\right) \approx N\left(f(\hat{\theta}^{\mathcal{Q}}_{\mathrm{ML}}), n^{-1}\mathbf{v}\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}^T\right). \tag{33}$$

Therefore, when n is large, the asymptotic approximation (31) holds.

Corollary 1. *Consider the feature function $f: \mathbb{R}^{D_\Theta} \to \mathbb{R}$ with $\mathbf{v} = \partial f(\theta_0)/\partial\theta$. Under the assumptions in Subsection 1.4, if we define*

$$\lambda \equiv \frac{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}{\mathbf{v}^T\mathbf{I}_{\mathcal{Q}}(\theta_0)^{-1}\mathbf{v}},$$

then it follows that

$$\underset{n\to+\infty}{\mathrm{wlim}} \frac{n}{\mathbf{v}^T\mathbf{I}_{\mathcal{Q}}(\theta_0)^{-1}\mathbf{v}}(f(\widehat{\theta}^{\mathcal{P}}) - f(\widehat{\theta}^{\mathcal{Q}}))^2 = (1-\lambda^{-1})\chi_1^2. \tag{34}$$

Moreover,

$$\underset{n\to+\infty}{\mathrm{wlim}} \mathbf{D}_{KL}(\pi_{\mathcal{Q}}(f(\theta)|\mathbf{x^n},\mathbf{y^n})||\pi_{\mathcal{P}}(f(\theta)|\mathbf{x^n})) = \frac{1-\lambda^{-1}}{2}(\chi_1^2 - 1) + \frac{1}{2}\ln(\lambda), \tag{35}$$

The asymptotic distribution result in (34) is directly from Proposition 11. The asymptotic result in (35) is based on Theorem 2, limit result (34), and the Slutsky Theorem.

Theorem 3 (Mutual Information)**.** *Consider a feature function $f: \mathbb{R}^{D_\Theta} \to \mathbb{R}$ with $\mathbf{v} = \nabla f(\theta_0)$. Under the assumptions in Subsection 1.4, if $m/n \to \varsigma \in (0,\infty)$ as both m and n approach infinity,*

$$\mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}}; f(\theta)|\mathbf{x^n}) - \frac{1}{2}\ln\left(\frac{m+n}{n}\right) \to 0, \tag{36}$$

where convergence is in probability under $\mathbb{Q}_n$.

There exist related approximation results for mutual information $\mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}}; \theta|\mathbf{x^n})$, which consider large m while holding the observed sample size n fixed. For more details, see Clarke and Barron (1990, 1994) and references therein. See also the case of non-identically distributed observations by Polson (1992), among others. Our results differ in that we allow both m and n to grow. Ours is a technically nontrivial extension of the existing results.

Here we provide an intuitive proof. The detailed and rigorous technical proof can be found in

Subsection 1.10. By definition in (27), the mutual information given $\mathbf{x^n}$ can be rewritten as

$$\begin{aligned}\mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}};\theta|\mathbf{x^n}) &= \int\int \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta)\pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \ln \frac{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x^n}|\theta)\pi_{\mathcal{P}}(\mathbf{x^n})}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x^n})\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)} \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta \\ &= \int\int \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta)\pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \ln \frac{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x^n}|\theta)}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x^n})} \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta - \int \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \ln \frac{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)}{\pi_{\mathcal{P}}(\mathbf{x^n})}\mathrm{d}\theta.\end{aligned}$$

The following approximation is standard in the information-theoretic literature (see, e.g., Clarke and Barron, 1990, 1994): when m and n are large, it holds that

$$\ln \frac{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x^n}|\theta)}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x^n})} \approx -\frac{1}{2}\tilde{S}_{m+n}(\theta)^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\tilde{S}_{m+n}(\theta) + \frac{1}{2}\ln\frac{m+n}{2\pi} + \ln\frac{1}{\pi_{\mathcal{P}}(\theta)} + \frac{1}{2}\ln|\mathbf{I}_{\mathcal{P}}(\theta)|,$$

where

$$\tilde{S}_{m+n}(\theta) \equiv \frac{1}{\sqrt{m+n}}\left[\sum_{t=1}^{n}\nabla\ln\pi_{\mathcal{P}}(\mathbf{x_t};\theta) + \sum_{t=1}^{m}\nabla\ln\pi_{\mathcal{P}}(\tilde{\mathbf{x}}_{\mathbf{t}};\theta)\right]. \tag{37}$$

Similarly, it also holds that

$$\ln \frac{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)}{\pi_{\mathcal{P}}(\mathbf{x^n})} \approx -\frac{1}{2}S_n(\theta)^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}S_n(\theta) + \frac{1}{2}\ln\frac{n}{2\pi} + \ln\frac{1}{\pi_{\mathcal{P}}(\theta)} + \frac{1}{2}\ln|\mathbf{I}_{\mathcal{P}}(\theta)|, \tag{38}$$

where

$$S_n(\theta) \equiv \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\nabla\ln\pi_{\mathcal{P}}(\mathbf{x_t};\theta). \tag{39}$$

It follows from (37) and (39) that

$$\tilde{S}_{m+n}(\theta) = \sqrt{\frac{n}{m+n}}S_n(\theta) + \sqrt{\frac{m}{m+n}}\tilde{S}_m(\theta), \tag{40}$$

where

$$\tilde{S}_m(\theta) \equiv \frac{1}{\sqrt{m}}\sum_{t=1}^{m}\nabla\ln\pi_{\mathcal{P}}(\tilde{\mathbf{x}}_{\mathbf{t}};\theta). \tag{41}$$

Now recall that the family of limited-information likelihoods $\pi_{\mathcal{P}}(\mathbf{x};\theta)$ satisfy the standard properties: $\forall\ \theta$,

$$\begin{aligned}&\mathbb{E}\left[\pi_{\mathcal{P}}(\mathbf{x};\theta)\right] = 1 \quad \text{and} \quad \mathbb{E}\left[\pi_{\mathcal{P}}(\mathbf{x};\theta)\nabla\ln\pi_{\mathcal{P}}(\mathbf{x};\theta)\right] = 0 \quad \text{and} \\ &\mathbb{E}\left\{\pi_{\mathcal{P}}(\mathbf{x};\theta)\left[\nabla\ln\pi_{\mathcal{P}}(\mathbf{x};\theta)\right]\left[\nabla\ln\pi_{\mathcal{P}}(\mathbf{x};\theta)\right]^T\right\} = -\mathbb{E}\left\{\pi_{\mathcal{P}}(\mathbf{x};\theta)\nabla^2\ln\pi_{\mathcal{P}}(\mathbf{x};\theta)\right\} = \mathbf{I}_{\mathcal{P}}(\theta).\end{aligned}$$

Thus, the following important identity can be derived:

$$\begin{aligned}&\int\int \tilde{S}_{m+n}(\theta)^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\tilde{S}_{m+n}(\theta)\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta)\pi_{\mathcal{P}}(\theta|\mathbf{x^n})\mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta \\ &\qquad\qquad = \frac{n}{m+n}\int S_n(\theta)^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}S_n(\theta)\pi_{\mathcal{P}}(\theta|\mathbf{x^n})\mathrm{d}\theta + \frac{m}{m+n}.\end{aligned}$$

Using the Taylor expansion of $S_n(\theta)$ around $\hat{\theta}^{\mathcal{P}}_{\text{ML}}$ and the normal approximation of posterior (32),

$$
\begin{aligned}
&\int S_n(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} S_n(\theta) \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \mathrm{d}\theta \\
&\qquad \approx n\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1} \left[\frac{1}{n}\sum_{t=1}^{n} \nabla^2 \ln \pi_{\mathcal{P}}(\mathbf{x_t}; \hat{\theta}^{\mathcal{P}}_{\text{ML}})\right]^2 \int \left(\theta - \hat{\theta}^{\mathcal{P}}_{\text{ML}}\right)^2 \varphi_{\mathcal{P},n}(\theta) \mathrm{d}\theta \\
&\qquad \approx 1
\end{aligned}
$$

where $\varphi_{\mathcal{P},n}(\theta) = \sqrt{\det\left[n^{-1}\mathbf{I}_{\mathcal{P}}(\theta_0)\right]/(2\pi)^{D_\Theta}} \exp\left\{-\frac{1}{2}(\theta - \hat{\theta}^{\mathcal{P}}_{\text{ML}})^T \left[n\mathbf{I}_{\mathcal{P}}(\theta_0)\right](\theta - \hat{\theta}^{\mathcal{P}}_{\text{ML}})\right\}$.

Therefore, when m and n are large, the mutual information can be approximated as follows

$$
\begin{aligned}
\mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}}; \theta|\mathbf{x^n}) &\approx \frac{1}{2}\left[\ln\frac{m+n}{n} + \frac{m}{m+n}\int S_n(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} S_n(\theta)\pi_{\mathcal{P}}(\theta|\mathbf{x^n})\mathrm{d}\theta - \frac{m}{m+n}\right] \\
&\approx \frac{1}{2}\ln\frac{m+n}{n}.
\end{aligned}
$$

Now we have completed the intuitive proof. The lengthy full proof in Subsection 1.9 follows the same main idea, yet with rigorous establishments of the approximation signs "≈" above.

1.3 Generic Notations and Definitions

First, we introduce some notations for the matrices. For any real symmetric non-negative definite matrix A, we define $\lambda_M(A)$ to be the largest eigenvalue of A and define $\lambda_m(A)$ to be the smallest eigenvalue of A. For a matrix A, we define the spectral norm of A to be $||A||_{\mathcal{S}}$. By definition of spectral norm, we know that for any real matrix A,

$$
||A||_{\mathcal{S}} \equiv \sqrt{\lambda_M(A^T A)}.
$$

Denote $\overline{\lambda}(\theta)$ and $\underline{\lambda}(\theta)$ to be the largest eigenvalue and the smallest eigenvalue of $\mathbf{I}_{\mathcal{P}}(\theta)$, respectively. That is,

$$
\overline{\lambda}(\theta) = \lambda_m(\mathbf{I}_{\mathcal{P}}(\theta)) \quad \text{and} \quad \underline{\lambda}(\theta) = \lambda_M(\mathbf{I}_{\mathcal{P}}(\theta)).
$$

If the matrix $\mathbf{I}_{\mathcal{P}}(\theta)$ is continuous in θ, $\overline{\lambda}(\theta)$ and $\underline{\lambda}(\theta)$ are continuous in θ. We define upper bound and lower bound to be

$$
\overline{\lambda} \equiv \sup_{\theta\in\Theta} \overline{\lambda}(\theta), \text{ and } \underline{\lambda} \equiv \inf_{\theta\in\Theta} \underline{\lambda}(\theta). \tag{42}
$$

Second, we introduce some notations related to subsets in Euclidean spaces. We define the "Euclidean distance" between two sets $S_1,\ S_2 \subset \mathbb{R}^{D_\Theta}$ as follows

$$
d_L(S_1, S_2) \equiv \inf\{|s_1 - s_2| :\ s_i \in S_i,\ i = 1, 2\}. \tag{43}
$$

For $\theta \in \mathbb{R}^{D_\Theta}$, we denote $\theta_{(1)}$ to be the first element of θ and denote $\theta_{(-1)}$ to be the $D_\Theta - 1$ dimensional vector containing all elements of θ other than $\theta_{(1)}$. Define the open ball centered at θ with radius r

to be

$$\Omega(\theta, r) \equiv \{\vartheta: \ |\vartheta - \theta| < r\} \ \text{ and } \ \overline{\Omega}(\theta, r) \equiv \{\vartheta: \ |\vartheta - \theta| \leq r\}$$

Denote

$$\Omega_{(1)}(\theta, r) = \Omega(\theta_{(1)}, r) \subset \mathbb{R}^1, \ \text{ and } \ \Omega_{(-1)}(\theta, r) = \Omega(\theta_{(-1)}, r) \subset \mathbb{R}^{D_\Theta - 1}.$$

In addition, we define

$$\Theta_{-1}(\theta_{(1)}) \equiv \left\{ \theta_{(-1)} \in \mathbb{R}^{D_\Theta - 1} \mid \begin{pmatrix} \theta_{(1)} \\ \theta_{(-1)} \end{pmatrix} \in \Theta \right\}, \tag{44}$$

and we denote

$$V_\Theta \equiv \mathrm{Vol}(\Theta) < +\infty, \quad V_\Theta(\theta_{(1)}) \equiv \mathrm{Vol}(\Theta_{-1}(\theta_{(1)})) < +\infty, \ \text{ and } \ V_{\Theta,1} \equiv \sup_{\theta_{(1)} \in \Theta_{(1)}} V_\Theta(\theta_{(1)}) < +\infty.$$

Third, we introduce some notations on metrics of probability measures. Consider two probability measures P and Q with densities p and q with respect to Lebesgue measure, respectively. The Hellinger affinity between P and Q is denoted as

$$\alpha_H(P, Q) \equiv \int \sqrt{p(x)q(x)} \mathrm{d}x.$$

The total variation distance between P and Q is denoted as

$$||P - Q||_{TV} \equiv \int |p(x) - q(x)| \mathrm{d}x.$$

Fourth, we introduce notations for time series. The maximal correlation coefficient and the uniform mixing coefficient are defined as

$$\rho_{\max}(\mathcal{F}_1, \mathcal{F}_2) \equiv \sup_{f_1 \in L^2_{\text{real}}(\mathcal{F}_1), f_2 \in L^2_{\text{real}}(\mathcal{F}_2)} |\mathrm{Corr}(f_1, f_2)|,$$

and

$$\phi(\mathcal{F}_1, \mathcal{F}_2) \equiv \sup_{A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2} |P(A_2|A_1) - P(A_2)|,$$

where $L^2_{\text{real}}(\mathcal{F}_i)$ denote the space of square-integrable. $\mathcal{F}_i$-measurable, real-valued random variables, for any sub $\sigma-$fields $\mathcal{F}_i \subset \mathcal{F}$.

1.4 Regularity Conditions for Theoretical Results

The regularity conditions we choose to impose on the behavior of the data are influenced by three major considerations. First, our assumptions are chosen to allow processes of sequential dependence. In particular, the processes allowed should be relevant to intertemporal asset pricing models. Second, our assumptions are required to meet the analytical tractability. Third, our assumptions are sufficient conditions in the sense that we are not trying to provide the weakest conditions or high level conditions to guarantee the results; but instead, we chose those regularity conditions which are relatively straightforward to check 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)
There exists constant $\lambda_D \geq 2d_D/(d_D - 1)$, where d_D is the constant in Assumption A3 (dominance condition), 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. Its definition 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, 1987), we adopt the mixing conditions as a convenient way of describing economic and financial data which allows time dependence and heteroskedasticity. The mixing conditions basically restrict the memory of a process to be weak, while allowing heteroskedasticity, so that large sample properties of the process are preserved. 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

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

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

Remark. *The dominating function assumption is widely adopted in the literature of generalized method of moments (Newey, 1985a,b; Newey and West, 1987). The dominating assumption, together with the uniform mixing assumption and stationarity assumption, imply the stochastic equicontinuity condition (iv) in Proposition 1 of Chernozhukov and Hong (2003). In the seminal GMM paper by Hansen (1982), the moment continuity condition can also be derived from the dominance conditions.*

Remark. *Following the discussion of Section 1, for each pair of j and k, it holds that for some constant $\zeta > 0$ and large constant $C > 0$,*

$$\mathbb{E}\sup_{\theta\in\Theta}\left|\frac{\partial^2}{\partial\theta_j\partial\theta_k}\ln\pi_{\mathcal{P}}(\mathbf{w};\theta)\right|^{2+\zeta} < C, \quad \textit{and} \quad \mathbb{E}\sup_{\theta\in\Theta}\left|\frac{\partial}{\partial\theta_j}\ln\pi_{\mathcal{P}}(\mathbf{w};\theta)\right|^{2+\zeta} < C, \quad \textit{and} \tag{45}$$

$$\mathbb{E}\sup_{\gamma\in\Theta\times\Psi}\left|\frac{\partial^2}{\partial\gamma_j\partial\gamma_k}\ln\pi_{\mathcal{Q}}(\mathbf{z};\gamma)\right|^{2+\zeta} < C, \quad \textit{and} \quad \mathbb{E}\sup_{\gamma\in\Theta\times\Psi}\left|\frac{\partial}{\partial\gamma_j}\ln\pi_{\mathcal{Q}}(\mathbf{z};\gamma)\right|^{2+\zeta} < C, \tag{46}$$

where $\gamma = (\theta, \psi)$, and $\pi_{\mathcal{P}}$ and $\pi_{\mathcal{Q}}$ are defined in Section 1. 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).

Assumption A4 (Nonsingular 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. *It 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 $\mathbb{E}\left[g_{\mathcal{P}}(\theta; \mathbf{x})\right] = 0$ only if $\theta = \theta_0$. And, the true parameters (θ_0, ψ_0) of the full model is identified by the moment conditions in the sense that $\mathbb{E}\left[g_{\mathcal{Q}}(\theta, \psi; \mathbf{x}, \mathbf{y})\right] = 0$ only if $\theta = \theta_0$ and $\psi = \psi_0$.

Remark. *Consider the discussion of limited-information likelihood in Section 1. The continuous differentiability of moment functions, together with the identification condition, imply that the parametric family of limited-information distributions $\mathbb{P}_\theta$ and $\mathbb{Q}_\gamma$, 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:*

$$\theta \to \theta_0 \Leftrightarrow \mathbb{P}_\theta \to \mathbb{P}_{\theta_0} \Leftrightarrow \mathbb{E}\left[\ln\left(d\mathbb{P}_\theta/d\mathbb{P}\right)\right] \to \mathbb{E}\left[\ln\left(d\mathbb{P}_{\theta_0}/d\mathbb{P}\right)\right] = 0, \quad \textit{and} \tag{47}$$

$$\gamma \to \gamma_0 \Leftrightarrow \mathbb{Q}_\gamma \to \mathbb{Q}_{\gamma_0} \Leftrightarrow \mathbb{E}\left[\ln\left(d\mathbb{Q}_\gamma/d\mathbb{Q}\right)\right] \to \mathbb{E}\left[\ln\left(d\mathbb{Q}_{\gamma_0}/d\mathbb{Q}\right)\right] = 0, \tag{48}$$

where $\gamma = (\theta, \psi)$ and $\gamma_0 = (\theta_0, \psi_0)$. And, the convergence of a sequence of parameter values is equivalent to the convergence of the moment conditions:

$$\gamma \to \gamma_0 \Leftrightarrow \mathbb{E}\left[g_{\mathcal{Q}}(\gamma; \mathbf{x}, \mathbf{y})\right] \to \mathbb{E}\left[g_{\mathcal{Q}}(\gamma_0; \mathbf{x}, \mathbf{y})\right] = 0. \tag{49}$$

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. And, the prior is absolutely continuous with respect to the Lebesgue measure with Radon-Nykodim 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 $\underline{\pi} \equiv \min_{\theta\in\Theta,\psi\in\Psi} \pi(\theta, \psi)$. The probability measure defined by the limited-information posterior density $\pi_{\mathcal{Q}}(\theta|\mathbf{x^n}, \mathbf{y^n})$ is dominated by the probability measure defined by the baseline limited-information posterior density $\pi_{\mathcal{P}}(\theta|\mathbf{x^n})$, for almost every $\mathbf{x^n}, \mathbf{y^n}$ under $\mathcal{Q}_0$.

Remark. *Compactness implies total boundness. In our diaster risk model, the parameter set for the prior is not compact due to the adoption of uninformative prior. However, in that numerical example, we can truncate the parameter set at very large values which will not affect the main numerical results.*

Remark. *The concept of dominating measure here is the one in measure theory. More precisely, this regularity condition requires that for any measurable set which has zero measure under $\pi_{\mathcal{Q}}(\theta|\mathbf{x^n}, \mathbf{y^n})$, it must also have zero measure under $\pi_{\mathcal{P}}(\theta|\mathbf{x^n})$. This assumption is just to guarantee that $\mathbf{D}_{KL}\left(\pi_{\mathcal{Q}}(\theta|\mathbf{x^n}, \mathbf{y^n})||\pi_{\mathcal{P}}(\theta|\mathbf{x^n})\right)$ to be well defined.*

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. 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 $\dfrac{\partial f(\theta)}{\partial(\theta_{(1)}, \cdots, \theta_{(D_f)})} > 0$ 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 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 $\mathbb{R}^{D_\Theta}$.*

Assumption A8 (Exponential Condition)
For sufficiently small $\delta > 0$, $\mathbb{E}\left[\sup_{\gamma'\in\Omega(\gamma,\delta)} \exp\left\{v^T g_{\mathcal{Q}}(\gamma; \mathbf{x}, \mathbf{y})\right\}\right] < \infty$, for all vectors v in a neighborhood of the origin.

Remark. *This assumption is needed to guarantee that the limited-information likelihood $\mathcal{Q}_\gamma$ in (16) is well-defined and the related Fenchel duality holds. While this assumption is stronger than the moment existence assumption in Hansen (1982), it is commonly adopted in the literature involving Kullback-Leibler divergence (see, e.g. Csiszár, 1975; Kitamura and Stutzer, 1997) and general exponential-family models (see, e.g. Berk, 1972; Dou, Pollard, and Zhou, 2012).*

1.5 Lemmas

As emphasized by White and Domowitz (1984), the mixing conditions serve as an operating assumption for economic and financial processes, because mixing assumptions are difficult to verify or test. However, White and Domowitz (1984) argues that the restriction is not so restrictive in the sense that a wide class of transformations of mixing processes are themselves mixing.

Lemma 1. *Let* $\mathbf{z_t} = Z\left((\mathbf{x_{t+\tau_0}}, \mathbf{y_{t+\tau_0}}), \cdots, (\mathbf{x_{t+\tau}}, \mathbf{y_{t+\tau}})\right)$ *where* Z *is a measurable function onto* $\mathbb{R}^{D_z}$ *and two integers* $\tau_0 < \tau$*. If* $\{\mathbf{x_t}, \mathbf{y_t}\}$ *is uniform mixing with uniform mixing coefficients* $\phi(m) \leq Cm^{-\lambda}$ *for some* $\lambda > 0$ *and* $C > 0$*, then* $\{\mathbf{z_t}\}$ *is uniform mixing with uniform mixing coefficients* $\phi_z(m) \leq C_z m^{-\lambda}$ *for some* $C_z > 0$*.*

Proof. It can be derived directly from the definition of uniform mixing. □

The following classical result is put here for easy reference. And, it easily leads to a corollary which will be used repeatedly.

Lemma 2. *For any two* $\sigma-$*fields* $\mathcal{F}_1$ *and* $\mathcal{F}_2$*, it holds that*

$$\rho_{max}(\mathcal{F}_1, \mathcal{F}_2) \leq 2\left[\phi(\mathcal{F}_1, \mathcal{F}_2)\right]^{1/2}.$$

Proof. The proof can be found in Ibragimov (1962) or Doob (1950, Lemma 7.1). □

Corollary 2. *Let* $\{\mathbf{z_t}\}$ *be strictly stationary process satisfying uniform mixing such that* $\phi(m) \leq Cm^{-\lambda}$ *for some* $\lambda > 0$ *and* $C > 0$*, then the autocorrelation function*

$$\max_{i,j} \left|\mathrm{Corr}(\mathbf{z_{(i),t}}, \mathbf{z_{(j),t+m}})\right| \leq 2\sqrt{C} m^{-\lambda/2}$$

where $\mathbf{z_{(i),t}}$ *is the i-th element of* $\mathbf{z_t}$*.*

Proof. It directly follows from Lemma 1 and Lemma 2. □

Lemma 3. *Let* $\{\mathbf{z_t}\}$ *be a sequence of strictly stationary random vectors such that* $\mathbb{E}_{\mathcal{P}_0}|\mathbf{z_t}|^2 < +\infty$ *and it satisfies the uniform mixing condition with* $\phi(m) \leq Cm^{-\lambda}$ *for some* $\lambda > 2$ *and* $C > 0$*. Then,* $\lim_{n\to\infty} n\mathrm{var}_0\left(n^{-1}\sum_{t=1}^{n} \mathbf{z_t}\right) = V_0 < +\infty$*.*

Proof. Let $\mathbf{z_{(i),t}}$ be the i-th element of vector $\mathbf{z_t}$. Denote $\sigma_i^2 \equiv \mathrm{var}_0(\mathbf{z_{(i),t}})$ for each i. And, we denote the cross correlation to be $\rho_{i,j}(\tau) \equiv \mathrm{Corr}(\mathbf{z_{(i),t}}, \mathbf{z_{(j),t+\tau}})$ for all t, τ, i and j. Then, we have, for each pair of i and j,

$$n\mathrm{Cov}_0\left(n^{-1}\sum_{t=1}^{n} \mathbf{z_{(i),t}}, n^{-1}\sum_{t=1}^{n} \mathbf{z_{(j),t}}\right) = \sigma_i\sigma_j\left[\rho_{i,j}(0) + 2\frac{n-1}{n}\rho_{i,j}(1) + \cdots + 2\frac{1}{n}\rho_{i,j}(n-1)\right].$$

According to Corollary 2, we know that $\rho_{i,j}(m) = o(m^{-1})$. Thus, by verifying the Cauchy condition, we know that $\rho_{i,j}(0) + 2\frac{n-1}{n}\rho_{i,j}(1) + \cdots + 2\frac{1}{n}\rho_{i,j}(n-1)$ converges to a finite constant. □

The following two lemmas are extensions of Propositions 6.1 - 6.2 in Clarke and Barron (1990, Page 468-470). Lemma 4 shows that analogs of the soundness condition for certain metrics on probability measures imply the existence of **strongly uniformly exponentially consistent (SUEC)** hypothesis tests. A composite hypothesis test is called **uniformly exponentially consistency (UEC)** if its type-I and type-II errors are uniformly upper bounded by $e^{-\xi n}$ for some positive ξ, over all alternatively (see e.g. Barron, 1989). A **strongly uniformly exponentially consistent (SUEC)** test is a hypothesis test whose type-I and type-II errors are upper bounded by $e^{-\xi n}$ for some positive constant ξ, uniformly over all alternatives and all null parametric models over two subsets in the probability measure space. Lemma 5 shows that metrics with the desirable consistency property exists, which extends Proposition 6.2 in Clarke and Barron (1990).

Lemma 4. *Suppose d_G is a metric on the space of probability measures on X with the property that for any $\epsilon > 0$, there exists $\xi > 0$ and $C > 0$ such that*

$$\mathbb{P}\left\{d_G(\widehat{\mathbb{P}}_n, \mathbb{P}) > \epsilon\right\} \leq Ce^{-\xi n}, \tag{50}$$

uniformly over all probability measures $\mathbb{P}$, where $\widehat{\mathbb{P}}_n$ is the empirical distribution. And, the metric d_G also satisfies

$$d_G(\mathbb{P}_{\theta'}, \mathbb{P}_\theta) \to 0 \ \Rightarrow \ \theta' \to \theta. \tag{51}$$

Then, for any $\delta > \delta_1 \geq 0$ and for each $\theta \in \overline{\Omega}(\theta_0, \delta_1)$, there exists a SUEC hypothesis test of $\theta \in \overline{\Omega}(\theta_0, \delta_1)$ versus alternative $\Omega(\theta_0, \delta)^{\complement}$.

Proof. The proof is an extension based on that of Lemma 6.1 in Clarke and Barron (1990). From (51), for any given $\delta > \delta_1 \geq 0$, there exists $\epsilon_1 > 0$ such that $d_L(\{\theta\}, \Omega(\theta_0, \delta)^{\complement}) > \delta - \delta_1 > 0$ implies that $d_G(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) > \epsilon_1$ for all $\theta' \in \Omega(\theta_0, \delta)^{\complement}$. Thus, for any $\delta > \delta_1 \geq 0$, there exists $\epsilon_1 > 0$ such that, $d_G(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) > \epsilon_1$ for all $\theta \in \Omega(\theta_0, \delta_1)$ and $\theta' \in \Omega(\theta_0, \delta)^{\complement}$. Therefore, for each $\theta \in \overline{\Omega}(\theta_0, \delta_1)$, if we have a SUEC test of

$$\mathbf{H}_0 : \ \mathbb{P} = \mathbb{P}_\theta \text{ versus } \mathbf{H}_A : \ \mathbb{P} \in \{\tilde{\mathbb{P}} : \ d_G(\tilde{\mathbb{P}}, \mathbb{P}_\theta) > \epsilon_1\},$$

then we have a SUEC test of

$$\mathbf{H}_0 : \ \theta' = \theta \text{ versus } \mathbf{H}_A : \ \mathbb{P} \in \{\mathbb{P}_{\theta'} : \ |\theta' - \theta| > \delta - \delta_1\}.$$

Let $\widehat{\mathbb{P}}_n$ be the empirical distribution. We choose $\epsilon = \epsilon_1/2$ and let

$$\mathcal{A}_{\theta,n} \equiv \{\mathbf{x^n} : \ d_G(\widehat{\mathbb{P}}_n, \mathbb{P}_\theta) \leq \epsilon\}$$

be the acceptance region. By the condition (50), we have that the probability of type-I error satisfies, for some $\xi > 0$ and $C > 0$,

$$\mathbb{P}_{\theta,n}\mathcal{A}_{\theta,n}^{\complement} \leq Ce^{-\xi n}, \quad \text{uniformly over } \theta \in \overline{\Omega}(\theta_0, \delta_1).$$

We want to show that the probability of type-II error $\mathbb{P}_{\theta'}\mathcal{A}_{\theta,n}$ is exponentially small uniformly over $\theta' \in \Omega(\theta_0, \delta)^{\complement}$ and $\theta \in \Omega(\theta_0, \delta_1)$.

More precisely, using triangular inequality, we know that for any $\theta \in \overline{\Omega}(\theta_0, \delta_1)$ and $\theta' \in \Omega(\theta_0, \delta)^{\complement}$

$$2\epsilon \leq d_G(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) \leq d_G(\mathbb{P}_\theta, \widehat{\mathbb{P}}_n) + d_G(\widehat{\mathbb{P}}_n, \mathbb{P}_{\theta'}). \tag{52}$$

By definition and inequality (52), for each $\theta \in \overline{\Omega}(\theta_0, \delta_1)$, on the acceptance region $\mathcal{A}_{\theta,n}$, we have

$$2\epsilon < \epsilon + d_G(\widehat{\mathbb{P}}_n, \mathbb{P}_{\theta'}) \text{ for any } \theta' \in \Omega(\theta_0, \delta)^{\complement}. \tag{53}$$

Thus, for each $\theta \in \overline{\Omega}(\theta_0, \delta_1)$, on the acceptance region $\mathcal{A}_{\theta,n}$, it holds that

$$d_G(\widehat{\mathbb{P}}_n, \mathbb{P}_{\theta'}) > \epsilon, \quad \text{for any } \theta' \in \Omega(\theta_0, \delta)^{\complement}.$$

Therefore, for each $\theta \in \overline{\Omega}(\theta_0, \delta_1)$ and each $\theta' \in \Omega(\theta_0, \delta)^{\complement}$, we have

$$\mathbb{P}_{\theta',n}\mathcal{A}_{\theta,n} \leq \mathbb{P}_{\theta',n}\left\{d_G(\widehat{\mathbb{P}}_n, \mathbb{P}_{\theta'}) > \epsilon\right\} \leq Ce^{-\xi n}.$$

□

Lemma 5. *For a space of probability measures denoted by $\mathbf{P}(\Psi, \lambda)$, on a separable metric space $\mathcal{X}$ such that the mixing coefficient $\phi(m) \leq \Psi m^{-\lambda}$ with $\Psi > 0$ and $\lambda \geq 2$ being universal constants, there exists a metric $d_G(P, Q)$ for probability measure P and Q in $\mathbf{P}(\Psi, \lambda)$ that satisfies the property (50) and such that convergence in d_G implies the weak convergence of the measures. Thus, in particular, for probability measures in a parametric family*

$$d_G(\mathbb{P}_{\theta'}, \mathbb{P}_\theta) \to 0 \Rightarrow \mathbb{P}_{\theta'} \to \mathbb{P}_\theta. \tag{54}$$

Proof. We extend the proof for the existence result in Proposition 6.2 in Clarke and Barron (1990) to allow for weak dependence. Let $\{F_i \,:\, i = 1, 2, \cdots\}$ be the countable field of sets generated by balls of the form $\{x :\ d_{\mathcal{X}}(x, s_{j_1}) \leq 1/j_2\}$ for $j_1, j_2 = 1, 2, \cdots$, where $d_{\mathcal{X}}$ denotes the metric for the space $\mathcal{X}$ and $s_1, s_2, \cdots$ is a countable dense sequence of points in $\mathcal{X}$. Define a metric on the space of probability measures as follows

$$d_G(P, Q) = \sum_{i=1}^{\infty} 2^{-i}|PF_i - QF_i|.$$

According to Gray (1988, Page 251–253), if $d_G(\widehat{\mathbb{P}}_n, \mathbb{P}) \to 0$, then $\widehat{\mathbb{P}}_n$ converges weakly to $\mathbb{P}$. Now, for any $\epsilon > 0$, we choose $k \geq 1 - \ln(\epsilon) / \ln(2)$. Thus,

$$d_G(\widehat{\mathbb{P}}_n, \mathbb{P}) \leq \sum_{i=1}^{k} 2^{-i}\left|\widehat{\mathbb{P}}_n F_i - \mathbb{P}F_i\right| + \sum_{i=k+1}^{\infty} 2^{-i} \leq \max_{1\leq i\leq k}\left|\widehat{\mathbb{P}}_n F_i - \mathbb{P}F_i\right| + \epsilon/2.$$

Then, we have

$$\mathbb{P}\left\{d_G(\widehat{\mathbb{P}}_n,\mathbb{P})>\epsilon\right\}\leq\mathbb{P}\left\{\max_{1\leq i\leq k}\left|\widehat{\mathbb{P}}_nF_i-\mathbb{P}F_i\right|>\epsilon/2\right\}\leq\sum_{i=1}^{k}\mathbb{P}\left\{\left|\widehat{\mathbb{P}}_nF_i-\mathbb{P}F_i\right|>\epsilon/2\right\}$$

The Hoeffding-type inequality for uniform mixing process (see, e.g. Roussas, 1996, Theorem 4.1) guarantees that there exists $C_1>0$ and $\xi>0$ such that

$$\sup_{\mathbb{P}\in\mathbf{P}(\Psi,\lambda)}\mathbb{P}\left\{\left|\widehat{\mathbb{P}}_nF_i-\mathbb{P}F_i\right|>\epsilon/2\right\}\leq C_1e^{-\xi n}.$$

Thus,

$$\mathbb{P}\left\{d_G(\widehat{\mathbb{P}}_n,\mathbb{P})>\epsilon\right\}\leq Ce^{-\xi n},\tag{55}$$

with $C=kC_1$. □

Lemma 6 extends the large deviation result of Schwartz (1965, Lemma 6.1) to allow time dependence in the data process. Let $\mathbf{z^n}=(\mathbf{z_1},\cdots,\mathbf{z_n})$ be strictly stationary and uniform mixing with $\phi(m)=O\left(m^{-\lambda}\right)$ for some positive λ. The process $\mathbf{z^n}$ have the joint density $\mathbb{P}_{\theta,n}$. Denote the mixture distribution of the parametric family $\mathbb{P}_{\theta,n}$ with respect to the conditional prior distribution $\pi_{\mathcal{P}}(\cdot|\mathcal{N}^{\complement})$ to be $\mathbb{P}_{\mathcal{N}^{\complement},n}$ with density $\pi_{\mathcal{P}}(\mathbf{z^n}|\mathcal{N}^{\complement})$. More precisely, we define

$$\pi_{\mathcal{P}}(\mathbf{z^n}|\mathcal{N}^{\complement})\equiv\int_{\mathcal{N}^{\complement}}\pi_{\mathcal{P}}(\mathbf{z^n}|\theta)\pi_{\mathcal{P}}(\theta|\mathcal{N}^{\complement})\mathrm{d}\theta.\tag{56}$$

Lemma 6. *Assume that the mixing coefficient power $\lambda>2$. Suppose there exist strongly uniformly exponentially consistent (SUEC) tests of hypothesis $\theta\in\mathcal{N}_0$ against the alternative $\theta\in\mathcal{N}^{\complement}$ such that $\mathcal{N}_0\subset\mathcal{N}$ with $d_L(\mathcal{N}_0,\mathcal{N}^{\complement})\geq\delta$ for some $\delta>0$. Then, there exists $\xi>0$ and a positive integer k such that, for all $\theta\in\mathcal{N}_0$,*

$$\left|\left|\mathbb{P}_{\theta,n}-\mathbb{P}_{\mathcal{N}^{\complement},n}\right|\right|_{TV}\geq 2(1-2e^{-\xi m}),\quad \textit{where } m+k\leq n.$$

Proof. We assume that there exists a sequence of SUEC tests, denoted by $\{\mathcal{A}_n\}$, for any sample with size n. Then, there exists a positive integer k such that for all $n\geq k$

$$\mathbb{P}_{\theta,n}\mathcal{A}_n<\frac{1}{8}\ \text{ for all }\theta\in\mathcal{N}_0\quad\text{and}\tag{57}$$

$$\mathbb{P}_{\theta',n}\mathcal{A}_n>1-\frac{1}{8}\ \text{ for all }\theta'\in\mathcal{N}^{\complement}.\tag{58}$$

For each $j=1,2,\cdots$, we define

$$\mathcal{A}_{k,j}=\mathcal{A}_k\left(\mathbf{z_{j+1}},\cdots,\mathbf{z_{j+k}}\right),\tag{59}$$

then, according to Lemma 1,

$$Y_m = \frac{1}{m}\sum_{j=1}^{m} \mathcal{A}_{k,j} \tag{60}$$

is an average of strictly stationary and uniform mixing such that $\phi(m) \le \phi^* m^{-\lambda}$. The expectation of Y_m, under distribution $\mathbb{P}_{\theta,n}$, is $\mu(\theta)$ with

$$\mu(\theta) = \begin{cases} < \frac{1}{8} & \text{if } \theta \in \mathcal{N}_0 \\ > \frac{7}{8} & \text{if } \theta \in \mathcal{N}^{\complement}. \end{cases} \tag{61}$$

By Corollary 2 and the uniform mixing conditions with $\lambda > 2$, together with the fact that $\mathcal{A}_{k,j} \in [0,1]$, we know that the assumptions of Theorem 2.4 in White and Domowitz (1984) holds and hence the CLT for the time series holds, i.e.

$$m^{1/2} Y_m \xrightarrow{d} N(\mu(\theta), V(\theta)), \tag{62}$$

where $V(\theta) = \lim_{m\to+\infty} \mathbb{P}_{\theta,n}\left[m^{1/2}\sum_{j=1}^{m}(\mathcal{A}_{k,j} - \mu(\theta))\right]^2$. From Lemma 3, we know that $V(\theta) \le V^* < \infty$. Therefore, the moment generating functions converge

$$m^{-1}\ln \mathbb{P}_{\theta,n} e^{tmY_m} \to \frac{1}{2}t^2 V(\theta). \tag{63}$$

On the one hand, when $\theta \in \mathcal{N}^{\complement}$, we have $\mu(\theta) > \frac{1}{4}$, according to Theorem 8.1.1 of Taniguchi and Kakizawa (2000), we can achieve the following large deviation result

$$\lim_{m\to+\infty} m^{-1}\ln \mathbb{P}_{\theta,n}\left\{Y_m \le \frac{1}{4}\right\} = -\frac{1}{32V(\theta)} \le -\frac{1}{32V^*}. \tag{64}$$

Therefore, there exists $\xi_1 > 0$ such that

$$\mathbb{P}_{\theta,n}\left\{Y_m \le \frac{1}{4}\right\} \le e^{-\xi_1 m} \quad \text{for all } \theta \in \mathcal{N}^{\complement}. \tag{65}$$

Thus,

$$\mathbb{P}_{\mathcal{N}^{\complement},n}\left\{Y_m \le \frac{1}{4}\right\} = \int_{\mathcal{N}^{\complement}} \mathbb{P}_{\theta,n}\left\{Y_m \le \frac{1}{4}\right\}\pi_{\mathcal{P}}(\theta|\mathcal{N}^{\complement})\mathrm{d}\theta \le e^{-\xi_1 m}, \quad \text{for } n \ge m+k. \tag{66}$$

On the other hand, when $\theta \in \mathcal{N}_0$, we have $\mu(\theta) < \frac{1}{4}$, according to Theorem 8.1.1 of Taniguchi and Kakizawa (2000), we obtain the large deviation result

$$\lim_{m\to+\infty} m^{-1}\ln \mathbb{P}_{\theta,n}\left\{Y_m \ge \frac{1}{4}\right\} = -\frac{1}{32V(\theta)} \le -\frac{1}{32V^*}. \tag{67}$$

Therefore, there exists $\xi_2 > 0$ such that

$$\mathbb{P}_{\theta,n}\left\{Y_m \ge \frac{1}{4}\right\} \le e^{-\xi_2 m}. \tag{68}$$

For $\xi = \min(\xi_1, \xi_2)$, it follows that

$$\left|\left|\mathbb{P}_{\theta,n} - \mathbb{P}_{\mathcal{N}^{\complement},n}\right|\right|_{TV} = 2 \sup_{A \in \mathcal{F}} \left|\mathbb{P}_{\theta,n} A - \mathbb{P}_{\mathcal{N}^{\complement},n} A\right| \geq 2(1 - 2e^{-\xi m}) \tag{69}$$

for $m + k \leq n$ by considering $A = \{Y_m \leq \frac{1}{4}\}$. □

We introduce Le Cam's theory on hypothesis testing (see, e.g. Le Cam and Yang, 2000, Chapter 8).

Lemma 7. *Under the regularity conditions in Subsection 1.4, there are test functions $\mathcal{A}_n$ and positive coefficients C, ξ, ϵ and K such that $\mathbb{P}_{0,n}(1 - \mathcal{A}_n) \to 0$ and $\mathbb{P}_{\theta,n} \mathcal{A}_n \leq C e^{-n\xi|\theta - \theta_0|^2/2}$ for all θ such that $K/\sqrt{n} \leq |\theta - \theta_0| \leq \epsilon$.*

Proof. For all $z \in \mathbb{R}^{K(D_x+D_y)}$, define the rectangular $F_z \equiv (-\infty, z_1] \times (-\infty, z_2] \times \cdots \times (-\infty, z_{K(D_x+D_y)}]$. The empirical process is defined as $\widehat{\Pi}_n(z) \equiv \widehat{\mathbb{P}}_n F_z = n^{-1} \sum_{t=1}^n \mathbf{1}_{\{\mathbf{z_t} \in F_z\}}$. We define $\Pi_\theta(z) \equiv \mathbb{P}_\theta F_z$. According to Le Cam and Yang (2000, Page 250), there exists positive constants c and ϵ such that $\sup_x |\Pi_{\theta_0}(x) - \Pi_\theta(x)| > c|\theta - \theta_0|$ for $|\theta - \theta_0| \leq \epsilon$. Denote the expectation to be $\mu_n(\theta) \equiv \mathbb{E}_{\mathbb{P}_\theta} \sup_x |\Pi_{\theta_0}(x) - \Pi_\theta(x)|$. By the classical result of weak convergence for the Kolmogorov-Smirnov statistic $\sqrt{n} \sup_x |\Pi_{\theta_0}(x) - \Pi_\theta(x)|$, we know that there exists a large constant M such that $\mu_n(\theta) \leq \frac{M}{2c\sqrt{n}}$ for all $|\theta - \theta_0| \leq \epsilon$. We choose $K \equiv \frac{4M}{c}$. Consider the test functions $\mathcal{A}_n = \{\sup_z |\Pi_{\theta_0}(z) - \Pi_\theta(z)| < K/\sqrt{n}\}$. Using triangular inequality, we obtain

$$\mathbb{P}_{\theta,n} \mathcal{A}_n \leq \mathbb{P}_{\theta,n} \left\{ \frac{c}{2} |\theta - \theta_0| \leq \sup_z \left| \widehat{\Pi}_n(z) - \Pi_\theta(z) \right| - \mu_n(\theta) \right\}$$

Using Dvoretzky-Kiefer-Wolfowitz type inequality for uniform mixing variables in Samson (2000, Theorem 3), we can show there exists $\xi > 0$ such that

$$\mathbb{P}_{\theta,n} \left\{ \frac{c}{2} |\theta - \theta_0| \leq \sup_z \left| \widehat{\Pi}_n(z) - \Pi_\theta(z) \right| - \mu_n(\theta) \right\} \leq e^{-\xi|\theta - \theta_0|^2/2}$$

for all $K/\sqrt{n} \leq |\theta - \theta_0| \leq \epsilon$. Thus, $\mathbb{P}_{\theta,n} \mathcal{A}_n \leq e^{-\xi|\theta-\theta_0|^2/2}$ for all $K/\sqrt{n} \leq |\theta - \theta_0| \leq \epsilon$. By the same inequality, it is straightforward to get $\mathbb{P}_{0,n}(1 - \mathcal{A}_n) \to 0$. □

1.6 Basic Properties of Limited-Information Likelihoods

The MLE for Limited-Information Likelihood $\mathbb{Q}_{\gamma,n}$ The empirical log-likelihood function is defined as

$$\widehat{\mathcal{L}}_n(\gamma) \equiv \frac{1}{n} \sum_{t=1}^n \ln \pi_{\mathbb{Q}}(\mathbf{x_t}, \mathbf{y_t}; \gamma), \quad \forall\ \gamma. \tag{70}$$

The maximum likelihood estimator $\hat{\theta}^{\mathbb{Q}}$ is defined as follows:

$$\hat{\gamma}^{\mathbb{Q}}_{\mathrm{ML}} \equiv \operatorname*{argmax}_\gamma \widehat{\mathcal{L}}_n(\gamma) = \operatorname*{argmax}_\gamma \eta_{\mathbb{Q}}(\gamma)^T \left[\frac{1}{n} \sum_{t=1}^n g_{\mathbb{Q}}(\gamma; \mathbf{x_t}, \mathbf{y_t}) \right] - A_{\mathbb{Q}}(\gamma). \tag{71}$$

We shall show that the MLE $\hat{\gamma}_{\text{ML}}^{\mathbb{Q}}$ is consistent and asymptotic normal with asymptotic efficient variance-covariance matrix. This is important to justify the use of our limited-information likelihood $\mathbb{Q}_\gamma$ of (16) in the Bayesian paradigm as a valid likelihood.

The first-order condition is $0 = \nabla \widehat{\mathcal{L}}_n(\hat{\gamma}_{\text{ML}}^{\mathbb{Q}})$ which is

$$0 = \left[\nabla \eta_{\mathbb{Q}}(\hat{\gamma}_{\text{ML}}^{\mathbb{Q}})\right]^T \left[\frac{1}{n}\sum_{t=1}^{n} g_{\mathbb{Q}}(\hat{\gamma}_{\text{ML}}^{\mathbb{Q}}; \mathbf{x_t}, \mathbf{y_t})\right] + \left[\frac{1}{n}\sum_{t=1}^{n} \nabla g_{\mathbb{Q}}(\hat{\gamma}_{\text{ML}}^{\mathbb{Q}}; \mathbf{x_t}, \mathbf{y_t})\right]^T \eta_{\mathbb{Q}}(\hat{\gamma}_{\text{ML}}^{\mathbb{Q}}) - \nabla A_{\mathbb{Q}}(\hat{\gamma}_{\text{ML}}^{\mathbb{Q}}). \tag{72}$$

The definition of $\hat{\gamma}_{\text{ML}}^{\mathbb{Q}}$ in (71) and the properties of expected log-likelihood $\mathcal{L}(\gamma)$ plays a vital role in establishing the consistency of $\hat{\gamma}_{\text{ML}}^{\mathbb{Q}}$. The first-order condition (72) is crucial for obtaining the asymptotic normality of $\hat{\gamma}_{\text{ML}}^{\mathbb{Q}}$.

Now, let's consider the expected log-likelihood of $\mathbb{Q}_\gamma$:

$$\mathcal{L}(\gamma) \equiv \eta_{\mathbb{Q}}(\gamma)^T g_{\mathbb{Q}}(\gamma) - A_{\mathbb{Q}}(\gamma). \tag{73}$$

where $g_{\mathbb{Q}}(\gamma) \equiv \mathbb{E}\left[g_{\mathbb{Q}}(\gamma; \mathbf{x}, \mathbf{y})\right]$. It holds that $\mathcal{L}(\gamma_0) = 0$ since $\eta_{\mathbb{Q}}(\gamma_0) = 0$ and $A_{\mathbb{Q}}(\gamma_0) = 0$. Further, it holds that $\nabla\mathcal{L}(\gamma_0) = 0$ since $\nabla\mathcal{L}(\gamma_0) = \left[\nabla\eta_{\mathbb{Q}}(\gamma_0)\right]^T g_{\mathbb{Q}}(\gamma_0) + \left[\nabla g_{\mathbb{Q}}(\gamma_0)\right]^T \eta_{\mathbb{Q}}(\gamma_0) - \nabla A_{\mathbb{Q}}(\gamma_0)$, and $g_{\mathbb{Q}}(\gamma_0) = 0$ and $\nabla A_{\mathbb{Q}}(\gamma_0) = 0$. The Jacobian matrix of $\mathcal{L}(\gamma)$ evaluated at γ_0 is equal to the negative Fisher information matrix $\mathbf{I}_{\mathbb{Q}}(\gamma_0)$:

$$\nabla^2\mathcal{L}(\gamma_0) = -\nabla^2 A_{\mathbb{Q}}(\gamma_0) = -\nabla\eta_{\mathbb{Q}}(\gamma_0)^T \nabla g_{\mathbb{Q}}(\gamma_0) = -G_{\mathbb{Q}}^T S_{\mathbb{Q}}^{-1} G_{\mathbb{Q}} = -\mathbf{I}_{\mathbb{Q}}(\gamma_0). \tag{74}$$

Here the first equality is simply implied by the formula of $\nabla^2\mathcal{L}(\gamma)$ from (73). The second equality of (74) is due to

$$\nabla A_{\mathbb{Q}}(\gamma) e^{A_{\mathbb{Q}}(\gamma)} = \mathbb{E}\left[\eta_{\mathbb{Q}}(\gamma)^T \nabla g_{\mathbb{Q}}(\gamma; \mathbf{x}, \mathbf{y}) e^{\eta_{\mathbb{Q}}(\gamma)^T g_{\mathbb{Q}}(\gamma; \mathbf{x}, \mathbf{y})}\right] \quad \forall\ \gamma, \tag{75}$$

which is implied by the definition of $A_{\mathbb{Q}}(\gamma)$ and the condition (17). The third equality of (74) is due to the relation $\nabla g_{\mathbb{Q}}(\gamma_0) = G_{\mathbb{Q}}$ and the following relation implied by (17):

$$\nabla\eta_{\mathbb{Q}}(\gamma_0) = S_{\mathbb{Q}}^{-1} G_{\mathbb{Q}}. \tag{76}$$

And, the fourth equality of (74) is simply due to the definition of $\mathbf{I}_{\mathbb{Q}}(\gamma_0)$.

Another important property of expected log-likelihood of $\mathbb{Q}_\gamma$ is the identification of γ_0 in the following sense:

$$\mathcal{L}(\gamma) \leq \mathcal{L}(\gamma_0) = 0, \quad \forall\ \gamma, \tag{77}$$

where the inequality is due to the Jensen's inequality applied to $A_{\mathbb{Q}}(\gamma)$. The Jensen's inequality holds with equality if and only if $\eta_{\mathbb{Q}}(\gamma)^T g_{\mathbb{Q}}(\gamma; \mathbf{x}, \mathbf{y})$ is constant almost surely under $\mathbb{Q}$ which is true only when $\gamma = \gamma_0$.

Consistency We first prove the consistency of the estimator $\hat{\gamma}^{\mathbb{Q}}_{\mathrm{ML}}$. To our knowledge, the theoretical result for the MLE of limited-information likelihood $\mathbb{Q}_\gamma$ based on the principle of minimum Kullback-Leibler divergence is new, though we appeal to the standard approach of Wald (1949) and Wolfowitz (1949).

Proposition 1. *Under Assumptions A1 - A5 and A8 - A9, $\hat{\gamma}^{\mathbb{Q}}_{ML}$ converges to γ_0 in probability.*

Proof. Our goal is to show that for any $\epsilon > 0$ there exists N such that

$$\mathbb{Q}_n\left\{\hat{\gamma}^{\mathbb{Q}}_{\mathrm{ML}} \in \Omega(\gamma_0, \epsilon)\right\} < \epsilon \text{ for } n \geq N, \tag{78}$$

where $\Omega(\gamma_0, \epsilon) \equiv \{\gamma : \ |\gamma - \gamma_0| > \epsilon\}$ is the open ball centered at γ_0 with radius ϵ.

For all $\epsilon > 0$, it holds that (due to the definition of $\hat{\gamma}^{\mathbb{Q}}_{\mathrm{ML}}$ in (71))

$$\mathbf{1}\{\hat{\gamma}^{\mathbb{Q}}_{\mathrm{ML}} \in \Omega(\gamma_0, \epsilon)\} \leq \mathbf{1}\{\sup_{\gamma \in \Omega(\gamma_0,\epsilon)^{\complement}} \widehat{\mathcal{L}}_n(\gamma) \geq \widehat{\mathcal{L}}_n(\gamma_0)\}. \tag{79}$$

Moreover, for all $\epsilon > 0$ and $h > 0$, it holds that

$$\begin{aligned}
\mathbf{1}\{\sup_{\gamma \in \Omega(\gamma_0,\epsilon)^{\complement}} \widehat{\mathcal{L}}_n(\gamma) \geq \widehat{\mathcal{L}}_n(\gamma_0)\} &= \mathbf{1}\{\widehat{\mathcal{L}}_n(\gamma_0) < -h\}\mathbf{1}\{\sup_{\gamma \in \Omega(\gamma_0,\epsilon)^{\complement}} \widehat{\mathcal{L}}_n(\gamma) \geq \widehat{\mathcal{L}}_n(\gamma_0)\} \\
&\quad + \mathbf{1}\{\widehat{\mathcal{L}}_n(\gamma_0) \geq -h\}\mathbf{1}\{\sup_{\gamma \in \Omega(\gamma_0,\epsilon)^{\complement}} \widehat{\mathcal{L}}_n(\gamma) \geq \widehat{\mathcal{L}}_n(\gamma_0)\} \\
&\leq \mathbf{1}\{\widehat{\mathcal{L}}_n(\gamma_0) < -h\} + \mathbf{1}\{\sup_{\gamma \in \Omega(\gamma_0,\epsilon)^{\complement}} \widehat{\mathcal{L}}_n(\gamma) \geq -h\}.
\end{aligned} \tag{80}$$

Combining (78) and (80), it suffices to show that for all $\epsilon > 0$ there exists $h > 0$ and N such that for $n \geq N$

$$\max\left\{\mathbb{Q}_n\left\{\widehat{\mathcal{L}}_n(\gamma_0) < -h\right\}, \mathbb{Q}_n\left\{\sup_{\gamma \in \Omega(\gamma_0,\epsilon)^{\complement}} \widehat{\mathcal{L}}_n(\gamma) \geq -h\right\}\right\} < \epsilon/2. \tag{81}$$

The first probabilistic bound in (81) is straightforward due to the LLN result

$$\operatorname*{wlim}_{n\to\infty} \widehat{\mathcal{L}}_n(\gamma_0) = \mathcal{L}(\gamma_0) = 0. \tag{82}$$

Now, we consider the second probabilistic bound in (81). From the dominance condition (Assumption A3) and the remark of identification condition (Assumption A5), we know that

$$\lim_{\delta\to 0} \sup_{\gamma' \in \Omega(\gamma,\delta)} \ln \pi_{\mathbb{Q}}(\mathbf{x}, \mathbf{y}; \gamma') = \ln \pi_{\mathbb{Q}}(\mathbf{x}, \mathbf{y}; \gamma) \quad \text{ a.s. } \mathbb{Q} \text{ for all } \gamma. \tag{83}$$

Thus, we have the following Dominance Convergence result:

$$\lim_{\delta\to 0} \mathbb{E}\left[\sup_{\gamma' \in \Omega(\gamma,\delta)} \ln \pi_{\mathbb{Q}}(\mathbf{x}, \mathbf{y}; \gamma')\right] = \mathcal{L}(\gamma). \tag{84}$$

Therefore, there exists a small constant $\epsilon' > 0$ such that for all γ in the compact set $\Omega(\gamma_0, \epsilon)^{\complement} \equiv (\Theta \times \Psi) \setminus \Omega(\gamma_0, \epsilon)$ (Assumption A6)

$$\lim_{\delta \to 0} \mathbb{E} \sup_{\gamma' \in \Omega(\gamma, \delta)} \ln \pi_{\mathbb{Q}}(\mathbf{x}, \mathbf{y}; \gamma') < -3\epsilon'. \tag{85}$$

Thus, for each $\gamma \in \Omega(\gamma_0, \epsilon)^{\complement}$ there exists $\delta_\gamma > 0$ such that

$$\mathbb{E} \sup_{\gamma' \in \Omega(\gamma, \delta_\gamma)} \ln \pi_{\mathbb{Q}}(\mathbf{x}, \mathbf{y}; \gamma') < -2\epsilon'. \tag{86}$$

The collection of small open balls $\Omega(\gamma, \delta_\gamma)$ cover the compact set $\Omega(\gamma_0, \epsilon)^{\complement}$, thus there exists a finite cover consists of small open balls $\{\Omega(\gamma_j, \delta_j) : \ j = 1, \cdots, m_c\}$ where m_c is a constant integer. Take $h = \epsilon'$ in the second probabilistic bound of (81), we have

$$\begin{aligned}
\mathbb{Q}_n \left\{ \sup_{\gamma \in \Omega(\gamma_0, \epsilon)^{\complement}} \widehat{\mathcal{L}}_n(\gamma) \geq -\epsilon' \right\} &\leq \sum_{j=1}^{m_g} \mathbb{Q}_n \left\{ \sup_{\gamma' \in \Omega(\gamma_j, \delta_j)} \widehat{\mathcal{L}}_n(\gamma') \geq -\epsilon' \right\} \\
&\leq \sum_{j=1}^{m_g} \mathbb{Q}_n \left\{ \frac{1}{n} \sum_{t=1}^{n} \sup_{\gamma' \in \Omega(\gamma_j, \delta_j)} \ln \pi_{\mathbb{Q}}(\mathbf{x_t}, \mathbf{y_t}; \gamma') \geq -\epsilon' \right\}.
\end{aligned}$$

According to the LLN and the inequality (86), it follows that there exists N such that for all $j = 1, \cdots, m_g$:

$$\mathbb{Q}_n \left\{ \frac{1}{n} \sum_{t=1}^{n} \sup_{\gamma' \in \Omega(\gamma_j, \delta_j)} \ln \pi_{\mathbb{Q}}(\mathbf{x_t}, \mathbf{y_t}; \gamma') \geq -\epsilon' \right\} < \frac{\epsilon}{2m_g}, \quad \forall \, n \geq N. \tag{87}$$

Thus, it holds that

$$\mathbb{Q}_n \left\{ \sup_{\gamma \in \Omega(\gamma_0, \epsilon)^{\complement}} \widehat{\mathcal{L}}_n(\gamma) \geq -\epsilon' \right\} < \frac{\epsilon}{2}, \quad \forall \, n \geq N. \tag{88}$$

Therefore, we have finished proving the proposition. □

Asymptotic Normality The asymptotic normality is also new, though the result is not surprising given the asymptotic normality result in Kitamura and Stutzer (1997) for the exponential tilted (ET) estimator.

Proposition 2. *Under Assumptions A1 - A5 and A8 - A9, the asymptotic normality of $\hat{\gamma}^{\mathbb{Q}}_{ML}$ holds:*

$$\operatorname*{wlim}_{n \to \infty} \sqrt{n} \left(\hat{\gamma}^{\mathbb{Q}}_{ML} - \gamma_0 \right) = N\left(0, \mathbf{I}_{\mathbb{Q}}(\gamma_0)^{-1}\right), \quad \textit{where } \mathbf{I}_{\mathbb{Q}}(\gamma_0) = G_{\mathbb{Q}}^T S_{\mathbb{Q}}^{-1} G_{\mathbb{Q}}. \tag{89}$$

Proof. Let's consider the following three Taylor expansions around γ_0:

$$\eta_{\mathcal{Q}}(\hat{\gamma}_{\mathrm{ML}}^{\mathcal{Q}}) = \nabla \eta_{\mathcal{Q}}(\tilde{\gamma}_\eta)(\hat{\gamma}_{\mathrm{ML}}^{\mathcal{Q}} - \gamma_0) \;\text{ and }\; \nabla A_{\mathcal{Q}}(\hat{\gamma}_{\mathrm{ML}}^{\mathcal{Q}}) = \nabla^2 A_{\mathcal{Q}}(\tilde{\gamma}_a)(\hat{\gamma}_{\mathrm{ML}}^{\mathcal{Q}} - \gamma_0) \;\text{ and}$$
$$\frac{1}{n}\sum_{t=1}^{n} g_{\mathcal{Q}}(\hat{\gamma}_{\mathrm{ML}}^{\mathcal{Q}}; \mathbf{x_t}, \mathbf{y_t}) = \frac{1}{n}\sum_{t=1}^{n} g_{\mathcal{Q}}(\gamma_0; \mathbf{x_t}, \mathbf{y_t}) + \left[\frac{1}{n}\sum_{t=1}^{n} \nabla g_{\mathcal{Q}}(\tilde{\gamma}_g; \mathbf{x_t}, \mathbf{y_t})\right](\hat{\gamma}_{\mathrm{ML}}^{\mathcal{Q}} - \gamma_0),$$

where $\tilde{\gamma}_\eta$, $\tilde{\gamma}_g$, and $\tilde{\gamma}_a$ are all between $\hat{\gamma}_{\mathrm{ML}}^{\mathcal{Q}}$ and γ_0.

Plugging the three Taylor expansions above into the first-order condition of $\hat{\gamma}_{\mathrm{ML}}^{\mathcal{Q}}$ in (72), it follows that

$$0 = \left[\nabla \eta_{\mathcal{Q}}(\hat{\gamma}_{\mathrm{ML}}^{\mathcal{Q}})\right]^T \left[\frac{1}{n}\sum_{t=1}^{n} g_{\mathcal{Q}}(\gamma_0; \mathbf{x_t}, \mathbf{y_t})\right] + \left[\nabla \eta_{\mathcal{Q}}(\hat{\gamma}_{\mathrm{ML}}^{\mathcal{Q}})\right]^T \left[\frac{1}{n}\sum_{t=1}^{n} \nabla g_{\mathcal{Q}}(\tilde{\gamma}_g; \mathbf{x_t}, \mathbf{y_t})\right](\hat{\gamma}_{\mathrm{ML}}^{\mathcal{Q}} - \gamma_0)$$
$$+ \left[\frac{1}{n}\sum_{t=1}^{n} \nabla g_{\mathcal{Q}}(\hat{\gamma}_{\mathrm{ML}}^{\mathcal{Q}}; \mathbf{x_t}, \mathbf{y_t})\right]^T \nabla \eta_{\mathcal{Q}}(\tilde{\gamma}_\eta)(\hat{\gamma}_{\mathrm{ML}}^{\mathcal{Q}} - \gamma_0) - \nabla^2 A(\tilde{\gamma}_a)(\hat{\gamma}_{\mathrm{ML}}^{\mathcal{Q}} - \gamma_0). \tag{90}$$

The consistency of $\hat{\gamma}_{\mathrm{ML}}^{\mathcal{Q}}$ implies that, in probability

$$\nabla \eta_{\mathcal{Q}}(\tilde{\gamma}_\eta) \to \nabla \eta_{\mathcal{Q}}(\gamma_0) = S_{\mathcal{Q}}^{-1} G_{\mathcal{Q}}, \text{ and } \nabla \eta_{\mathcal{Q}}(\hat{\gamma}_{\mathrm{ML}}^{\mathcal{Q}}) \to \nabla \eta_{\mathcal{Q}}(\gamma_0) = S_{\mathcal{Q}}^{-1} G_{\mathcal{Q}}, \text{ and}$$
$$\nabla^2 A_{\mathcal{Q}}(\tilde{\gamma}_a) \to \nabla^2 A_{\mathcal{Q}}(\gamma_0) = \mathbf{I}_{\mathcal{Q}}(\gamma_0) = G_{\mathcal{Q}}^T S_{\mathcal{Q}}^{-1} G_{\mathcal{Q}}.$$

The uniform law of large numbers (ULLN) (see, e.g., White and Domowitz, 1984) implies that, in probability,

$$\frac{1}{n}\sum_{t=1}^{n} \nabla g_{\mathcal{Q}}(\tilde{\gamma}_g; \mathbf{x_t}, \mathbf{y_t}) \to G_{\mathcal{Q}}. \tag{91}$$

Thus, according to (90), it holds that

$$\sqrt{n}(\hat{\gamma}_{\mathrm{ML}}^{\mathcal{Q}} - \gamma_0) = -\left(G_{\mathcal{Q}}^T S_{\mathcal{Q}}^{-1} G_{\mathcal{Q}}\right)^{-1} G_{\mathcal{Q}}^T S_{\mathcal{Q}}^{-1} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_{\mathcal{Q}}(\gamma_0; \mathbf{x_t}, \mathbf{y_t}) + o_p(1), \tag{92}$$

where the CLT implies

$$\operatorname*{wlim}_{n\to\infty} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_{\mathcal{Q}}(\gamma_0; \mathbf{x_t}, \mathbf{y_t}) = N(0, S_{\mathcal{Q}}). \tag{93}$$

Therefore, we have finished proving the result of the proposition. □

The MLE for Limited-Information Likelihood $\mathbb{P}_{\theta,n}$

Proposition 3. *Under Assumptions A1 - A5 and A8 - A9, the asymptotic normality of $\hat{\theta}_{ML}^{\mathcal{Q}}$ holds:*

$$\operatorname*{wlim}_{n\to\infty} \sqrt{n}\left(\hat{\theta}_{ML}^{\mathcal{P}} - \theta_0\right) = N\left(0, \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\right), \quad \textit{where } \mathbf{I}_{\mathcal{P}}(\theta_0) = G_{\mathcal{P}}^T S_{\mathcal{P}}^{-1} G_{\mathcal{P}}. \tag{94}$$

Proof. It is the same as the proofs for the full model $\mathcal{Q}$ above. □

Asymptotic Normality of LIL Posteriors

Proposition 4. *Under Assumptions A1 - A6 in Appendix 1.4, it holds that*

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

Proof. We extend the proof of Theorem 2.1 in Clarke (1999) which is under the i.i.d. condition. However, we have to adjust two parts of their proof, to extend the result to the case that the observations are time series with uniform mixing. The first part is to show that $\sup_{\theta\in\Theta}|\widehat{H}_{\mathcal{P},n}(\theta)| = O_p(1)$ where

$$\widehat{H}_{\mathcal{P},n}(\theta) \equiv -\frac{1}{n}\sum_{t=1}^{n} \ln \pi_{\mathcal{P}}(\mathbf{x_t};\theta). \tag{95}$$

When n is large enough, we obtain that

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

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 $\sup_{\theta\in\Theta}|\widehat{H}_{\mathcal{P},n}(\theta)| = O_p(1)$. The second part is to show that

$$\int u^T u \left|\pi_{\mathcal{P}}(\hat{\theta}^{\mathcal{P}}_{\text{ML}} + u/\sqrt{n}|\mathbf{x^n}) - \varphi_{\mathcal{P}}(u)\right| \mathrm{d}u \to 0 \quad \text{in} \quad \mathbb{P}_n \tag{96}$$

where $\varphi_{\mathcal{P}}(u) = \sqrt{\det\mathbf{I}_{\mathcal{P}}(\theta_0)/(2\pi)^{D_\Theta}}\exp\left[-\frac{1}{2}u^T\mathbf{I}_{\mathcal{P}}(\theta_0)u\right]$. In Clarke (1999), it shows that when $\mathbf{x_1},\cdots,\mathbf{x_n}$ are i.i.d., the limit result (96) is satisfied under the regularity conditions in Assumptions A3 - A6. To extend this limit result to allow 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 1.4. □

Proposition 5. *Under Assumptions A1 - A6 in Appendix 1.4, it holds that*

$$\mathbf{D}_{KL}(\pi_{\mathcal{Q}}(\gamma|\mathbf{x^n},\mathbf{y_n})||N(\hat{\gamma}^{\mathcal{Q}}_{ML}, n^{-1}\mathbf{I}_{\mathcal{Q}}(\gamma_0)^{-1})) \to 0 \quad in\ \mathbb{Q}_n.$$

Proof. It is the same as the proofs of Proposition 4. □

Proposition 6. *Denote $\theta_{(1)}$ to be the first element in θ. Then, we have*

$$\mathbf{D}_{KL}(\pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})||\pi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n})) \leq \mathbf{D}_{KL}(\pi_{\mathcal{Q}}(\theta|\mathbf{x^n},\mathbf{y^n})||\pi_{\mathcal{P}}(\theta|\mathbf{x^n})). \tag{97}$$

Proof. Denote $\theta_{(-1)}$ to be the vector containing all parameters other than $\theta_{(1)}$ in θ. Then, we have

$$\begin{aligned}&\mathbf{D}_{KL}\left(\pi_{\mathcal{Q}}(\theta|\mathbf{x}^{\mathbf{n}},\mathbf{y}^{\mathbf{n}})||\pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}})\right)\\&\quad=\mathbb{E}^{\theta_{(1)}}\mathbf{D}_{KL}\left(\pi_{\mathcal{Q}}(\theta_{(-1)}|\mathbf{x}^{\mathbf{n}},\mathbf{y}^{\mathbf{n}},\theta_{(1)})||\pi_{\mathcal{P}}(\theta_{(-1)}|\mathbf{x}^{\mathbf{n}},\theta_{(1)})\right)+\mathbf{D}_{KL}\left(\pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x}^{\mathbf{n}},\mathbf{y}^{\mathbf{n}})||\pi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x}^{\mathbf{n}})\right).\end{aligned} \tag{98}$$

Because the term (98) is nonnegative, the result of the proposition is proved. □

Corollary 3. *Denote* $\mathbf{v}\equiv\nabla f(\theta_0)$*. Under the assumptions in Subsection 1.4, we have*

$$\mathbf{D}_{KL}\left(\pi_{\mathcal{P}}(f(\theta)|\mathbf{x}^{\mathbf{n}})||N(f(\widehat{\theta}^{\mathcal{P}}),n^{-1}\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v})\right)\to 0 \quad \text{in } \mathbb{P}_n, \tag{99}$$

and

$$\mathbf{D}_{KL}\left(\pi_{\mathcal{Q}}(f(\theta)|\mathbf{x}^{\mathbf{n}},\mathbf{y}^{\mathbf{n}})||N(f(\widehat{\theta}^{\mathcal{Q}}),n^{-1}\mathbf{v}^T\mathbf{I}_{\mathcal{Q}}(\theta_0)^{-1}\mathbf{v})\right)\to 0 \quad \text{in } \mathbb{Q}_n. \tag{100}$$

Proof. Because of Assumption **FF** and the assumptions in Subsection 1.4 are invariant under invertible and second-order smooth transformations, without loss of generality, we assume that $f(\theta)=(\theta_{(1)},\theta_{(2)},\cdots,\theta_{(D_f)})$. Applying Proposition 4, Proposition 5 and Proposition 6, we know that the results hold. □

1.7 Information Matrices of Limited-Information Likelihoods

Definition 2. *We define the following quantities which are expected negative log (limited information) likelihood will be used repeatedly in the proofs.*

$$H_{\mathcal{Q}}(\gamma)\equiv-\int\pi_{\mathcal{Q}}(\mathbf{x},\mathbf{y}|\gamma_0)\ln\pi_{\mathcal{Q}}(\mathbf{x},\mathbf{y};\gamma)\,d\mathbf{x}d\mathbf{y}, \quad \text{and} \quad H_{\mathcal{P}}(\theta)\equiv-\int\pi_{\mathcal{P}}(\mathbf{x}|\theta_0)\ln\pi_{\mathcal{P}}(\mathbf{x};\theta)\,d\mathbf{x}. \tag{101}$$

Define the sample correspondences as

$$\widehat{H}_{\mathcal{Q},n}(\gamma)\equiv-\frac{1}{n}\sum_{t=1}^{n}\ln\pi_{\mathcal{Q}}(\mathbf{x_t},\mathbf{y_t};\gamma) \quad \text{and} \quad \widehat{H}_{\mathcal{P},n}(\theta)\equiv-\frac{1}{n}\sum_{t=1}^{n}\ln\pi_{\mathcal{P}}(\mathbf{x_t};\theta). \tag{102}$$

Proposition 7. *Under the regularity conditions in Subsection 1.4, we have*

$$\widehat{H}_{\mathcal{Q},n}(\gamma)\to H_{\mathcal{Q}}(\gamma) \quad \text{and} \quad \widehat{H}_{\mathcal{P},n}(\theta)\to H_{\mathcal{P}}(\theta) \quad \text{a.s. uniformly in } \theta,\gamma.$$

Proof. Simply follows the uniform law of large numbers (ULLN) in White and Domowitz (1984, Theorem 2.3). □

Definition 3. *We define the observed Fisher information matrices as*

$$\widehat{\mathbf{I}}_{\mathcal{P},n}(\theta)\equiv-\frac{1}{n}\nabla^2\ln\pi_{\mathcal{P}}(\mathbf{x}^{\mathbf{n}}|\theta)=-\frac{1}{n}\sum_{t=1}^{n}\nabla^2\ln\pi_{\mathcal{P}}(\mathbf{x_t};\theta)+o_p(1), \tag{103}$$

and

$$\widehat{\mathbf{I}}_{\mathcal{Q},n}(\gamma) \equiv -\frac{1}{n}\nabla^2 \ln \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma) = -\frac{1}{n}\sum_{t=1}^{n} \nabla^2 \ln \pi_{\mathcal{Q}}(\mathbf{x_t}, \mathbf{y_t}; \gamma) + o_p(1), \tag{104}$$

Definition 4. *The empirical score functions are*

$$s_{\mathcal{P},n}(\theta) \equiv \frac{1}{n}\nabla \ln \pi_{\mathcal{P}}(\mathbf{x^n}|\theta) = \frac{1}{n}\sum_{i=1}^{n} \nabla \ln \pi_{\mathcal{P}}(\mathbf{x_t}; \theta) + o_p(1),$$

and

$$s_{\mathcal{Q},n}(\gamma) \equiv \frac{1}{n}\nabla \ln \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma) = \frac{1}{n}\sum_{i=1}^{n} \nabla \ln \pi_{\mathcal{Q}}(\mathbf{x_t}, \mathbf{y_t}; \gamma) + o_p(1),$$

The standardized empirical score functions are

$$S_{\mathcal{P},n}(\theta) = \sqrt{n} s_{\mathcal{P},n}(\theta), \quad \textit{and} \quad S_{\mathcal{Q},n}(\theta) = \sqrt{n} s_{\mathcal{Q},n}(\theta).$$

Proposition 8. *Under the assumptions in Subsection 1.4, the uniform law of large numbers (ULLN) holds:*

$$\sup_{\gamma\in\Gamma} \left|\left|\widehat{\mathbf{I}}_{\mathcal{Q},n}(\gamma) - \mathbf{I}_{\mathcal{Q}}(\gamma)\right|\right|_{\mathcal{S}} \to 0 \quad \textit{in} \quad \mathbb{Q}_n, \tag{105}$$

where

$$\mathbf{I}_{\mathcal{Q}}(\gamma) \equiv -\mathbb{E}\left[\nabla^2 \ln \pi_{\mathcal{Q}}(\mathbf{x}, \mathbf{y}; \gamma)\right].$$

Proof. By applying Theorem 2.3 of White and Domowitz (1984), the ULLN gives

$$-\frac{1}{n}\sum_{t=1}^{n} \nabla^2 \ln \pi_{\mathcal{Q}}(\mathbf{x_t}, \mathbf{y_t}; \gamma) \to \mathbf{I}_{\mathcal{Q}}(\gamma) \quad \text{a.s. uniformly in } \gamma.$$

□

Proposition 9. *Under the assumptions in Subsection 1.4, the uniform law of large numbers (ULLN) holds:*

$$\sup_{\theta\in\Theta} \left|\left|\widehat{\mathbf{I}}_{\mathcal{P},n}(\theta) - \mathbf{I}_{\mathcal{P}}(\theta)\right|\right|_{\mathcal{S}} \to 0 \quad \textit{in} \quad \mathbb{P}_n, \tag{106}$$

where

$$\mathbf{I}_{\mathcal{P}}(\theta) \equiv -\mathbb{E}\left[\nabla^2 \ln \pi_{\mathcal{P}}(\mathbf{x}; \theta)\right].$$

Proof. By applying Theorem 2.3 of White and Domowitz (1984), the ULLN gives

$$-\frac{1}{n}\sum_{t=1}^{n} \nabla^2 \ln \pi_{\mathcal{P}}(\mathbf{x_t}; \theta) \to \mathbf{I}_{\mathcal{P}}(\theta) \quad \text{a.s. uniformly in } \Theta.$$

□

Corollary 4. *Under the assumptions in Subsection 1.4, for any sequence of random variables* $\tilde{\gamma}_n$ *and* $\tilde{\theta}_n$ *such that* $\tilde{\gamma}_n \to \gamma_0$ *and* $\tilde{\theta}_n \to \theta_0$ *in probability, we have*

$$\widehat{\mathbf{I}}_{\mathcal{Q},n}(\tilde{\gamma}_n) \to \mathbf{I}_{\mathcal{Q}}(\gamma_0) \quad \text{in} \quad \mathbb{Q}_n, \quad \text{and} \quad \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}_n) \to \mathbf{I}_{\mathcal{P}}(\theta_0) \quad \text{in} \quad \mathbb{P}_n,$$

where $\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})$ *and* $\widehat{\mathbf{I}}_{\mathcal{Q},n}(\tilde{\theta})$ *are observed Fisher Information matrixes defined in (103) and (104), respectively.*

Proposition 10. *Under the assumptions in Subsection 1.4, it holds that the Hausman test statistic based on ML estimation of the limited-information likelihood satisfies*

$$n(\widehat{\theta}^{\mathcal{P}}_{ML} - \widehat{\theta}^{\mathcal{Q}}_{ML})^T \left[\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1} - \mathbf{I}_{\mathcal{Q}}(\theta_0)^{-1}\right]^{-1} (\widehat{\theta}^{\mathcal{P}}_{ML} - \widehat{\theta}^{\mathcal{Q}}_{ML}) \rightsquigarrow \chi^2_{D_\Theta}$$

Proof. Under the constrained model $\mathcal{Q}$, the estimator $\widehat{\theta}^{\mathcal{Q}}_{\mathrm{ML}}$ is asymptotic efficient with asymptotic variance $n^{-1}\mathbf{I}_{\mathcal{Q}}(\theta_0)^{-1}$, while the estimator $\widehat{\theta}^{\mathcal{P}}_{\mathrm{ML}}$ is asymptotic normal but not asymptotic efficient. The estimator $\widehat{\theta}^{\mathcal{Q}}_{\mathrm{ML}}$ becomes inconsistent when $\mathcal{Q}$ is false, while the estimator $\widehat{\theta}^{\mathcal{P}}_{\mathrm{ML}}$ is always consistent since $\mathcal{P}$ is assumed always to be true. Thus, the statistic $n(\widehat{\theta}^{\mathcal{P}}_{\mathrm{ML}} - \widehat{\theta}^{\mathcal{Q}}_{\mathrm{ML}})^T \left[\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1} - \mathbf{I}_{\mathcal{Q}}(\theta_0)^{-1}\right]^{-1} (\widehat{\theta}^{\mathcal{P}}_{\mathrm{ML}} - \widehat{\theta}^{\mathcal{Q}}_{\mathrm{ML}})$ is effectively the Hausman specification test statistic based on MLE of limited-information likelihoods for subset of moments. In fact, the result directly follows from Theorem 3 of Newey (1985b) with D_Θ to be the rank of $\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1} - \mathbf{I}_{\mathcal{Q}}(\theta_0)^{-1}$. □

Proposition 11. *Denote* $\mathbf{v} \equiv \nabla f(\theta_0)$*. Under the assumptions in Subsection 1.4, it holds that the Hausman test statistic based on MLEs satisfies*

$$n(f(\widehat{\theta}^{\mathcal{P}}_{ML}) - f(\widehat{\theta}^{\mathcal{Q}}_{ML}))^T \left[\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v} - \mathbf{v}^T\mathbf{I}_{\mathcal{Q}}(\theta_0)^{-1}\mathbf{v}\right]^{-1} (f(\widehat{\theta}^{\mathcal{P}}_{ML}) - f(\widehat{\theta}^{\mathcal{Q}}_{ML})) \rightsquigarrow \chi^2_{D_f} \tag{107}$$

Proof. By using the Delta method and Proposition 10, it follows that $\sqrt{n}(f(\widehat{\theta}^{\mathcal{P}}_{\mathrm{ML}}) - f(\widehat{\theta}^{\mathcal{Q}}_{\mathrm{ML}}))$ has asymptotic normal distribution with the asymptotic covariance matrix $\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v} - \mathbf{v}^T\mathbf{I}_{\mathcal{Q}}(\theta_0)^{-1}\mathbf{v}$. According to continuous mapping theorem, we know that the result holds. □

Proposition 12. *Suppose that the assumptions in Subsection 1.4 are satisfied. Define the sets*

$$\mathcal{J}_{1,n}(\delta,\eta) \equiv \left\{ \left|\left|\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathcal{P}}(\theta)\right|\right|_{\mathcal{S}} \le \eta \left|\left|\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\right|\right|_{\mathcal{S}}^{-1}, \quad \forall\ \theta \in \Omega(\theta_0,\delta) \text{ and } \tilde{\theta} \in \Omega(\theta,\delta) \right\},$$

and

$$\mathcal{J}_{2,n}(\theta,\delta,\eta) \equiv \left\{ \left|\left|\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathcal{P}}(\theta)\right|\right|_{\mathcal{S}} \le \eta \left|\left|\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\right|\right|_{\mathcal{S}}^{-1}, \quad \forall\ \tilde{\theta} \in \Omega(\theta,\delta) \right\}.$$

Then, for any $\eta > 0$ *there exists small enough positive constants* δ_1 *and* δ *such that*

$$\mathbb{P}_n\mathcal{J}_{1,n}(\delta,\eta)^{\complement} = o\left(\frac{1}{n}\right) \quad \text{and} \quad \sup_{\theta\in\Omega(\theta_0,\delta_1)} \mathbb{P}_{\theta,n}\mathcal{J}_{2,n}(\theta,\delta,\eta)^{\complement} = o\left(\frac{1}{n}\right).$$

Proof. Appealing to the fact that the Spectral norm and the Frobenius norm are equivalent for the $D_\Theta \times D_\Theta$ matrixes and following the argument on page 49 of Clarke and Barron (1994) or page

465 of Clarke and Barron (1990), we can prove the results very similarly. We omit the detailed proofs to avoid tedious repetition of the proofs in Clarke and Barron (1990) and Clarke and Barron (1994). ☐

Corollary 5. *Suppose that the assumptions in Subsection 1.4 are satisfied. Define the sets*

$$\mathcal{J}_{3,n}(\delta,\eta) \equiv \left\{1-\eta \leq \left|\left|\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\right|\right|_{\mathcal{S}} \leq 1+\eta, \quad \forall\ \theta \in \Omega(\theta_0,\delta) \text{ and } \tilde{\theta} \in \Omega(\theta,\delta)\right\},$$

and

$$\mathcal{J}_{4,n}(\theta,\delta,\eta) \equiv \left\{1-\eta \leq \left|\left|\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\right|\right|_{\mathcal{S}} \leq 1+\eta, \quad \forall\ \tilde{\theta} \in \Omega(\theta,\delta)\right\}.$$

Then, for any $\eta > 0$ there exists small enough positive constants δ_1 and δ such that

$$\mathbb{P}_n \mathcal{J}_{3,n}(\delta,\eta)^{\complement} = o\left(\frac{1}{n}\right), \quad \text{and} \quad \sup_{\theta\in\Omega(\theta_0,\delta_1)} \mathbb{P}_{\theta,n}\mathcal{J}_{4,n}(\theta,\delta,\eta)^{\complement} = o\left(\frac{1}{n}\right).$$

Proof. We have

$$\begin{aligned}
\left|\left|\left|\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\right|\right|_{\mathcal{S}} - 1\right| &\leq \left|\left|\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2} - I\right|\right|_{\mathcal{S}} \\
&= \left|\left|\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\left[\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathcal{P}}(\theta)\right]\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\right|\right|_{\mathcal{S}} \\
&\leq \left|\left|\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathcal{P}}(\theta)\right|\right|_{\mathcal{S}} \left|\left|\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\right|\right|_{\mathcal{S}} \qquad (108)
\end{aligned}$$

The first inequality is due to the triangular inequality for spectral norm. The second inequality is because for each unit vector $\mathbf{v}$ *in* $\mathbb{R}^d$,

$$\begin{aligned}
\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\left[\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathcal{P}}(\theta)\right]\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\mathbf{v} &\leq \lambda_M\left(\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathcal{P}}(\theta)\right)\left|\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\right|^2 \\
&= \left|\left|\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathcal{P}}(\theta)\right|\right|_{\mathcal{S}} \mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v} \\
&\leq \left|\left|\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathcal{P}}(\theta)\right|\right|_{\mathcal{S}} \left|\left|\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\right|\right|_{\mathcal{S}}.
\end{aligned}$$

Therefore, the results of this corollary follow directly from the inequality (108) and the results of Proposition 12. ☐

Proposition 13. *Under the assumptions in Subsection 1.4, for any $\eta > 0$ there exists $\delta > 0$ such that*

$$\mathbb{P}_n\left\{\sup_{\theta\in\Omega(\theta_0,\delta)} s_{\mathcal{P},n}(\theta)^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}s_{\mathcal{P},n}(\theta) < \eta\right\}^{\complement} = o(1).$$

Proof. Due to the continuity, we know that there exists $\delta_1 > 0$ such that for all $\theta \in \Omega(\theta_0,\delta_1)$,

$$\frac{1}{2} < \lambda_m\left(\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\mathbf{I}_{\mathcal{P}}(\theta_0)\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\right) \leq \lambda_M\left(\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\mathbf{I}_{\mathcal{P}}(\theta_0)\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\right) < 2.$$

It follows that for all $\theta \in \Omega(\theta_0, \delta_1)$

$$\begin{aligned} s_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} s_{\mathcal{P},n}(\theta) &\le 2 s_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1} s_{\mathcal{P},n}(\theta) \\ &\le 4 s_{\mathcal{P},n}(\theta_0)^T \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1} s_{\mathcal{P},n}(\theta_0) \\ &\quad + 4 \left[s_{\mathcal{P},n}(\theta) - s_{\mathcal{P},n}(\theta_0)\right]^T \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1} \left[s_{\mathcal{P},n}(\theta) - s_{\mathcal{P},n}(\theta_0)\right]. \end{aligned}$$

By Taylor's expansion of the score function $s_{\mathcal{P},n}(\theta)$ around θ_0, we know that there exists $\tilde{\theta}$ between θ_0 and θ such that

$$\begin{aligned} &\left[s_{\mathcal{P},n}(\theta) - s_{\mathcal{P},n}(\theta_0)\right]^T \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1} \left[s_{\mathcal{P},n}(\theta) - s_{\mathcal{P},n}(\theta_0)\right] \\ &= (\theta - \theta_0)^T \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}) \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1} \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})(\theta - \theta_0) \\ &\le \underline{\lambda}^{-1} (\theta - \theta_0)^T \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})^2 (\theta - \theta_0). \end{aligned} \tag{109}$$

where the inequality above is due to the fact that $\underline{\lambda} \le \underline{\lambda}(\theta_0)$. According to Proposition 12, there exists $\delta_2 \in (0, \delta_1)$ such that

$$\mathbb{P}_n \mathcal{J}_{1,n}(\delta_2, 1)^{\complement} = o\left(\frac{1}{n}\right),$$

where

$$\mathcal{J}_{1,n}(\delta_2, 1) \equiv \left\{ \left|\left|\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathcal{P}}(\theta)\right|\right|_{\mathcal{S}} \le \left|\left|\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\right|\right|_{\mathcal{S}}^{-1}, \ \ \forall\, \theta \in \Omega(\theta_0, \delta_2) \text{ and } \tilde{\theta} \in \Omega(\theta, \delta_2) \right\}.$$

Therefore, we only need to focus on the big probability set $\mathcal{J}_{1,n}(\delta_2, 1)$. Thus, by the triangular inequality, for any $\theta \in \Omega(\theta_0, \delta_2)$, we know that

$$\left|\left|\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})\right|\right|_{\mathcal{S}} \le ||\mathbf{I}_{\mathcal{P}}(\theta)||_{\mathcal{S}} + \left|\left|\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\right|\right|_{\mathcal{S}}^{-1} \le \overline{\lambda} + \underline{\lambda}.$$

Then, following the inequality (109), if we restrict on the big probability set $\mathcal{J}_{1,n}(\delta_2, 1)$, it follows that

$$\left[s_{\mathcal{P},n}(\theta) - s_{\mathcal{P},n}(\theta_0)\right]^T \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1} \left[s_{\mathcal{P},n}(\theta) - s_{\mathcal{P},n}(\theta_0)\right] \le \underline{\lambda}^{-1} (\overline{\lambda} + \underline{\lambda})^2 |\theta - \theta_0|^2.$$

Therefore, we choose

$$\delta = \min\left\{\delta_1, \delta_2, \sqrt{\frac{\eta}{8\underline{\lambda}^{-1}(\overline{\lambda} + \underline{\lambda})^2}}\right\},$$

and when $\theta \in \Omega(\theta_0, \delta)$ and $\mathbf{x^n} \in \mathcal{J}_{1,n}(\delta, 1)$,

$$s_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} s_{\mathcal{P},n}(\theta) \le 4 s_{\mathcal{P},n}(\theta_0)^T \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1} s_{\mathcal{P},n}(\theta_0) + \frac{\eta}{2}.$$

By Markov's inequality, it is straightforward to see that

$$s_{\mathcal{P},n}(\theta_0)^T \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1} s_{\mathcal{P},n}(\theta_0) \to 0 \ \text{ in } \ \mathbb{P}_n.$$

Therefore, we have shown that

$$\mathbb{P}_n \left\{ \sup_{\theta \in \Omega(\theta_0, \delta)} s_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} s_{\mathcal{P},n}(\theta) < \eta \right\}^{\complement} = o(1).$$

□

1.8 Properties of Posteriors Based on Limited-Information Likelihoods

Proposition 14. *Let's define*

$$\mathcal{S}_n(\delta, \eta) \equiv \left\{ \left| \int_{\Omega(\theta_0, \delta)} \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) S_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} S_{\mathcal{P},n}(\theta) d\theta - D_{\Theta} \right| < \eta \right\}.$$

Suppose that the assumptions in Subsection 1.4 hold. For any $\eta > 0$, there exists $\delta > 0$ such that

$$\mathbb{P}_n \mathcal{S}_n(\delta, \eta)^{\complement} = o(1).$$

Proof. We first show that for any $\eta > 0$

$$\mathbb{P}_n \left\{ \int_{\Omega(\theta_0, \delta)} \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) S_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} S_{\mathcal{P},n}(\theta) \mathrm{d}\theta > D_{\Theta} + \eta \right\} = o(1). \tag{110}$$

According to Corollary 5, we know that there exists $\delta_1 > 0$ such that

$$\mathbb{P}_n \mathcal{I}_{3,n} \left(\delta_1, \frac{\eta}{2D_{\Theta}} \right)^{\complement} = o\left(\frac{1}{n}\right),$$

where

$$\mathcal{I}_{3,n} \left(\delta_1, \frac{\eta}{2D_{\Theta}} \right) \equiv \left\{ \left(1 - \frac{\eta}{2D_{\Theta}}\right)^{1/2} \leq \left|\left| \mathbf{I}_{\mathcal{P}}(\theta)^{-1/2} \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}) \mathbf{I}_{\mathcal{P}}(\theta)^{-1/2} \right|\right|_{\mathcal{S}} \leq \left(1 + \frac{\eta}{2D_{\Theta}}\right)^{1/2}, \ \ \forall\, \theta \in \Omega(\theta_0, \delta_1) \text{ and } \tilde{\theta} \in \Omega(\theta, \delta_1) \right\}.$$

According to the consistency of the estimators, the set $\mathcal{A}_n(\delta_1) \equiv \left\{ \widehat{\theta}^{\mathcal{P}}_{\mathrm{ML}} \in \Omega(\theta_0, \delta_1) \right\}$ has probability going to 1. Thus, on the big probability event $\mathcal{I}_{3,n}\left(\delta_1, \frac{\eta}{2D_{\Theta}}\right) \cap \mathcal{A}_n(\delta_1)$, by Taylor's expansion, we have

$$\begin{aligned}
S_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} S_{\mathcal{P},n}(\theta) &= n(\theta - \widehat{\theta}^{\mathcal{P}}_{\mathrm{ML}})^T \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}) \mathbf{I}_{\mathcal{P}}(\theta)^{-1} \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}) (\theta - \widehat{\theta}^{\mathcal{P}}_{\mathrm{ML}}) \\
&= n(\theta - \widehat{\theta}^{\mathcal{P}}_{\mathrm{ML}})^T \mathbf{I}_{\mathcal{P}}(\theta)^{1/2} \left[\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2} \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}) \mathbf{I}_{\mathcal{P}}(\theta)^{-1/2} \right]^2 \mathbf{I}_{\mathcal{P}}(\theta)^{1/2} (\theta - \widehat{\theta}^{\mathcal{P}}_{\mathrm{ML}}) \\
&\leq n \left(1 + \frac{\eta}{2D_{\Theta}}\right)^{1/2} (\theta - \widehat{\theta}^{\mathcal{P}}_{\mathrm{ML}})^T \mathbf{I}_{\mathcal{P}}(\theta) (\theta - \widehat{\theta}^{\mathcal{P}}_{\mathrm{ML}})
\end{aligned}$$

where $\tilde{\theta}$ is between θ and $\widehat{\theta}^{\mathcal{P}}_{\mathrm{ML}}$. By the continuity, we know that there exists $\delta_2 > 0$ such that for all

$\theta \in \Omega(\theta_0, \delta_2)$

$$\left|\left|\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1/2}\mathbf{I}_{\mathcal{P}}(\theta)\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1/2}\right|\right|_{\mathcal{S}} \leq \left(1 + \frac{\eta}{2D_\Theta}\right)^{1/2}.$$

Choose $\delta \equiv \min\{\delta_1, \delta_2\}$. Thus, when considering $\theta \in \Omega(\theta_0, \delta)$ and restricting on the event $\mathcal{I}_{3,n}\left(\delta, \frac{\eta}{2D_\Theta}\right) \cap \mathcal{A}_n(\delta)$, we have

$$S_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}^{-1}(\theta) S_{\mathcal{P},n}(\theta) \leq \left(1 + \frac{\eta}{2D_\Theta}\right) n(\theta - \widehat{\theta}_{\mathrm{ML}}^{\mathcal{P}})^T \mathbf{I}_{\mathcal{P}}(\theta_0)(\theta - \widehat{\theta}_{\mathrm{ML}}^{\mathcal{P}}).$$

Therefore, we have

$$\int_{\Omega(\theta_0,\delta)} \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) S_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}^{-1}(\theta) S_{\mathcal{P},n}(\theta) \mathrm{d}\theta \leq \left(1 + \frac{\eta}{2D_\Theta}\right) \int_{\Omega(\theta_0,\delta)} \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) n(\theta - \widehat{\theta}_{\mathrm{ML}}^{\mathcal{P}})^T \mathbf{I}_{\mathcal{P}}(\theta_0)(\theta - \widehat{\theta}_{\mathrm{ML}}^{\mathcal{P}}) \mathrm{d}\theta.$$

According to Theorem 1 and Proposition 3 of Chernozhukov and Hong (2003), we know that

$$\left(1 + \frac{\eta}{2D_\Theta}\right) \int_{\Omega(\theta_0,\delta)} \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) n(\theta - \widehat{\theta}_{\mathrm{ML}}^{\mathcal{P}})^T \mathbf{I}_{\mathcal{P}}(\theta_0)(\theta - \widehat{\theta}_{\mathrm{ML}}^{\mathcal{P}}) \mathrm{d}\theta \to D_\Theta + \frac{\eta}{2} \text{ in } \mathbb{P}_n.$$

Therefore, the limit result in (110) holds. The proof of the following limit result is quite similar

$$\mathbb{P}_n \left\{ \int_{\Omega(\theta_0,\delta)} \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) S_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} S_{\mathcal{P},n}(\theta) \mathrm{d}\theta > D_\Theta - \eta \right\} = o(1), \quad \forall \ \eta > 0. \tag{111}$$

So, we ignore the detailed proof. ☐

Proposition 15. *Let's define*

$$\mathcal{S}_{\mathbf{v},n}(\delta, \eta) \equiv \left\{ \left| \int_{\Omega(\theta_0,\delta)} \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \frac{\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} S_{\mathcal{P},n}(\theta) S_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta) \mathbf{v}}{\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} \mathbf{v}} d\theta - 1 \right| < \eta \right\}.$$

Suppose that the assumptions in Subsection 1.4 hold. For any $\eta > 0$, there exists $\delta > 0$ such that

$$\mathbb{P}_n \mathcal{S}_{\mathbf{v},n}(\delta, \eta)^{\complement} = o(1).$$

Proof. The proof is similar to that of Proposition 14. ☐

Proposition 16. *Under the assumptions in Subsection 1.4, for any open subset $\mathcal{N} \subset \Theta$ and open set $\mathcal{N}_0 \subset \mathcal{N}$ such that $d_L(\mathcal{N}^{\complement}, \mathcal{N}_0) > \delta$ for some $\delta > 0$, there exist positive constants C, ξ_1 and ξ_2 such that*

$$\sup_{\theta \in \mathcal{N}_0} \mathbb{P}_{\theta,n} \left\{ \pi_{\mathcal{P}}(\mathbf{x^n}|\theta) \leq e^{\xi_1 n} \int_{\mathcal{N}^{\complement}} \pi_{\mathcal{P}}(\vartheta) \pi_{\mathcal{P}}(\mathbf{x^n}|\vartheta) \, d\vartheta \right\} \leq C e^{-\xi_2 n}.$$

Proof. According to Lemma 4 and Lemma 5 at the end of Section 1.6, the assumptions in Subsection 1.4 guarantee the existence of **strongly uniformly exponentially consistent (SUEC)** hypothesis tests.

In particular, for any open subset $\mathcal{N} \subset \Theta$ and $\mathcal{N}_0 \subset \mathcal{N}$ such that $d_L(\mathcal{N}, \mathcal{N}_0) > \delta$ for some $\delta > 0$, for each $\theta \in \mathcal{N}_0$, there exists a sequence of tests with acceptance region $\mathcal{A}_{\theta,n}$ for null hypothesis $\theta' = \theta$ versus $\theta' \in \mathcal{N}^{\complement}$ such that

$$\sup_{\theta \in \mathcal{N}_0} \mathbb{P}_{\theta,n} \mathcal{A}_{\theta,n}^{\complement} \leq C e^{-\xi n} \quad \text{and} \quad \sup_{\theta \in \mathcal{N}_0} \sup_{\theta' \in \mathcal{N}^{\complement}} \mathbb{P}_{\theta',n} \mathcal{A}_{\theta,n} < C e^{-\xi n}, \quad \text{for some } \xi > 0.$$

Denote the mixture distribution of $\mathbb{P}_{\theta,n}$ with respect to the conditional prior distribution $\pi_{\mathcal{P}}(\cdot|\mathcal{N}^{\complement})$ by $\mathcal{P}_{\mathcal{N}^{\complement},n}$ with density $\pi_{\mathcal{P}}(\mathbf{x^n}|\mathcal{N}^{\complement})$. More precisely, we define

$$\pi_{\mathcal{P}}(\mathbf{x^n}|\mathcal{N}^{\complement}) \equiv \int_{\mathcal{N}^{\complement}} \pi_{\mathcal{P}}(\mathbf{x^n}|\theta) \pi_{\mathcal{P}}(\theta|\mathcal{N}^{\complement}) \mathrm{d}\theta. \tag{112}$$

Following Lemma 6, we can show that there exist a real number $r > 0$ such that

$$\left|\left|\mathbb{P}_{\theta,n} - \mathcal{P}_{\mathcal{N}^{\complement},n}\right|\right|_{TV} \geq 2(1 - 2e^{-rn}), \quad \forall \ \ \theta \in \mathcal{N}_0.$$

For any positive sequence ϵ_n, by Markov's inequality, it follows that for each $\theta \in \mathcal{N}_0$

$$\begin{aligned}
&\mathbb{P}_{\theta,n} \left\{ \frac{\pi_{\mathcal{P}}(\mathbf{x^n}|\mathcal{N}^{\complement})}{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)} > \epsilon_n \right\} \leq \frac{1}{\epsilon_n^{1/2}} \int \pi_{\mathcal{P}}(\mathbf{x^n}|\theta)^{1/2} \pi_{\mathcal{P}}(\mathbf{x^n}|\mathcal{N}^{\complement})^{1/2} \mathrm{d}\mathbf{x^n} = \frac{1}{\epsilon_n^{1/2}} \alpha_H(\mathbb{P}_{\theta,n}, \mathbb{P}_{\mathcal{N}^{\complement},n}) \\
&\leq \frac{1}{\epsilon_n^{1/2}} \sqrt{1 - \left(\frac{1}{2}\left|\left|\mathbb{P}_{\theta,n} - \mathbb{P}_{\mathcal{N}^{\complement},n}\right|\right|_{TV}\right)^2} \leq \frac{1}{\epsilon_n^{1/2}} \sqrt{1 - (1 - 2e^{-rn})^2} = \frac{2e^{-\frac{rn}{2}}}{\epsilon_n^{1/2}} \sqrt{1 - e^{-rn}} \leq \frac{2e^{-\frac{rn}{2}}}{\epsilon_n^{1/2}}.
\end{aligned}$$

If we choose $\epsilon_n = e^{-\frac{rn}{4}}$, then we have

$$\sup_{\theta \in \mathcal{N}_0} \mathbb{P}_{\theta,n} \left\{ \frac{\pi_{\mathcal{P}}(\mathbf{x^n}|\mathcal{N}^{\complement})}{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)} > e^{-\frac{rn}{4}} \right\} \leq 2e^{-\frac{rn}{4}}.$$

□

Proposition 17. *Under the assumptions in Subsection 1.4, for any open subset $\mathcal{N} \subset \Theta$ and open set $\mathcal{N}_0 \subset \mathcal{N}$ such that $d_L(\mathcal{N}^{\complement}, \mathcal{N}_0) > \delta$ for some $\delta > 0$, then there exist positive constants C, ξ_1, ξ_2 such that*

$$\sup_{\theta \in \mathcal{N}_0} \mathbb{P}_{\theta,n} \left\{ \pi_{\mathcal{P}}(\mathbf{x^n}|\theta) \leq e^{\xi_1 n} \int_{\mathcal{N}_{-1}(\theta_{(1)})^{\complement}} \pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)}) \pi_{\mathcal{P}}(\mathbf{x^n}|\theta_{(1)}, \theta'_{(-1)}) d\theta'_{(-1)} \right\} \leq C e^{-\xi_2 n}. \tag{113}$$

Proof. The proof is the same as that of Proposition 16. □

Proposition 18. *Under the assumptions in Subsection 1.4, for any open neighborhood $\mathcal{N} \subset \Theta$ of θ_0 there exist positive constants C and ξ such that*

$$\mathbb{Q}_n \left\{ \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_0) \leq e^{\xi n} \int_{\mathcal{N}^{\complement}} \pi_{\mathcal{P}}(\vartheta) \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\vartheta) d\vartheta \right\} = o(1). \tag{114}$$

Proof. For any open neighborhood $\mathcal{N}$ of θ_0, from the identification assumption (i.e. **Assumption ID**) and the compactness of Θ, it follows that there exists $\epsilon > 0$ such that $\min_{\theta\in\mathcal{N}^\complement} H_{\mathcal{Q}}(\theta) \geq \epsilon$ where $H_{\mathcal{Q}}(\theta)$ is defined in (101). Consider the large probability set

$$\mathcal{A}_n \equiv \left\{\sup_{\theta\in\Theta} |\widehat{H}_{\mathcal{Q},n}(\theta) - H_{\mathcal{Q}}(\theta)| < \epsilon/2\right\}.$$

From Proposition 7, we know that $\mathbb{Q}_n\mathcal{A}_n \to 1$ as $n \to \infty$. Thus, we only need to focus on event $\mathcal{A}_n$. Then, we have

$$\begin{aligned}
&\mathbb{Q}_n\mathcal{A}_n\left\{\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_0) \leq e^{n\epsilon/4}\int_{\mathcal{N}^\complement} \pi_{\mathcal{P}}(\theta')\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta')\mathrm{d}\theta'\right\}\\
&= \mathbb{Q}_n\mathcal{A}_n\left\{e^{-n\widehat{H}_{\mathcal{Q},n}(\theta_0)} \leq e^{n\epsilon/4}\int_{\mathcal{N}^\complement} \pi_{\mathcal{P}}(\theta')e^{-n\widehat{H}_{\mathcal{Q},n}(\theta)}\mathrm{d}\theta'\right\}\\
&\leq \mathbb{Q}_n\left\{e^{-n\widehat{H}_{\mathcal{Q},n}(\theta_0)} \leq e^{n\epsilon/4}\int_{\mathcal{N}^\complement} \pi_{\mathcal{P}}(\theta')e^{-n[H_{\mathcal{Q}}(\theta')-\epsilon/2]}\mathrm{d}\theta'\right\}\\
&\leq \mathbb{Q}_n\left\{e^{-n\widehat{H}_{\mathcal{Q},n}(\theta_0)} \leq e^{n\epsilon/4}\int_{\mathcal{N}^\complement} \pi_{\mathcal{P}}(\theta')e^{-n\epsilon/2}\mathrm{d}\theta'\right\}\\
&\leq \mathbb{Q}_n\left\{e^{-n\widehat{H}_{\mathcal{Q},n}(\theta_0)} \leq e^{-n\epsilon/4}\right\} \leq e^{-n\epsilon/16}\mathbb{E}e^{n\widehat{H}_{\mathcal{Q},n}(\theta_0)/4}
\end{aligned}$$

Because $n\widehat{H}_{\mathcal{Q},n}(\theta_0)$ converges to a chi-squire random variable with degree of freedom D_Θ in distribution, we know that $\mathbb{E}e^{n\widehat{H}_{\mathcal{Q},n}(\theta_0)/4} \to 2^{D_\Theta/2}$. Thus,

$$\mathbb{Q}_n\mathcal{A}_n\left\{\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_0) \leq e^{n\epsilon/4}\int_{\mathcal{N}^\complement} \pi_{\mathcal{P}}(\theta')\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta')\mathrm{d}\theta'\right\} \leq Ce^{-n\epsilon/16}$$

for some constant $C > 0$. Therefore, if we take $\xi = \epsilon/4$, the proof is completed. ☐

Proposition 19. *Under the assumptions in Subsection 1.4, for any open subsets $\mathcal{N} \subset \Theta$ and any positive constant ξ, there exists a neighborhood $\mathcal{N}_0$ of θ_0 such that*

$$\sup_{\theta\in\mathcal{N}_0} \mathbb{P}_{\theta,n}\left\{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta) \geq e^{\xi n}\int_{\mathcal{N}} \pi_{\mathcal{P}}(\vartheta)\pi_{\mathcal{P}}(\mathbf{x^n}|\vartheta)\,d\vartheta\right\} = o\left(\frac{1}{n}\right). \tag{115}$$

Proof. Let $r_n = 1/\sqrt{n}$ and it is sufficient to show that

$$\sup_{\theta\in\mathcal{N}_0} \mathbb{P}_{\theta,n}\left\{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta) \geq e^{\xi n}\int_{\Omega(\theta,r_n)} \pi_{\mathcal{P}}(\vartheta)\pi_{\mathcal{P}}(\mathbf{x^n}|\vartheta)\mathrm{d}\vartheta\right\} = o\left(\frac{1}{n}\right).$$

It is equivalent to show that

$$\sup_{\theta\in\mathcal{N}_0} \mathbb{P}_{\theta,n}\left\{\ln\frac{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)}{\pi_{\mathcal{P}}(\mathbf{x^n}|\Omega(\theta,r_n))} \geq \xi_n n\right\} = o\left(\frac{1}{n}\right),$$

where

$$\xi_n \equiv \xi - \frac{1}{n}\ln \pi_{\mathcal{P}}(\Omega(\theta, r_n))$$

with

$$\pi_{\mathcal{P}}(\mathbf{x^n}|\Omega(\theta, r_n)) \equiv \int_{\Omega(\theta,r_n)} \frac{\pi_{\mathcal{P}}(\vartheta)}{\pi_{\mathcal{P}}(\Omega(\theta, r_n))}\pi_{\mathcal{P}}(\mathbf{x^n}|\vartheta)\mathrm{d}\vartheta.$$

In fact, we have

$$\pi_{\mathcal{P}}(\Omega(\theta, r_n)) = \int_{\Omega(\theta,r_n)} \pi_{\mathcal{P}}(\vartheta)\mathrm{d}\vartheta \geq m_\pi \Gamma_{D_\Theta}\left(\frac{1}{n}\right)^{D_\Theta/2},$$

where Γ_{D_Θ} is the volume of the unit ball in $\mathbb{R}^{D_\Theta}$. Thus,

$$\xi_n = \xi - O(n^{-1}\ln n).$$

Therefore, for all large n, $\xi_n \geq \xi/2$ and hence it suffices to show that

$$\sup_{\theta\in\mathcal{N}_0} \mathbb{P}_{\theta,n}\left\{\ln \frac{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)}{\pi_{\mathcal{P}}(\mathbf{x^n}|\Omega(\theta, r_n))} \geq \xi n/2\right\} = o\left(\frac{1}{n}\right).$$

For each $\theta \in \Theta$, by Markov's inequality, we have

$$\mathbb{P}_{\theta,n}\left\{\ln \frac{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)}{\pi_{\mathcal{P}}(\mathbf{x^n}|\Omega(\theta, r_n))} \geq \xi n/2\right\} \leq \frac{4}{n^2\xi^2}\mathbb{E}_{\mathcal{P}_\theta}\left[\ln \frac{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)}{\pi_{\mathcal{P}}(\mathbf{x^n}|\Omega(\theta, r_n))}\right]^2 \tag{116}$$

We consider the set

$$\mathcal{I}_{4,n}(\theta, \delta, 1) \equiv \left\{\left|\left|\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\right|\right|_{\mathcal{S}} \leq 2, \quad \forall\, \tilde{\theta} \in \Omega(\theta, \delta)\right\}.$$

According to Corollary 5, it follows that there exist positive constants δ and δ_0 such that

$$\sup_{\theta\in\Omega(\theta_0,\delta_0)} \mathbb{P}_{\theta,n}\mathcal{I}_{4,n}(\theta, \delta, 1)^{\complement} = o\left(\frac{1}{n}\right).$$

Therefore, we only need to focus on the big probability set $\mathcal{I}_{4,n}(\theta, \delta, 1)$ for each $\theta \in \Omega(\theta_0, \delta_0)$.

We choose $\mathcal{N}_0 = \Omega(\theta_0, \delta_0)$. We have for each $\theta \in \mathcal{N}_0$ the following equality holds

$$\begin{aligned}\ln \frac{\pi_{\mathcal{P}}(\mathbf{x^n}|\Omega(\theta, r_n))}{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)} &= \ln \int_{\Omega(\theta,r_n)} \pi_{\mathcal{P}}(\vartheta|\Omega(\theta, r_n))\frac{\pi_{\mathcal{P}}(\mathbf{x^n}|\vartheta)}{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)}\mathrm{d}\vartheta \\ &= \ln \int_{\Omega(\theta,r_n)} \pi_{\mathcal{P}}(\vartheta|\Omega(\theta, r_n))e^{\sqrt{n}S_{\mathcal{P},n}(\theta)^T(\vartheta-\theta)-\frac{1}{2}n(\vartheta-\theta)^T\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})(\vartheta-\theta)}\mathrm{d}\vartheta,\end{aligned}$$

where $\tilde{\theta}$ is between θ and ϑ.

On the one hand, on the event $\mathcal{I}_{4,n}(\theta, \delta, 1)$, we have

$$\ln \frac{\pi_{\mathcal{P}}(\mathbf{x^n}|\Omega(\theta, r_n))}{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)} \leq \ln \int_{\Omega(\theta, r_n)} \pi_{\mathcal{P}}(\vartheta|\Omega(\theta, r_n)) e^{|S_{\mathcal{P},n}(\theta)| + ||\mathbf{I}_{\mathcal{P}}(\theta)||_{\mathcal{S}}} \mathrm{d}\vartheta = |S_{\mathcal{P},n}(\theta)| + ||\mathbf{I}_{\mathcal{P}}(\theta)||_{\mathcal{S}} .$$

Thus, on the other hand, on the event $\mathcal{I}_{4,n}(\theta, \delta, 1)$, we have by Jensen's inequality

$$\begin{aligned} \ln \frac{\pi_{\mathcal{P}}(\mathbf{x^n}|\Omega(\theta, r_n))}{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)} &\geq \int_{\Omega(\theta, r_n)} \left[\sqrt{n} S_{\mathcal{P},n}(\theta)^T (\vartheta - \theta) - \frac{1}{2} (\vartheta - \theta)^T \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})(\vartheta - \theta) \right] \pi_{\mathcal{P}}(\vartheta|\Omega(\theta, r_n)) \mathrm{d}\vartheta \\ &\geq \int_{\Omega(\theta, r_n)} \left[\sqrt{n} S_{\mathcal{P},n}(\theta)^T (\vartheta - \theta) - (\vartheta - \theta)^T \mathbf{I}_{\mathcal{P}}(\theta)(\vartheta - \theta) \right] \pi_{\mathcal{P}}(\vartheta|\Omega(\theta, r_n)) \mathrm{d}\vartheta \\ &\geq -|S_{\mathcal{P},n}(\theta)| - ||\mathbf{I}_{\mathcal{P}}(\theta)||_{\mathcal{S}} \end{aligned}$$

Therefore, we have

$$\left[\ln \frac{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)}{\pi_{\mathcal{P}}(\mathbf{x^n}|\Omega(\theta, r_n))} \right]^2 \leq \left[|S_{\mathcal{P},n}(\theta)| + ||\mathbf{I}_{\mathcal{P}}(\theta)||_{\mathcal{S}} \right]^2 \leq 2|S_{\mathcal{P},n}(\theta)|^2 + 2\, ||\mathbf{I}_{\mathcal{P}}(\theta)||_{\mathcal{S}}^2 . \tag{117}$$

Combining (116) and (117), we know that

$$\begin{aligned} \mathbb{P}_{\theta,n} \left\{ \ln \frac{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)}{\pi_{\mathcal{P}}(\mathbf{x^n}|\Omega(\theta, r_n))} \geq \xi n/2 \right\} &\leq \frac{8}{n^2 \xi^2} \left[\mathbb{E}_{\mathcal{P}_\theta} |S_{\mathcal{P},n}(\theta)|^2 + \mathbb{E}_{\mathcal{P}_\theta} ||\mathbf{I}_{\mathcal{P}}(\theta)||_{\mathcal{S}}^2 \right] \\ &\leq \frac{8}{n^2 \xi^2} \left[\mathrm{tr}\left(\mathbf{I}_{\mathcal{P}}(\theta) \right) + \lambda_M \left(\mathbf{I}_{\mathcal{P}}(\theta) \right)^2 \right] \leq \frac{8(D_\Theta \overline{\lambda} + \overline{\lambda}^2)}{n^2 \xi^2}. \end{aligned}$$

□

Proposition 20. *Under the assumptions in Subsection 1.4, for any open subsets $\mathcal{N} \subset \Theta$ and any positive constant ξ, we have*

$$\mathbb{Q}_n \left\{ \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma) \geq e^{\xi n} \int_{\mathcal{N}} \pi_{\mathcal{P}}(\gamma') \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma') d\gamma' \right\} = o\left(\frac{1}{n} \right), \quad \textit{for every } \gamma \in \mathcal{N}.$$

Proof. Let $r_n = 1/\sqrt{n}$ and it is sufficient to show that

$$\mathbb{Q}_n \left\{ \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma) \geq e^{\xi n} \int_{\Omega(\gamma, r_n)} \pi_{\mathcal{P}}(\gamma') \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma') \mathrm{d}\gamma' \right\} = o\left(\frac{1}{n} \right).$$

It is equivalent to show that

$$\mathbb{Q}_n \left\{ \ln \frac{\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma)}{\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\Omega(\gamma, r_n))} \geq \xi_n n \right\} = o\left(\frac{1}{n} \right),$$

where

$$\xi_n \equiv \xi - \frac{1}{n} \ln \pi_{\mathcal{P}}(\Omega(\gamma, r_n))$$

with

$$\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\Omega(\gamma, r_n)) \equiv \int_{\Omega(\gamma,r_n)} \frac{\pi_{\mathcal{P}}(\gamma')}{\pi_{\mathcal{P}}(\Omega(\gamma, r_n))} \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma')\mathrm{d}\gamma'.$$

Thus, we have

$$\pi_{\mathcal{P}}(\Omega(\gamma, r_n)) = \int_{\Omega(\gamma,r_n)} \pi_{\mathcal{P}}(\gamma')\mathrm{d}\gamma' \geq m_\pi B_D \left(\frac{1}{n}\right)^{D/2},$$

where B_D is the volume of the unit ball in $\mathbb{R}^D$. Thus,

$$\xi_n = \xi - O(n^{-1}\ln n).$$

Therefore, for all large n, $\xi_n \geq \xi/2$ and hence it suffices to show that

$$\mathbb{Q}_n \left\{ \ln \frac{\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma)}{\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\Omega(\gamma, r_n))} \geq \xi n/2 \right\} = o\left(\frac{1}{n}\right).$$

By Markov's inequality, we have

$$\mathbb{Q}_n \left\{ \ln \frac{\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma)}{\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\Omega(\gamma, r_n))} \geq \xi n/2 \right\} \leq \frac{4}{n^2\xi^2} \mathbb{E}\left[\ln \frac{\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma)}{\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\Omega(\gamma, r_n))} \right]^2 \tag{118}$$

We consider the set

$$\mathcal{I}_n \equiv \left\{ \left\| \mathbf{I}_{\mathcal{Q}}(\gamma_0)^{-1/2} \widehat{\mathbf{I}}_{\mathcal{Q},n}(\tilde{\gamma}) \mathbf{I}_{\mathcal{Q}}(\gamma_0)^{-1/2} \right\|_{\mathcal{S}} \leq 2, \quad \forall\, \tilde{\gamma} \in \Omega(\gamma_0, r_n) \right\}.$$

According to Proposition 8, it follows that $\mathbb{Q}_n \mathcal{I}_n^{\complement} = o(1)$. Therefore, we only need to focus on the big probability set $\mathcal{I}_n$. It holds that

$$\begin{aligned}
\ln \frac{\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\Omega(\gamma_0, r_n))}{\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma_0)} &= \ln \int_{\Omega(\gamma_0,r_n)} \pi_{\mathcal{P}}(\gamma|\Omega(\gamma_0, r_n)) \frac{\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma)}{\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma_0)} \mathrm{d}\gamma \\
&= \ln \int_{\Omega(\gamma_0,r_n)} \pi_{\mathcal{P}}(\gamma|\Omega(\gamma_0, r_n)) e^{n\widehat{H}_{\mathcal{Q},n}(\gamma_0) - n\widehat{H}_{\mathcal{Q},n}(\gamma)} \mathrm{d}\gamma.
\end{aligned} \tag{119}$$

By a second-order Taylor expansion around γ_0 and Cauchy-Schwarz inequality, it holds that

$$\begin{aligned}
\left| n\widehat{H}_{\mathcal{Q},n}(\gamma_0) - n\widehat{H}_{\mathcal{Q},n}(\gamma) \right| &\leq \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \nabla \ln \pi_{\mathcal{Q}}(\gamma_0; \mathbf{z_t}) \right| + \frac{1}{2}(\gamma - \gamma_0)^T \left[-\frac{1}{n} \sum_{t=1}^{n} \nabla^2 \ln \pi_{\mathcal{Q}}(\tilde{\gamma}; \mathbf{z_t}) \right] (\gamma - \gamma_0) \\
&= \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \nabla \ln \pi_{\mathcal{Q}}(\gamma_0; \mathbf{z_t}) \right| + \frac{n}{2}(\gamma - \gamma_0)^T \widehat{\mathbf{I}}_{\mathcal{Q},n}(\tilde{\gamma})(\gamma - \gamma_0)
\end{aligned}$$

where $\tilde{\gamma}$ is between γ_0 and γ. Thus, on the event $\mathcal{I}_n$ and when $\gamma \in \Omega(\gamma_0, r_n)$, we have

$$\left| n\widehat{H}_{\mathcal{Q},n}(\gamma_0) - n\widehat{H}_{\mathcal{Q},n}(\gamma) \right| \leq ||\mathbf{I}_{\mathcal{Q}}(\gamma_0)||_{\mathcal{S}} + \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \nabla \ln \pi_{\mathcal{Q}}(\gamma_0; \mathbf{z_t}) \right|.$$

Combining (119), we know that on the event $\mathcal{I}_n$, it holds that

$$\left[\frac{\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n} | \Omega(\gamma_0, r_n))}{\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n} | \gamma_0)} \right]^2 \leq 2\, ||\mathbf{I}_{\mathcal{Q}}(\gamma_0)||_{\mathcal{S}}^2 + 2 \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \nabla \ln \pi_{\mathcal{Q}}(\gamma_0; \mathbf{z_t}) \right|^2.$$

In fact, it easy to see that

$$\mathbb{E} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \nabla \ln \pi_{\mathcal{Q}}(\gamma_0; \mathbf{z_t}) \right|^2 \to \operatorname{tr}\left[\mathbf{I}_{\mathcal{Q}}(\gamma_0)^{-1} \right], \quad \text{as } n \to \infty.$$

The trace $\operatorname{tr}\left[\mathbf{I}_{\mathcal{Q}}(\gamma_0)^{-1}\right]$ is upper bounded by $\lambda_m\left(\mathbf{I}_{\mathcal{Q}}(\gamma_0)\right)^{-1}$. Then, combining (118) and (119), we know that

$$\limsup_{n\to\infty} \mathbb{Q}_n \left\{ \ln \frac{\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n} | \gamma)}{\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n} | \Omega(\gamma, r_n))} \geq \xi n/2 \right\} \leq \frac{8}{n^2 \xi^2} \left\{ ||\mathbf{I}_{\mathcal{Q}}(\gamma_0)||_{\mathcal{S}}^2 + \lambda_m\left(\mathbf{I}_{\mathcal{Q}}(\gamma_0)\right)^{-1} \right\}.$$

□

Proposition 21. *Assume the regularity conditions in Subsection 1.4 hold. For any open neighborhood $\mathcal{N}$ of θ_0, there is an $\xi > 0$ and a open neighborhood $\mathcal{N}_0$ of θ_0 such that*

$$\sup_{\theta \in \mathcal{N}_0} \mathbb{P}_{\theta,n} \left\{ \int_{\mathcal{N}} \pi_{\mathcal{P}}(\theta') \pi_{\mathcal{P}}(\mathbf{x^n} | \theta')\, d\theta' \geq e^{\xi n} \int_{\mathcal{N}^{\complement}} \pi_{\mathcal{P}}(\theta') \pi_{\mathcal{P}}(\mathbf{x^n} | \theta')\, d\theta' \right\} = o(1).$$

Proof. According to Proposition 19, for any positive constant ξ_1, there exists a neighborhood $\mathcal{N}_0$ of θ_0 such that $d_L(\mathcal{N}^{\complement}, \mathcal{N}_0) > \delta$ for some $\delta > 0$ and

$$\sup_{\theta \in \mathcal{N}_0} \mathbb{P}_{\theta,n} \mathcal{A}_n(\theta, \xi_1) = o\left(\frac{1}{n}\right),$$

with

$$\mathcal{A}_n(\theta, \xi_1) \equiv \left\{ \pi_{\mathcal{P}}(\mathbf{x^n} | \theta) \geq e^{\xi_1 n} \int_{\mathcal{N}} \pi_{\mathcal{P}}(\vartheta) \pi_{\mathcal{P}}(\mathbf{x^n} | \vartheta) \mathrm{d}\vartheta \right\}.$$

By Proposition 16, there exist positive constants C, ξ' and ξ_2 such that

$$\sup_{\theta \in \mathcal{N}_0} \mathbb{P}_{\theta,n} \left\{ \pi_{\mathcal{P}}(\mathbf{x^n} | \theta) \leq e^{\xi' n} \int_{\mathcal{N}^{\complement}} \pi_{\mathcal{P}}(\vartheta) \pi_{\mathcal{P}}(\mathbf{x^n} | \vartheta) \mathrm{d}\vartheta \right\} \leq C e^{-\xi_2 n}.$$

We take a constant $0 < \xi < \xi'$ and let $\xi'' = \xi' - \xi$. Then,

$$
\begin{aligned}
&\sup_{\theta \in \mathcal{N}_0} \mathbb{P}_{\theta,n} \left\{ \int_{\mathcal{N}} \pi_{\mathcal{P}}(\theta') \pi_{\mathcal{P}}(\mathbf{x^n}|\theta') \mathrm{d}\theta' \leq e^{\xi n} \int_{\mathcal{N}^{\complement}} \pi_{\mathcal{P}}(\theta') \pi_{\mathcal{P}}(\mathbf{x^n}|\theta') \mathrm{d}\theta' \right\} \\
&\quad \leq \sup_{\theta \in \mathcal{N}_0} \mathbb{P}_{\theta,n} \mathcal{A}_n(\theta, \xi'') \left\{ \int_{\mathcal{N}} \pi_{\mathcal{P}}(\theta') \pi_{\mathcal{P}}(\mathbf{x^n}|\theta') \mathrm{d}\theta' \leq e^{\xi n} \int_{\mathcal{N}^{\complement}} \pi_{\mathcal{P}}(\theta') \pi_{\mathcal{P}}(\mathbf{x^n}|\theta') \mathrm{d}\theta' \right\} + \sup_{\theta \in \mathcal{N}_0} \mathbb{P}_{\theta,n} \mathcal{A}_n(\theta, \xi'')^{\complement} \\
&\quad \leq \sup_{\theta \in \mathcal{N}_0} \mathbb{P}_{\theta,n} \left\{ \pi_{\mathcal{P}}(\mathbf{x^n}|\theta) \leq e^{n(\xi - \xi'')} \int_{\mathcal{N}^{\complement}} \pi_{\mathcal{P}}(\theta') \pi_{\mathcal{P}}(\mathbf{x^n}|\theta') \mathrm{d}\theta' \right\} + o(1) \\
&\quad \leq \sup_{\theta \in \mathcal{N}_0} \mathbb{P}_{\theta,n} \left\{ \pi_{\mathcal{P}}(\mathbf{x^n}|\theta) \leq e^{n\xi'} \int_{\mathcal{N}^{\complement}} \pi_{\mathcal{P}}(\theta') \pi_{\mathcal{P}}(\mathbf{x^n}|\theta') \mathrm{d}\theta' \right\} + o(1) \\
&\quad \leq C e^{-\xi_2 n} + o(1) = o(1).
\end{aligned}
$$

□

Proposition 22. *Assume the regularity conditions in Subsection 1.4 hold. For any open neighborhood $\mathcal{N}$ of θ_0, there is an $\xi > 0$ such that*

$$
\mathbb{Q}_n \left\{ \int_{\mathcal{N}} \pi_{\mathcal{P}}(\gamma') \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma') d\gamma' \geq e^{\xi n} \int_{\mathcal{N}^{\complement}} \pi_{\mathcal{P}}(\gamma') \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma') d\gamma' \right\} = o(1).
$$

Proof. According to Proposition 20, for any positive constant ξ_1, there exists a neighborhood $\mathcal{N}_0$ of θ_0 such that $d_L(\mathcal{N}^{\complement}, \mathcal{N}_0) > \delta$ for some $\delta > 0$ and

$$
\mathbb{Q}_n \mathcal{A}_n(\xi_1) = o\left(\frac{1}{n}\right),
$$

with

$$
\mathcal{A}_n(\xi_1) \equiv \left\{ \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma_0) \geq e^{\xi_1 n} \int_{\mathcal{N}} \pi_{\mathcal{P}}(\vartheta) \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma') \mathrm{d}\gamma' \right\}.
$$

By Proposition 18, there exist positive constants C and ξ' such that

$$
\mathbb{Q}_n \left\{ \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma_0) \leq e^{\xi' n} \int_{\mathcal{N}^{\complement}} \pi_{\mathcal{P}}(\gamma') \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma') \mathrm{d}\gamma' \right\} = o(1).
$$

We take a constant $0 < \xi < \xi'$ and let $\xi'' = \xi' - \xi$. Then,

$$
\begin{aligned}
&\mathbb{Q}_n \left\{ \int_{\mathcal{N}} \pi_{\mathcal{P}}(\gamma') \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma') \mathrm{d}\gamma' \leq e^{\xi n} \int_{\mathcal{N}^{\complement}} \pi_{\mathcal{P}}(\gamma') \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma') \mathrm{d}\gamma' \right\} \\
&\quad \leq \mathbb{Q}_n \mathcal{A}_n(\xi'') \left\{ \int_{\mathcal{N}} \pi_{\mathcal{P}}(\gamma') \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma') \mathrm{d}\gamma' \leq e^{\xi n} \int_{\mathcal{N}^{\complement}} \pi_{\mathcal{P}}(\gamma') \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma') \mathrm{d}\gamma' \right\} + \mathbb{Q}_n \mathcal{A}_n(\xi'')^{\complement} \\
&\quad \leq \mathbb{Q}_n \left\{ \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma_0) \leq e^{n(\xi - \xi'')} \int_{\mathcal{N}^{\complement}} \pi_{\mathcal{P}}(\gamma') \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma') \mathrm{d}\gamma' \right\} + o(1) \\
&\quad \leq \mathbb{Q}_n \left\{ \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma_0) \leq e^{n\xi'} \int_{\mathcal{N}^{\complement}} \pi_{\mathcal{P}}(\gamma') \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\gamma') \mathrm{d}\gamma' \right\} + o(1) = o(1).
\end{aligned}
$$

□

1.9 Proof of Theorem 1

It follows immediately from the results of Theorem 2 and Theorem 3, which are proved in Section 1.10 and Section 1.11, respectively.

1.10 Proof of Theorem 2

Because of Assumption **FF** and the assumptions in Subsection 1.4 are invariant under invertible and second-order smooth transformations, without loss of generality, we assume that $f(\theta) = \theta_{(1)}$ and hence $\mathbf{v} = (1, 0, \cdots, 0)^T$. Let us denote

$$\varphi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n}) = \frac{1}{\sqrt{2\pi\frac{1}{n}\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}} \exp\left\{-\frac{1}{2\frac{1}{n}\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}(\theta_{(1)} - \widehat{\theta}^{\mathcal{P}}_{(1)})^2\right\}$$

$$\varphi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n}) = \frac{1}{\sqrt{2\pi\frac{1}{n}\mathbf{v}^T\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}}} \exp\left\{-\frac{1}{2\frac{1}{n}\mathbf{v}^T\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}}(\theta_{(1)} - \widehat{\theta}^{\mathcal{Q}}_{(1)})^2\right\}$$

where $\widehat{\theta}^{\mathcal{P}}$ and $\widehat{\theta}^{\mathcal{Q}}$ are MLE estimators for baseline model and full model, respectively, and $\widehat{\theta}^{\mathcal{P}}_{(1)}$ and $\widehat{\theta}^{\mathcal{Q}}_{(1)}$ are the first elements of $\widehat{\theta}^{\mathcal{P}}$ and $\widehat{\theta}^{\mathcal{Q}}$, respectively. Let's now focus on the decomposition of the relative entropy between constrained and unconstrained posterior distributions:

$$\mathbf{D}_{KL}\left(\pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n})||\pi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n})\right) = A_n + B_n + C_n \tag{120}$$

where

$$A_n = \int \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n}) \ln \frac{\pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n})}{\varphi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n})} \mathrm{d}\theta_{(1)}, \tag{121}$$

$$B_n = \int \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n}) \ln \frac{\varphi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n})}{\varphi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n})} \mathrm{d}\theta_{(1)}, \tag{122}$$

$$C_n = \int \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n}) \ln \frac{\varphi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n})}{\pi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n})} \mathrm{d}\theta_{(1)}. \tag{123}$$

We shall show that

$$A_n \to 0 \ \text{ in } \ \mathbb{Q}_n, \tag{124}$$

and

$$B_n - \frac{n}{2\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}\left(\widehat{\theta}^{\mathcal{P}}_{(1)} - \widehat{\theta}^{\mathcal{Q}}_{(1)}\right)^2 \to \frac{1}{2}\ln\frac{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}{\mathbf{v}^T\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}} + \frac{1}{2}\frac{\mathbf{v}^T\mathbf{I}_{\mathcal{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}} - \frac{1}{2} \ \text{ in } \mathbb{Q}_n, \tag{125}$$

and

$$C_n \to 0 \ \text{ in } \ \mathbb{Q}_n. \tag{126}$$

Step 1: we prove the weak convergence of A_n in (124). In fact, $A_n = \mathbf{D}_{KL}\left(\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})||\varphi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\right)$. And, according to Corollary 3, we know $A_n \to 0$ in $\mathbb{Q}_n$.

Step 2: we prove the weak convergence of B_n in (125). We know that

$$B_n = \frac{1}{2}\int \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\left[\ln\frac{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}{\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}} - n\frac{(\theta_{(1)}-\widehat{\theta}^{\mathbb{Q}}_{(1)})^2}{\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}} + n\frac{(\theta_{(1)}-\widehat{\theta}^{\mathcal{P}}_{(1)})^2}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}\right]\mathrm{d}\theta_{(1)} \quad (127)$$

Now let's define

$$\Delta_{B,n} \equiv B_n - \frac{n}{2\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}\left(\widehat{\theta}^{\mathcal{P}}_{(1)} - \widehat{\theta}^{\mathbb{Q}}_{(1)}\right)^2 - \frac{1}{2}\ln\frac{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}{\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}} - \frac{1}{2}\frac{\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}} + 1/2. \quad (128)$$

We consider the following decomposition for $\Delta_{B,n}$:

$$\Delta_{B,n} = \frac{1}{2}\int\left[\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) - \varphi_{\mathbb{Q}}\left(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}\right)\right]\mathrm{d}\theta_{(1)}\ln\frac{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}{\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}} \quad (129)$$

$$-\frac{1}{2}\int\left[\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) - \varphi_{\mathbb{Q}}\left(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}\right)\right] n\frac{(\theta_{(1)}-\widehat{\theta}^{\mathbb{Q}}_{(1)})^2}{\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}}\mathrm{d}\theta_{(1)} \quad (130)$$

$$+\frac{1}{2}\int\left[\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) - \varphi_{\mathbb{Q}}\left(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}\right)\right] n\frac{(\theta_{(1)}-\widehat{\theta}^{\mathcal{P}}_{(1)})^2}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}\mathrm{d}\theta_{(1)} \quad (131)$$

The term in (129) is denoted as $B_{n,1}$, the term in (130) is denoted as $B_{n,2}$, and the term (131) is denoted as $B_{n,3}$.

Step 2.1: We show that $B_{n,1} \to 0$ in $\mathbb{Q}_n$. We know that, in $\mathbb{Q}_n$,

$$|B_{n,1}| \le \frac{1}{2}\ln\frac{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}{\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}}\int\left|\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) - \varphi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\right|\mathrm{d}\theta_{(1)} \to 0 \quad (132)$$

where the convergence result in (132) is due to the fact that the squared total variation distance[5] is upper bounded by the relative entropy (see e.g. Kullback, 1967) and due to the result in Proposition 3.

Step 2.2: We show that $B_{n,2} \to 0$ in $\mathbb{Q}_n$. Equivalently, we show that

$$\int_{\Theta}\left[\pi_{\mathbb{Q}}(\theta|\mathbf{x^n},\mathbf{y^n}) - \varphi_{\mathbb{Q}}\left(\theta|\mathbf{x^n},\mathbf{y^n}\right)\right] n\frac{(\theta_{(1)}-\widehat{\theta}^{\mathbb{Q}}_{(1)})^2}{\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0|\psi_0)^{-1}\mathbf{v}}\mathrm{d}\theta \to 0 \text{ in } \mathbb{Q}_n.$$

[5] The total variation distance between the constrained posterior on $\theta_{(1)}$ and normal distribution is $\int\left|\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) - \varphi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\right|\mathrm{d}\theta_{(1)}$.

This is actually a direct implication from Theorem 1 and Proposition 1 of Chernozhukov and Hong (2003).

Step 2.3: Similar argument can be used to prove that $B_{n,3} \to 0$ in $\mathbb{Q}_n$. More precisely, it is equivalent to show that

$$\frac{1}{2}\int \left[\pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) - \varphi_{\mathcal{Q}}\left(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}\right)\right] n\frac{(\theta_{(1)}-\widehat{\theta}^{\mathcal{P}}_{(1)})^2}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}\mathrm{d}\theta_{(1)} \to 0 \text{ in } \mathbb{Q}_n,$$

and

$$\frac{1}{2}\int \left[\pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) - \varphi_{\mathcal{Q}}\left(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}\right)\right] n\frac{2(\theta_{(1)}-\widehat{\theta}^{\mathcal{Q}}_{(1)})(\theta^{\mathcal{Q}}_{(1)}-\widehat{\theta}^{\mathcal{P}}_{(1)})}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}\mathrm{d}\theta_{(1)} \to 0 \text{ in } \mathbb{Q}_n,$$

and

$$\frac{1}{2}\int \left[\pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) - \varphi_{\mathcal{Q}}\left(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}\right)\right] n\frac{(\theta^{\mathcal{Q}}_{(1)}-\widehat{\theta}^{\mathcal{P}}_{(1)})^2}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}\mathrm{d}\theta_{(1)} \to 0 \text{ in } \mathbb{Q}_n.$$

According to Chernozhukov and Hong (2003, Theorem 1 and Proposition 1), all the three limiting conditions above are satisfied.

Step 3: We prove the weak convergence of C_n in (126). For a constant $r > 0$, we decompose the term C_n as follows

$$C_n = C_{n,1} + C_{n,2} - C_{n,3}, \tag{133}$$

where

$$C_{n,1} = \int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \ln\frac{\varphi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n})}{\pi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n})}\mathrm{d}\theta_{(1)} \tag{134}$$

$$C_{n,2} = \int_{\Omega(\theta_{0,(1)},r)^{\complement}} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \ln\varphi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n})\mathrm{d}\theta_{(1)} \tag{135}$$

$$C_{n,3} = \int_{\Omega(\theta_{0,(1)},r)^{\complement}} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \ln\pi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n})\mathrm{d}\theta_{(1)}. \tag{136}$$

Step 3.1: we show that $C_{n,3} \to 0$ in $\mathbb{Q}_n$. Equivalently, we show that for any $\epsilon > 0$

$$\limsup_{n\to+\infty} \mathbb{Q}_n\left\{|C_{n,3}| > \epsilon\right\} < \epsilon. \tag{137}$$

Let

$$\mathcal{A}_n(\eta) \equiv \left\{\frac{\int_{\Omega(\theta_{0,(1)},r)^{\complement}}\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{P}}(\theta_{(1)},\theta_{(-1)})\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_{(1)},\theta_{(-1)})\mathrm{d}\theta_{(1)}\mathrm{d}\theta_{(-1)}}{\int_{\Omega(\theta_{0,(1)},r)}\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{P}}(\theta_{(1)},\theta_{(-1)})\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_{(1)},\theta_{(-1)})\mathrm{d}\theta_{(1)}\mathrm{d}\theta_{(-1)}} < \eta\right\}.$$

By Proposition 21, we know that for any $\eta > 0$

$$\mathbb{Q}_n\mathcal{A}_n(\eta)^{\complement} = o(1).$$

Define the set

$$\mathcal{B}_n(\eta) \equiv \left\{ \sup_{\theta \in \Theta} \left| \widehat{H}_{\mathcal{P},n}(\theta) - H_{\mathcal{P}}(\theta) \right| < \eta \right\},$$

where $H_{\mathcal{P}}(\theta)$ and $\widehat{H}_{\mathcal{P},n}(\theta)$ are defined in (101) and (95), respectively. By Proposition 7, we know that for any $\eta > 0$

$$\mathbb{Q}_n \mathcal{B}_n(\eta)^{\complement} = o(1).$$

Define the set

$$\mathcal{I}_{1,n}(\delta, \eta) \equiv \left\{ \left|\left| \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathcal{P}}(\theta) \right|\right|_{\mathcal{S}} \leq \eta \left|\left| \mathbf{I}_{\mathcal{P}}(\theta)^{-1} \right|\right|_{\mathcal{S}}^{-1}, \ \ \forall\, \theta \in \Omega(\theta_0, \delta) \text{ and } \tilde{\theta} \in \Omega(\theta, \delta) \right\},$$

By Proposition 12, we know that for any $\eta > 0$ there exists $\delta > 0$ such that

$$\mathbb{Q}_n \mathcal{I}_{1,n}(\delta, \eta)^{\complement} = o\left(\frac{1}{n}\right).$$

Define the set

$$\mathcal{E}_n(\delta) \equiv \left\{ \widehat{\theta}^{\mathcal{P}} \in \Omega(\theta_0, \delta) \right\}.$$

By consistency of MLE for limited-information likelihoods, we know that

$$\mathbb{Q}_n \mathcal{E}_n(\delta)^{\complement} = o(1).$$

Let

$$\mathcal{K}_n(\delta, \eta) \equiv \left\{ \pi_{\mathcal{P}}(\mathbf{x^n}) \leq (1 + \eta) \int_{\Omega(\theta_0, \delta)} \pi_{\mathcal{P}}(\theta) \pi_{\mathcal{P}}(\mathbf{x^n} | \theta) \mathrm{d}\theta \right\}.$$

According to Proposition 21, we know that for any $\delta > 0$ and $\eta > 0$ we have

$$\mathbb{Q}_n \mathcal{K}_n(\delta, \eta)^{\complement} \to 0.$$

We define

$$\mathcal{M}_n(\delta, \eta) \equiv \mathcal{A}_n(\eta) \cap \mathcal{B}_n(\eta) \cap \mathcal{I}_{1,n}(\delta, \eta) \cap \mathcal{E}_n(\delta) \cap \mathcal{K}_n(\delta, \eta),$$

then, we have

$$\mathbb{Q}_n \mathcal{M}_n(\delta, \eta)^{\complement} = o(1).$$

Let's consider the decomposition

$$\mathbb{Q}_n \left\{ |C_{n,3}| > \epsilon \right\} \leq \mathbb{Q}_n \mathcal{M}_n(\delta, \eta) \cap \left\{ |C_{n,3}| > \epsilon \right\} + \mathbb{Q}_n \mathcal{M}_n(\delta, \eta)^{\complement}.$$

Our strategy of proving the result (137) is to find random variables $\overline{C}_{n,3}$ and $\underline{C}_{n,3}$ such that

$$\underline{C}_{n,3} \leq C_{n,3} \leq \overline{C}_{n,3} \quad \text{on } \ \mathcal{A}_n(\eta) \ \ \text{(of course on) } \mathcal{M}_n(\delta, \eta)$$

and

$$\underline{C}_{n,3} \to 0 \text{ in } \mathbb{Q}_n \text{ and } \overline{C}_{n,3} \to 0 \text{ in } \mathbb{Q}_n.$$

Thus, we have

$$\begin{aligned}
\limsup_{n\to+\infty} \mathbb{Q}_n \{|C_{n,3}| > \epsilon\} &\le \limsup_{n\to+\infty} \mathcal{Q}_0^n \mathcal{M}_n(\delta,\eta) \cap \{|C_{n,3}| > \epsilon\} \\
&\le \limsup_{n\to+\infty} \mathbb{Q}_n \mathcal{M}_n(\delta,\eta) \cap \left\{\max\{|\overline{C}_{n,3}|,\ |\underline{C}_{n,3}|\} > \epsilon\right\} \\
&\le \limsup_{n\to+\infty} \mathbb{Q}_n \mathcal{M}_n(\delta,\eta) \cap \left\{|\overline{C}_{n,3}| > \epsilon\right\} + \limsup_{n\to+\infty} \mathbb{Q}_n \mathcal{M}_n(\delta,\eta) \cap \left\{|\underline{C}_{n,3}| > \epsilon\right\} = 0.
\end{aligned}$$

Now, let's figure out the limits of $\overline{C}_{n,3}$ and $\underline{C}_{n,3}$. On the event $\mathcal{M}_n(\delta,\eta)$, we have that

$$\begin{aligned}
\pi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n}) &= \frac{\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{P}}(\theta)\pi_{\mathcal{P}}(\theta|\mathbf{x^n})\mathrm{d}\theta_{(-1)}}{\int_{\Theta} \pi_{\mathcal{P}}(\theta)\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta} && (138)\\
&= \frac{\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{P}}(\theta)\pi_{\mathcal{P}}(\theta|\mathbf{x^n})\mathrm{d}\theta_{(-1)}}{\int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathcal{P}}(\theta)\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta \left(1 + \frac{\int_{\Omega(\theta_{0,(1)},r)^{\complement}} \pi_{\mathcal{P}}(\theta)\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta}{\int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathcal{P}}(\theta)\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta}\right)} && (139)\\
&\ge \frac{\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{P}}(\theta)\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta_{(-1)}}{(1+\eta)\int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathcal{P}}(\theta)\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta} && (140)
\end{aligned}$$

Because $\widehat{\theta}^{\mathcal{P}}$ is the MLE for limited-information likelihood of baseline model $\mathbb{P}_\theta$, we have

$$\pi_{\mathcal{P}}(\mathbf{x^n}|\theta) \le \pi_{\mathcal{P}}(\mathbf{x^n}|\widehat{\theta}^{\mathcal{P}}) = \exp\left\{-\ln\frac{1}{\pi_{\mathcal{P}}(\mathbf{x^n}|\widehat{\theta}^{\mathcal{P}})}\right\} = \exp\left\{-n\widehat{H}_{\mathcal{P},n}(\widehat{\theta}^{\mathcal{P}})\right\} \tag{141}$$

Thus, we have

$$\pi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n}) \ge \frac{\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{P}}(\theta)\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta_{(-1)}}{(1+\eta)\pi_{\mathcal{P}}(\Omega(\theta_{0,(1)},r))\exp\{-n\widehat{H}_{\mathcal{P},n}(\widehat{\theta}^{\mathcal{P}})\}} \tag{142}$$

Plug (142) into the expression for $C_{n,3}$, we have

$$
\begin{aligned}
C_{n,3} &\geq \int_{\Omega(\theta_{0,(1)},r)^c} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \ln \frac{\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{P}}(\theta)\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta_{(-1)}}{(1+\eta)\pi_{\mathcal{P}}(\Omega(\theta_{0,(1)},r))\exp\{-n\widehat{H}_{\mathcal{P},n}(\widehat{\theta}^{\mathcal{P}})\}}\mathrm{d}\theta_{(1)} \\
&\geq \int_{\Omega(\theta_{0,(1)},r)^\complement} \int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\pi_{\mathcal{P}}(\theta_{(-1)}|\theta_{(1)}) \ln\left[\pi_{\mathcal{P}}(\theta_{(1)})\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)\right]\mathrm{d}\theta_{(1)}\mathrm{d}\theta_{(-1)} \\
&\quad - \int_{\Omega(\theta_{0,(1)},r)^\complement} \pi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \left[\ln(1+\eta) + \ln \pi_{\mathcal{P}}(\Omega(\theta_{0,(1)},r)) - nH_{\mathcal{P},n}(\widehat{\theta}^{\mathcal{P}})\right]\mathrm{d}\theta_{(1)} \\
&= \int_{\Omega(\theta_{0,(1)},r)^\complement} \int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\pi_{\mathcal{P}}(\theta_{(-1)}|\theta_{(1)}) \ln \pi_{\mathcal{P}}(\theta_{(1)})\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta_{(1)}\mathrm{d}\theta_{(-1)} \\
&\quad - \pi_{\mathcal{P}}(\Omega(\theta_{0,(1)},r)^\complement|\mathbf{x^n},\mathbf{y^n}) \left[\ln(1+\eta) + \ln \pi_{\mathcal{P}}(\Omega(\theta_{0,(1)},r)) - n\widehat{H}_{\mathcal{P},n}(\widehat{\theta}^{\mathcal{P}})\right] \qquad (143)
\end{aligned}
$$

We define the term in (143) to be $\underline{C}_{n,3}$. Thus, we can further decompose $\underline{C}_{n,3}$ as follows,

$$\underline{C}_{n,3} = \underline{C}_{n,3,1} - \underline{C}_{n,3,2}$$

where

$$\underline{C}_{n,3,1} = \int_{\Omega(\theta_{0,(1)},r)^\complement} \int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\pi_{\mathcal{P}}(\theta_{(-1)}|\theta_{(1)}) \ln\left[\pi_{\mathcal{P}}(\theta_{(1)})\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)\right]\mathrm{d}\theta_{(1)}\mathrm{d}\theta_{(-1)}$$

and

$$\underline{C}_{n,3,2} = \pi_{\mathcal{Q}}(\Omega(\theta_{0,(1)},r)^\complement|\mathbf{x^n},\mathbf{y^n}) \left[\ln(1+\eta) + \ln \pi_{\mathcal{P}}(\Omega(\theta_{0,(1)},r)) - n\widehat{H}_{\mathcal{P},n}(\widehat{\theta}^{\mathcal{P}})\right].$$

We have

$$
\begin{aligned}
&|\underline{C}_{n,3,1}| \leq \left|\int_{\Omega(\theta_{0,(1)},r)^\complement} \int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\pi_{\mathcal{P}}(\theta_{(-1)}|\theta_{(1)}) \ln \pi_{\mathcal{P}}(\theta_{(1)})\mathrm{d}\theta_{(1)}\mathrm{d}\theta_{(-1)}\right| \\
&+ \left|\int_{\Omega(\theta_{0,(1)},r)^\complement} \int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\pi_{\mathcal{P}}(\theta_{(-1)}|\theta_{(1)}) \ln \pi_{\mathcal{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta_{(1)}\mathrm{d}\theta_{(-1)}\right| \\
&\leq \int_{\Omega(\theta_{0,(1)},r)^\complement} \int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\pi_{\mathcal{P}}(\theta_{(-1)}|\theta_{(1)}) \left|\ln \pi_{\mathcal{P}}(\theta_{(1)})\right|\mathrm{d}\theta_{(1)}\mathrm{d}\theta_{(-1)} \qquad (144) \\
&+ \int_{\Omega(\theta_{0,(1)},r)^\complement} \int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\pi_{\mathcal{P}}(\theta_{(-1)}|\theta_{(1)}) n|\widehat{H}_{\mathcal{P},n}(\theta) - H_{\mathcal{P}}(\theta)|\mathrm{d}\theta_{(1)}\mathrm{d}\theta_{(-1)} \qquad (145) \\
&+ \int_{\Omega(\theta_{0,(1)},r)^\complement} \int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\pi_{\mathcal{P}}(\theta_{(-1)}|\theta_{(1)}) n|H_{\mathcal{P}}(\theta)|\mathrm{d}\theta_{(1)}\mathrm{d}\theta_{(-1)} \qquad (146) \\
&+ o_p(1).
\end{aligned}
$$

The term (144) can be bounded from above by

$$M_1 \int_{\Omega(\theta_{0,(1)},r)^\complement} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\mathrm{d}\theta_{(1)} = M_1 \pi_{\mathcal{Q}}(\Omega(\theta_{0,(1)},r)^\complement|\mathbf{x^n},\mathbf{y^n}) \to 0 \text{ in } \mathbb{Q}_n, \qquad (147)$$

where the existence of such constant M_1 is due to the compactness of $\Theta \subset \mathbb{R}^d$ and the continuity of $\pi_{\mathcal{P}}(\theta_{(-1)}|\theta_{(1)})|\ln \pi_{\mathcal{P}}(\theta_{(1)})|$ and we have

$$\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{P}}(\theta_{(-1)}|\theta_{(1)})|\ln \pi_{\mathcal{P}}(\theta_{(1)})|\mathrm{d}\theta_{(-1)} \leq M_1. \tag{148}$$

The term (145), for large enough n, is bounded from above by

$$n\int_{\Omega(\theta_{0,(1)},r)^{\complement}} \int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\pi_{\mathcal{P}}(\theta_{(-1)}|\theta_{(1)})\mathrm{d}\theta_{(1)}\mathrm{d}\theta_{(-1)}, \tag{149}$$

because

$$\sup_{\theta\in\Theta} |\widehat{H}_{\mathcal{P},n}(\theta) - H_{\mathcal{P}}(\theta)| < 1 \ \text{ for large enough } \ n. \tag{150}$$

The term (149) can be further bounded from above by

$$M_2 n\int_{\Omega(\theta_{0,(1)},r)^{\complement}} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\mathrm{d}\theta_{(1)} = M_2 n\pi_{\mathcal{Q}}(\Omega(\theta_{0,(1)},r)^{\complement}|\mathbf{x^n},\mathbf{y^n}) \to 0 \ \text{ in } \ \mathbb{Q}_n, \tag{151}$$

where the existence of such constant M_2 is due to the compactness of the $\Theta \subset \mathbb{R}^d$ and the continuity of $\pi_{\mathcal{P}}(\theta_{(-1)}|\theta_{(1)}) = \frac{\pi_{\mathcal{P}}(\theta_{(1)},\theta_{(-1)})}{\pi_{\mathcal{P}}(\theta_{(1)})}$,

$$\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{P}}(\theta_{(-1)}|\theta_{(1)})\mathrm{d}\theta_{(-1)} \leq M_2. \tag{152}$$

The term (146) is bounded from above by

$$M_3 n\int_{\Omega(\theta_{0,(1)},r)^{\complement}} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) = M_3 n\pi_{\mathcal{Q}}(\Omega(\theta_{0,(1)},r)|\mathbf{x^n},\mathbf{y^n}) \to 0 \ \text{ in } \ \mathbb{Q}_n. \tag{153}$$

Therefore, the term $\underline{C}_{n,3,1} \to 0$ in $\mathbb{Q}_n$. It is straightforward to see that $\underline{C}_{n,3,2}$ converges to zero in probability, because $n\pi_{\mathcal{Q}}(\Omega(\theta_{0,(1)},r)|\mathbf{x^n},\mathbf{y^n}) \to 0$ in $\mathbb{Q}_n$ and $\widehat{H}_{\mathcal{P},n}(\widehat{\theta}^{\mathcal{P}}) \to H_{\mathcal{P}}(\theta_0)$ in $\mathbb{Q}_n$.

Now, let's construct $\overline{C}_{n,3}$ and show it indeed converges to zero in probability. By restricting the domain to $\Omega(\theta_0,\delta)$, we have

$$\pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \leq \frac{\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{P}}(\theta)\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta_{(-1)}}{\int_{\Omega(\theta_0,\delta)} \pi_{\mathcal{P}}(\theta)\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta}. \tag{154}$$

By Taylor expansion of $\ln \pi_{\mathcal{P}}(\mathbf{x^n}|\theta)$ around $\widehat{\theta}^{\mathcal{P}}$, we have

$$\begin{aligned}
\pi_{\mathcal{P}}(\theta|\mathbf{x^n}) &\leq \frac{\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{P}}(\theta)\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta_{(-1)}}{m_\pi \pi_{\mathcal{P}}(\mathbf{x^n}|\widehat{\theta}^{\mathcal{P}}) \int_{\Omega(\theta_0,\delta)} \exp\{-n(\theta-\widehat{\theta}^{\mathcal{P}})^T \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})(\theta-\widehat{\theta}^{\mathcal{P}})\}\mathrm{d}\theta} \\
&\leq \frac{\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{P}}(\theta)\mathrm{d}\theta_{(-1)}}{m_\pi \int_{\Omega(\theta_0,\delta)} \exp\{-n(\theta-\widehat{\theta}^{\mathcal{P}})^T \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})(\theta-\widehat{\theta}^{\mathcal{P}})\}\mathrm{d}\theta} \\
&\leq \frac{M_4}{\int_{\Omega(\theta_0,\delta)} \exp\{-n(\theta-\widehat{\theta}^{\mathcal{P}})^T \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})(\theta-\widehat{\theta}^{\mathcal{P}})\}\mathrm{d}\theta}
\end{aligned}$$

where $\tilde{\theta}$ is on the segment between $\widehat{\theta}^{\mathcal{P}}$ and θ. And, the existence of the constant M_4 such that

$$\frac{\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathcal{P}}(\theta)\mathrm{d}\theta_{(-1)}}{m_\pi} \leq M_4. \tag{155}$$

Thus, we have

$$C_{n,3} \leq \int_{\Omega(\theta_{0,(1)},r)^{\complement}} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \ln \frac{M_4 \mathrm{d}\theta_{(1)}}{\int_{\Omega(\theta_0,\delta)} \exp\{-\frac{n}{2}(\theta-\widehat{\theta}^{\mathcal{P}})^T \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})(\theta-\widehat{\theta}^{\mathcal{P}})\}\mathrm{d}\theta} \tag{156}$$

We define $\overline{C}_{n,3}$ to be the term on the right hand side of the inequality (156). On the event $\mathcal{M}_n(\delta,\eta)$, we have

$$(\theta-\widehat{\theta}^{\mathcal{P}})^T \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})(\theta-\widehat{\theta}^{\mathcal{P}}) \leq 2(\theta-\widehat{\theta}^{\mathcal{P}})^T \mathbf{I}_{\mathcal{P}}(\theta_0)(\theta-\widehat{\theta}^{\mathcal{P}})$$

and

$$C_{n,3} \leq \int_{\Omega(\theta_{0,(1)},r)^{\complement}} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \ln \frac{M_4 \mathrm{d}\theta_{(1)}}{\int_{\Omega(\theta_0,\delta)} \exp\{-n(\theta-\widehat{\theta}^{\mathcal{P}})^T \mathbf{I}_{\mathcal{P}}(\theta_0)(\theta-\widehat{\theta}^{\mathcal{P}})\}\mathrm{d}\theta}$$

By the normal distribution and the $\sqrt{n}-$ consistency of MLE $\widehat{\theta}^{\mathcal{P}}$, we know that for any $\nu > 0$, it follows that

$$\mathbb{Q}_n \left\{ \frac{\int_{\Omega(\theta_0,\delta)} \exp\{-n(\theta-\widehat{\theta}^{\mathcal{P}})^T \mathbf{I}_{\mathcal{P}}(\theta_0)(\theta-\widehat{\theta}^{\mathcal{P}})\}\mathrm{d}\theta}{\int_{\mathbb{R}^{D_\Theta}} \exp\{-n(\theta-\widehat{\theta}^{\mathcal{P}})^T \mathbf{I}_{\mathcal{P}}(\theta_0)(\theta-\widehat{\theta}^{\mathcal{P}})\}\mathrm{d}\theta} < 1-\nu \right\} \to 0. \tag{157}$$

Thus, we have

$$\begin{aligned}
&\limsup_{n\to+\infty} \mathbb{Q}_n\{|\overline{C}_{n,3}| > \epsilon\} \\
&\leq \limsup_{n\to+\infty} \mathbb{Q}_n \left\{ \int_{\Omega(\theta_{0,(1)},r)^{\complement}} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \ln \frac{(1-\nu)^{-1} M_4 \mathrm{d}\theta_{(1)}}{\int_{\mathbb{R}^{D_\Theta}} \exp\{-n(\theta-\widehat{\theta}^{\mathcal{P}})^T \mathbf{I}_{\mathcal{P}}(\theta_0)(\theta-\widehat{\theta}^{\mathcal{P}})\}\mathrm{d}\theta} > \epsilon \right\} \\
&= \limsup_{n\to+\infty} \mathbb{Q}_n \left\{ \int_{\Omega(\theta_{0,(1)},r)^{\complement}} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \left| \ln \frac{M_4 (2\pi)^{D_\Theta/2}}{(1-\nu)|2n\mathbf{I}_{\mathcal{P}}(\theta_0)|^{1/2}} \right| > \epsilon \right\} \\
&= \limsup_{n\to+\infty} \mathbb{Q}_n \left\{ \pi_{\mathcal{Q}}(\Omega(\theta_{0,(1)},r)^{\complement}|\mathbf{x^n},\mathbf{y^n}) \left| \ln \frac{M_4 (2\pi)^{D_\Theta/2}}{(1-\nu)|2n\mathbf{I}_{\mathcal{P}}(\theta_0)|^{1/2}} \right| > \epsilon \right\} = 0,
\end{aligned}$$

where the last limiting result is a direct implication of Proposition 18. Thus, $\underline{C}_{n,3} \to 0$ in $\mathbb{Q}_n$. Therefore, we have $C_{n,3} \to 0$ in $\mathbb{Q}_n$.

Step 3.2: We show $C_{n,2}$ goes to zero in $\mathbb{Q}_n$. The expression (135) for $C_{n,2}$ can be rewritten as

$$C_{n,2} = \pi_{\mathcal{Q}}(\Omega(\theta_{0,(1)}, r)^{\complement}|\mathbf{x^n}, \mathbf{y^n}) \ln \frac{|n\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)\mathbf{v}|^{1/2}}{(2\pi)^{1/2}} - \int_{\Omega(\theta_{0,(1)},r)^{\complement}} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n}) \frac{n(\theta_{(1)} - \widehat{\theta}^{\mathcal{P}}_{(1)})^2}{2\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}} \mathrm{d}\theta_{(1)}$$

Thus, the term $C_{n,2}$ can be decomposed as follows

$$C_{n,2} = \pi_{\mathcal{Q}}(\Omega(\theta_{0,(1)}, r)^{\complement}|\mathbf{x^n}, \mathbf{y^n}) \ln \frac{|n\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)\mathbf{v}|^{1/2}}{(2\pi)^{1/2}} \tag{158}$$

$$- \int_{\Omega(\theta_{0,(1)},r)^{\complement}} \left[\pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n}) - \varphi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n})\right] \frac{n(\theta_{(1)} - \widehat{\theta}^{\mathcal{P}}_{(1)})^2}{2\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)\mathbf{v}} \mathrm{d}\theta_{(1)} \tag{159}$$

$$- \int_{\Omega(\theta_{0,(1)},r)^{\complement}} \varphi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n}) \frac{n(\theta_{(1)} - \widehat{\theta}^{\mathcal{P}}_{(1)})^2}{2\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}} \mathrm{d}\theta_{(1)} \tag{160}$$

It is easy to see that the first term (158) goes to zero in $\mathbb{Q}_n$. The second term (159) and the third term (160) go to zero in probability according to Theorem 1 and Proposition 1 in Chernozhukov and Hong (2003) and the fact that $n(\widehat{\theta}^{\mathcal{Q}} - \widehat{\theta}^{\mathcal{P}})^2 = O_p(1)$. Therefore, we have shown that $C_{n,2} \to 0$ in $\mathbb{Q}_n$.

Step 3.3: We need to prove $C_{n,1}$ goes to zero in $\mathbb{Q}_n$. According to Corollary 5, we know that for any $\eta > 0$ there exists $\delta_0 > 0$ such that

$$\mathbb{P}_n \mathcal{I}_n(\delta_0, \eta) \to 1, \quad \text{as } n \to \infty,$$

with

$$\mathcal{I}_n(\delta_0, \eta) \equiv \left\{1 - \eta \leq \left|\left|\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1/2}\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1/2}\right|\right|_{\mathcal{S}} \leq 1 + \eta, \text{for all } \tilde{\theta} \in \Omega(\theta_0, \delta_0)\right\}.$$

Also, by the continuity and positivity of the prior density $\pi_{\mathcal{P}}(\theta)$, we know that for any $\eta > 0$, there exists δ_1 small enough it holds that $1 - \eta \leq \pi_{\mathcal{P}}(\theta)/\pi_{\mathcal{P}}(\theta') \leq 1 + \eta$ for all $\theta, \theta' \in \Omega(\theta_0, \delta_1)$. According to the consistency of MLE $\widehat{\theta}^{\mathcal{P}}$, we shall only focus on the event $\mathcal{A}_n(\delta) \equiv \widehat{\theta}^{\mathcal{P}} \in \Omega(\theta_0, \delta)$ with $\delta = \min(\delta_0, \delta_1)$. On the joint large probability event $\mathcal{I}_n(\delta, \eta) \cap \mathcal{A}_n(\eta)$, we have for each $\theta \in \Omega(\theta_0, r)$ with $r < \delta$, when n is large enough,

$$\pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \leq \frac{\pi_{\mathcal{P}}(\theta)\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)}{\int_{\Omega(\theta_0,r)} \pi_{\mathcal{P}}(\vartheta)\pi_{\mathcal{P}}(\mathbf{x^n}|\vartheta)\mathrm{d}\vartheta} \leq (1+\eta)^2 \left[\frac{n(1+\eta)}{2\pi}\right]^{D_\Theta/2} [\det \mathbf{I}_{\mathcal{P}}(\theta_0)]^{1/2} e^{-\frac{1}{2}(1-\eta)n(\theta-\widehat{\theta}^{\mathcal{P}})^T\mathbf{I}_{\mathcal{P}}(\theta_0)(\theta-\widehat{\theta}^{\mathcal{P}})}.$$

On the event $\mathcal{I}_n(\delta,\eta)\cap\mathcal{A}_n(\eta)$ we have

$$
\begin{aligned}
C_{n,1} \geq & \int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \ln \frac{1}{\sqrt{2\pi n^{-1}|\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}|}} \mathrm{d}\theta_{(1)} \\
& - \int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \frac{n(\theta_{(1)}-\widehat{\theta}^{\mathcal{P}}_{(1)})^2}{2\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}} \mathrm{d}\theta_{(1)} \\
& - \int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \ln\left[\int_{\Theta_{-1}(\theta_{(1)})} e^{-\frac{n}{2}(1-\eta)(\theta-\widehat{\theta}^{\mathcal{P}})^T\mathbf{I}_{\mathcal{P}}(\theta_0)(\theta-\widehat{\theta}^{\mathcal{P}})} \frac{[(1-\eta)n]^{D_\Theta/2}[\det\mathbf{I}_{\mathcal{P}}(\theta_0)]^{1/2}}{(2\pi)^{D_\Theta/2}} \mathrm{d}\theta_{(-1)}\right] \mathrm{d}\theta_{(1)} \\
& - \left[\frac{D_\Theta}{2}+2\right]\ln(1+\eta) + \frac{D_\Theta}{2}\ln(1-\eta).
\end{aligned}
$$

There exists open square centered at θ_0 which is denoted as $\mathcal{O}=\Omega(\theta_{0,(1)},r)\otimes\Omega(\theta_{0,(-1)},\delta)$. First, we have

$$
\int_{\Theta_{-1}(\theta_{(1)})} e^{-\frac{n}{2}(1-\eta)(\theta-\widehat{\theta}^{\mathcal{P}})^T\mathbf{I}_{\mathcal{P}}(\theta_0)(\theta-\widehat{\theta}^{\mathcal{P}})} \frac{[(1-\eta)n]^{D_\Theta/2}[\det\mathbf{I}_{\mathcal{P}}(\theta_0)]^{1/2}}{(2\pi)^{D_\Theta/2}} \mathrm{d}\theta_{(-1)} \leq \frac{e^{-(1-\eta)\frac{n(\theta_{(1)}-\widehat{\theta}^{\mathcal{P}}_{(1)})^2}{2\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}}}{\sqrt{2\pi n^{-1}\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}/(1-\eta)}}.
$$

Thus, on the large probability event $\mathcal{I}_n(\delta,\eta)\cap\mathcal{A}_n(\eta)$ we have, when n is large

$$
C_{n,1} \geq -\left[\frac{D_\Theta}{2}+2\right]\ln(1+\eta) + \left[\frac{D_\Theta-1}{2}\right]\ln(1-\eta) + \eta\int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \frac{n(\theta_{(1)}-\widehat{\theta}^{\mathcal{P}}_{(1)})^2}{2\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}} \mathrm{d}\theta_{(1)}.
$$

According to Theorem 1 and Proposition 1 of Chernozhukov and Hong (2003), we know that

$$
\int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \frac{n(\theta_{(1)}-\widehat{\theta}^{\mathcal{P}}_{(1)})^2}{2\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}} \mathrm{d}\theta_{(1)} = O_p(1).
$$

Therefore, it follows that

$$
C_{n,1} \geq -\left[\frac{D_\Theta}{2}+2\right]\ln(1+\eta) + \left[\frac{D_\Theta-1}{2}\right]\ln(1-\eta) + \eta O_p(1). \tag{161}
$$

On the other hand, by Proposition 21, we know that for any $\eta'>0$ and $\delta'>0$

$$
\mathbb{P}_n\mathcal{X}_n(\delta',\eta') \to 1, \quad \text{as } n\to\infty,
$$

with

$$
\mathcal{X}_n(\delta',\eta') \equiv \left\{\pi_{\mathcal{P}}(\mathbf{x^n}) \leq (1+\eta')\int_{\Omega(\theta_0,\delta')} \pi_{\mathcal{P}}(\theta)\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta\right\}.
$$

Then, on the large probability event $\mathcal{I}_n(\delta,\eta)\cap\mathcal{A}_n(\eta)\cap\mathcal{X}_n(\delta,\eta)$ and taking $r<\delta$, it holds that

$$\begin{aligned}
C_{n,1} &= \int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \left[\ln \frac{\varphi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n})\pi_{\mathcal{P}}(\mathbf{x^n})}{\int_{\Theta_{-1}(\theta_{(1)})}\pi_{\mathcal{P}}(\theta)\pi_{\mathcal{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta_{(-1)}}\right]\mathrm{d}\theta_{(1)} \\
&\leq \int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \ln\left[\frac{(2\pi)^{-1/2}|n^{-1}\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}|^{-1/2}e^{-\frac{n(\theta_{(1)}-\widehat{\theta}^{\mathcal{P}}_{(1)})^2}{2\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}}}{\int_{\Theta_{-1}(\theta_{(1)})}(1-\eta)e^{-(1+\eta)\frac{n}{2}(\theta-\widehat{\theta}^{\mathcal{P}})^T\mathbf{I}_{\mathcal{P}}(\theta_0)(\theta-\widehat{\theta}^{\mathcal{P}})}\mathrm{d}\theta_{(-1)}}\right]\mathrm{d}\theta_{(1)} \\
&\quad + \int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \ln\left[\int_{\Omega(\theta_0,\delta)}(1+\eta)e^{-(1-\eta)\frac{n}{2}(\vartheta-\widehat{\theta}^{\mathcal{P}})^T\mathbf{I}_{\mathcal{P}}(\theta_0)(\vartheta-\widehat{\theta}^{\mathcal{P}})}\mathrm{d}\vartheta\right]\mathrm{d}\theta_{(1)}.
\end{aligned}$$

Calculating the integrations, we have

$$C_{n,1} \leq \frac{D_\Theta+1}{2}\ln(1+\eta) - \frac{D_\Theta+2}{2}\ln(1-\eta) - \inf_{\theta_{(1)}\in\Omega(\theta_{0,(1)},r)}\ln\Phi_n(\Theta_{-1}(\theta_{(1)})|\theta_{(1)}) + \eta\int \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\frac{n(\theta_{(1)}-\widehat{\theta}^{\mathcal{P}}_{(1)})^2}{2\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)\mathbf{v}}\mathrm{d}\theta.$$

where $\Phi_n(\cdot|\theta_{(1)})$ is the multivariate normal probability measure on $\mathbb{R}^{D_\Theta-1}$ and effectively it is the conditional distribution of $\theta_{(-1)}$ given $\theta_{(1)}$ where $\theta=(\theta_{(1)},\theta_{(-1)})\sim N(\widehat{\theta}^{\mathcal{P}},(1+\eta)^{-1}n^{-1}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1})$. In fact, $\Phi_n(\cdot|\theta_{(1)})$ is multivariate normal with distribution

$$N\left(\widehat{\theta}^{\mathcal{P}}_{(-1)}+\Sigma_{21}\Sigma_{11}^{-1}(\theta_{(1)}-\widehat{\theta}^{\mathcal{P}}_{(1)}),(1+\eta)^{-1}n^{-1}(\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})\right)$$

with

$$\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1} = \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}.$$

We choose δ and r small enough such that $\widehat{\theta}^{\mathcal{P}}_{(-1)}+\Sigma_{21}\Sigma_{11}^{-1}(\theta_{(1)}-\widehat{\theta}^{\mathcal{P}}_{(1)})$ is in the interior of $\Theta_{-1}(\theta_{(1)})$ and there exists $\tau_0>0$ such that $d_L(\widehat{\theta}^{\mathcal{P}}_{(-1)}+\Sigma_{21}\Sigma_{11}^{-1}(\theta_{(1)}-\widehat{\theta}^{\mathcal{P}}_{(1)}),\partial\Theta_{-1}(\theta_{(1)}))>\tau_0$ for all $\theta_{(1)}\in\Omega(\theta_{0,(1)})$. Thus,

$$\inf_{\theta_{(1)}\in\Omega(\theta_{0,(1)},r)}\Phi_n(\Theta_{-1}(\theta_{(1)})|\theta_{(1)})\to 1.$$

Therefore, when n is large enough, we have

$$\inf_{\theta_{(1)}\in\Omega(\theta_{0,(1)},r)}\Phi_n(\Theta_{-1}(\theta_{(1)})|\theta_{(1)}) > \frac{1}{2}\ln(1-\eta).$$

And hence, we have

$$C_{n,1} \leq \frac{D_\Theta+1}{2}\ln(1+\eta) - \frac{D_\Theta+3}{2}\ln(1-\eta) + \eta\int \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\frac{n(\theta_{(1)}-\widehat{\theta}^{\mathcal{P}}_{(1)})^2}{2\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)\mathbf{v}}\mathrm{d}\theta.$$

According to Theorem 1 and Proposition 1 of Chernozhukov and Hong (2003), we know that

$$\int \pi_{\mathcal{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\frac{n(\theta_{(1)}-\widehat{\theta}^{\mathcal{P}}_{(1)})^2}{2\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)\mathbf{v}}\mathrm{d}\theta = O_p(1).$$

Therefore, we can get

$$C_{n,1} \leq \frac{D_\Theta + 1}{2} \ln(1+\eta) - \frac{D_\Theta + 3}{2} \ln(1-\eta) + \eta O_p(1). \tag{162}$$

Combine the bounds in (161) and (162) where the constant η can be arbitrarily small, we know that $C_{n,1} \to 0$ in $\mathbb{Q}_n$.

1.11 Proof of Theorem 3

Because of Assumption **FF** and the assumptions in Subsection 1.4 are invariant under invertible and second-order smooth transformations, without loss of generality, we assume that $f(\theta) = \theta_{(1)}$ and hence $\mathbf{v} = (1, 0, \cdots, 0)^T$. We want to show when m and n go to infinity and $m/n \to \varrho$, we have for any $\epsilon > 0$ that

$$\limsup_{n\to+\infty} \mathbb{P}_n \left\{ \left| \int \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \int \pi(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \ln \frac{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\mathbf{x^n})}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta_{(1)})} \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta - \frac{1}{2} \ln \frac{n}{m+n} \right| > \epsilon \right\} < \epsilon.$$

Denote

$$R_n \equiv \int \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \int \pi(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \ln \frac{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\mathbf{x^n})}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta_{(1)})} \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta - \frac{1}{2} \ln \frac{n}{m+n}.$$

Then, we are going to show for any $\epsilon > 0$ it holds that

$$\limsup_{n\to+\infty} \mathbb{P}_n \left\{ |R_n| > \epsilon \right\} < \epsilon \tag{163}$$

We can further decompose R_n as follows:

$$\begin{aligned}
R_n = & \int \pi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n}) \ln \frac{\pi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n})}{\pi_{\mathcal{P}}(\theta_{(1)})} \mathrm{d}\theta_{(1)} && (164)\\
& + \int \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \int \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \ln \frac{\int \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta')\pi_{\mathcal{P}}(\theta')\mathrm{d}\theta'}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta)} \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta && (165)\\
& - \int \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \int \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \ln \frac{\int \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta_{(1)}, \theta'_{(-1)})\pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta)} \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta && (166)\\
& - \frac{1}{2} \ln \frac{n}{m+n}.
\end{aligned}$$

We denote the term in (164) as $R_{n,1}$, denote the term in (165) as $R_{n,2}$, and denote the term in (166) as $R_{n,3}$.

For the term $R_{n,1}$, we can further decompose it as follows

$$\begin{aligned}
R_{n,1} = & \int \pi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n}) \ln \frac{\pi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n})}{\varphi(\theta_{(1)}|\mathbf{x^n})} \mathrm{d}\theta_{(1)} + \frac{1}{2} \ln(n) - \frac{1}{2} \ln\left(2\pi \mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}\right) \\
& - \int \pi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n}) \frac{n(\theta_{(1)} - \widehat{\theta}^{\mathcal{P}}_{(1)})^2}{2\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}} \mathrm{d}\theta_{(1)} - \int \pi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n}) \ln \pi_{\mathcal{P}}(\theta_{(1)}) \mathrm{d}\theta_{(1)},
\end{aligned}$$

where

$$\varphi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n}) = \frac{1}{\sqrt{2\pi\frac{1}{n}\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}}\exp\left\{-\frac{1}{2\frac{1}{n}\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}(\theta_{(1)}-\widehat{\theta}^{\mathcal{P}}_{(1)})^2\right\}.$$

According to Corollary 3, we know that

$$\int \pi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n})\ln\frac{\pi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n})}{\varphi(\theta_{(1)}|\mathbf{x^n})}\mathrm{d}\theta_{(1)} \to 0 \ \text{ in } \ \mathbb{P}_n. \tag{167}$$

The same argument as in proving the weak convergence of the term involving B_n in (127) can be used to show that

$$\int \pi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n})\frac{n(\theta_{(1)}-\widehat{\theta}^{\mathcal{P}}_{(1)})^2}{2\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}}\mathrm{d}\theta_{(1)} \to \frac{1}{2} \ \text{ in } \ \mathbb{P}_n. \tag{168}$$

Note that the prior $\pi_{\mathcal{P}}(\theta_{(1)})$ is assumed to be continuous on the compact domain. Again, because of Corollary 3 and the fact that total variation distance is bounded by relative entropy, we know that

$$\int \pi_{\mathcal{P}}(\theta_{(1)}|\mathbf{x^n})\ln\pi_{\mathcal{P}}(\theta_{(1)})\mathrm{d}\theta_{(1)} \to \ln\pi_{\mathcal{P}}(\theta_{0,(1)}) \ \text{ in } \ \mathbb{P}_n. \tag{169}$$

Thus, following (167 – 169), it holds that

$$R_{n,1} - \frac{1}{2}\ln(n) - \frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}\right) - \frac{1}{2} - \ln\pi_{\mathcal{P}}(\theta_{0,(1)}) \to 0 \text{ in } \mathbb{P}_n. \tag{170}$$

We shall also show that, in $\mathbb{P}_n$,

$$R_{n,2} + \frac{D_\Theta}{2}\ln(n+m) - \frac{D_\Theta}{2}\ln(2\pi) - \frac{1}{2}\ln\det\left[\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\right] - \ln\pi_{\mathcal{P}}(\theta_0) - \frac{D_\Theta}{2} \to 0. \tag{171}$$

and

$$R_{n,3} + \frac{D_\Theta-1}{2}\ln(n+m) - \frac{D_\Theta-1}{2}\ln(2\pi) - \frac{1}{2}\ln\frac{\det\left[\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\right]}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}} - \ln\pi_{\mathcal{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) - \frac{D_\Theta-1}{2} \to 0 \text{ in } \mathbb{P}_n. \tag{172}$$

Combining the weak convergence results (170 – 172), we can achieve the weak convergence of R_n, or equivalently the result in (163).

The proofs of the result (171) and the result (172) are quite similar, though the proof of the result (172) is a little bit more involving. Without tedious repeating the same proofs, we shall only provide the proof for the result (172). We further define the left-hand side of (172) as $R^*_{n,3}$, that is,

$$R^*_{n,3} \equiv R_{n,3} + \frac{D_\Theta-1}{2}\ln(n+m) - \frac{D_\Theta-1}{2}\ln(2\pi) - \frac{1}{2}\ln\frac{\det\left[\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\right]}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}} - \ln\pi_{\mathcal{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) - \frac{D_\Theta-1}{2}. \tag{173}$$

In the rest of the proof, we shall show that for any $\epsilon > 0$, it holds that

$$\limsup_{n\to+\infty} \mathbb{P}_n \left\{ |R^*_{n,3}| > \epsilon \right\} < \epsilon \tag{174}$$

Step 1: We first define the big probability events which we shall focus on in order to show (174). We define the big probability event

$$\mathcal{A}_n(\theta_0, \delta, \xi) \equiv \left\{ \pi_{\mathcal{P}}(\Omega(\theta_0, \delta)^{\complement} | \mathbf{x^n}) \le e^{-\xi n} \right\}. \tag{175}$$

According to Theorem 1 and Proposition 3 in Chernozhukov and Hong (2003), we know that it suffices to show that

$$\limsup_{n\to+\infty} \mathbb{P}_n \mathcal{A}_n(\theta_0, \delta, \xi) \left\{ |R^*_{n,3}| > \epsilon \right\} < \epsilon. \tag{176}$$

Define the sets

$$\mathcal{I}_{1,n}(\delta, \eta) \equiv \left\{ \left|\left| \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathcal{P}}(\theta) \right|\right|_{\mathcal{S}} \le \eta \left|\left| \mathbf{I}_{\mathcal{P}}(\theta)^{-1} \right|\right|_{\mathcal{S}}^{-1}, \ \ \forall\, \theta \in \Omega(\theta_0, \delta) \text{ and } \tilde{\theta} \in \Omega(\theta, \delta) \right\},$$

$$\mathcal{I}_{2,m}(\theta, \delta, \eta) \equiv \left\{ \left|\left| \widehat{\mathbf{I}}_{\mathcal{P},m}(\tilde{\theta}) - \mathbf{I}_{\mathcal{P}}(\theta) \right|\right|_{\mathcal{S}} \le \eta \left|\left| \mathbf{I}_{\mathcal{P}}(\theta)^{-1} \right|\right|_{\mathcal{S}}^{-1}, \ \ \forall\, \tilde{\theta} \in \Omega(\theta, \delta) \right\}.$$

Appealing to Propositions 12, we know that for any $\eta > 0$ there exists small enough positive constants δ_1 and δ such that

$$\mathbb{P}_n \mathcal{I}_{1,n}(\delta, \eta)^{\complement} = o\left(\frac{1}{n}\right) \quad \text{and} \quad \sup_{\theta\in\Omega(\theta_0,\delta_1)} \mathcal{P}_{\theta,m} \mathcal{I}_{2,m}(\theta, \delta, \eta)^{\complement} = o\left(\frac{1}{m}\right).$$

Define

$$\mathcal{H}_{1,n}(\delta) \equiv \left\{ \left| \widehat{H}_{\mathcal{P},n}(\widehat{\theta}^{\mathcal{P}}) - H(\theta_0) \right| < \delta \right\}, \quad \text{and} \quad \mathcal{H}_{2,n}(\delta) \equiv \left\{ \sup_{\theta\in\Theta} \left| \widehat{H}_{\mathcal{P},n}(\theta) - H(\theta) \right| < \delta \right\}.$$

According to Proposition 7, we know that

$$\mathbb{P}_n \mathcal{H}_{1,n}(\delta)^{\complement} = o(1) \quad \text{and} \quad \mathbb{P}_n \mathcal{H}_{2,n}(\delta)^{\complement} = o(1).$$

Define

$$\mathcal{B}_m(\theta, \delta, \xi) \equiv \left\{ \pi_{\mathcal{P}}(\mathbf{x^m}|\theta) > e^{\xi m} \int_{\Omega_{(-1)}(\theta,\delta)^{\complement}} \pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)}) \pi_{\mathcal{P}}(\mathbf{x^m}|\theta_{(1)}, \theta'_{(-1)}) \mathrm{d}\theta'_{(-1)} \right\}. \tag{177}$$

According to Proposition 16, for any $\delta_0 \in (0, \delta)$, there exists $\xi > 0$ such that

$$\sup_{\theta\in\Omega(\theta_0,\delta_0)} \mathcal{P}_{\theta,m} \mathcal{B}_m(\theta, \delta, \xi)^{\complement} = O(e^{-\xi m}).$$

Define

$$\mathcal{C}_m(\theta,\delta,\xi) \equiv \left\{ \pi_{\mathcal{P}}(\mathbf{x}^{\mathbf{m}}|\theta) < e^{\xi m} \int_{\Omega(\theta,\delta)} \pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)}) \pi_{\mathcal{P}}(\mathbf{x}^{\mathbf{m}}|\theta_{(1)},\theta'_{(-1)}) \mathrm{d}\theta'_{(-1)} \right\}. \tag{178}$$

According to Proposition 19, for any $\delta_0 \in (0,\delta)$, there exists $\xi > 0$ such that

$$\sup_{\theta\in\Omega(\theta_0,\delta_0)} \mathcal{P}_{\theta,m}\mathcal{C}_m(\theta,\delta_0,\xi)^{\complement} = o\left(\frac{1}{n}\right).$$

We define

$$\mathcal{L}_n(\delta,\eta) \equiv \left\{ s_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} s_{\mathcal{P},n}(\theta) < \eta, \quad \forall\, \theta \in \Omega(\theta_0,\delta) \right\}. \tag{179}$$

According to Proposition 13, we know that for any $\eta > 0$, there exists $\delta > 0$ such that

$$\mathbb{P}_n \mathcal{L}_n(\delta,\eta)^{\complement} = o(1).$$

Step 2: We capture the asymptotically essential component in $R_{n,3}$. For any $\eta \in (0,1/2)$, according to the discussion in Step 1, we know that there exists $\delta,\delta_1 \in (0,\eta)$ such that

$$\mathbb{P}_n \mathcal{I}_{1,n}(\delta,\eta)^{\complement} = o\left(\frac{1}{n}\right) \quad \text{and} \quad \sup_{\theta\in\Omega(\theta,\delta_1)} \mathcal{P}_{\theta,m}\mathcal{I}_{2,m}(\theta,\delta,\eta)^{\complement} = o\left(\frac{1}{m}\right).$$

For the given δ above, we know that there exist positive constants $\delta_2 < \delta$, ξ_1 and ξ_2 such that

$$\sup_{\theta\in\Omega(\theta_0,\delta_2)} \mathcal{P}_{\theta,m}\mathcal{B}_m(\theta,\delta,\xi_1/\varrho)^{\complement} \le e^{-\xi_2 m} \quad \text{and} \quad \sup_{\theta\in\Omega(\theta_0,\delta_2)} \mathcal{P}_{\theta,m}\mathcal{C}_m(\theta,\delta,\xi_1/8)^{\complement} = o\left(\frac{1}{m}\right)$$

where $\mathcal{B}_m(\theta,\delta,\xi_1)$ and $\mathcal{C}_m(\theta,\delta,\xi_1)$ are defined in (177) and (178), respectively. Because $H_{\mathcal{P}}(\theta)$ is continuous in θ, then there exists $\delta_3 > 0$ such that

$$\sup_{\theta\in\Omega(\theta_0,\delta_3)} |H_{\mathcal{P}}(\theta) - H_{\mathcal{P}}(\theta_0)| < \xi_1/8.$$

For the given δ, according to Proposition 13, we know that there exists $\delta_4 > 0$ such that

$$\mathbb{P}_n \mathcal{L}_n(\delta_0, \frac{1}{2}\underline{\lambda}\delta^2\eta)^{\complement} = o(1).$$

We choose $\delta_0 \equiv \min\{\delta_1,\delta_2,\delta_3,\delta_4\}$. According to Proposition 16, there exist $\xi_0 > 0$ such that $\mathbb{P}_n \mathcal{A}_n(\delta_0,\xi_0)^{\complement} \le e^{-\xi_0 n}$. By restricting on the event

$$\mathcal{M}_n \equiv \mathcal{A}_n(\theta_0,\delta_0,\xi_0) \cap \mathcal{H}_{1,n}(\xi_1/8) \cap \mathcal{H}_{2,n}(\xi_1/8) \cap \mathcal{I}_{1,n}(\delta,\eta) \cap \mathcal{L}_n(\delta_0, \frac{1}{2}\underline{\lambda}\delta^2\eta)$$

and then focusing on the event, for a given $\theta \in \Omega(\theta_0, \delta_0)$,

$$\mathcal{N}_m(\theta) \equiv \mathcal{B}_m(\theta, \delta, \xi_1) \cap \mathcal{I}_{2,m}(\theta, \delta, \eta) \cap \mathcal{C}_m(\theta, \delta, \xi_1).$$

We show that the following term, denoted as $R^e_{n,3}$, is the asymptotically essential term of $R_{n,3}$

$$\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \ln \frac{\int_{\Omega_{(-1)}(\theta,\delta)} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta_{(1)}, \theta'_{(-1)})\pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta)} \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta. \tag{180}$$

That is, there exists a function $a(\eta)$ with $\lim_{\eta\to 0} a(\eta) = 0$ such that on the event $\mathcal{M}_n$

$$|R_{n,3} - R^e_{n,3}| \leq a(\eta) + o_p(1).$$

We consider the decomposition

$$\begin{aligned}
&\frac{\int \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta_{(1)}, \theta'_{(-1)})\pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta)} \\
&= \frac{\left(\int_{\Omega_{(-1)}(\theta,\delta)} + \int_{\Omega_{(-1)}(\theta,\delta)^{\complement}}\right) \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta_{(1)}, \theta'_{(-1)})\pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta)}
\end{aligned} \tag{181}$$

On the event $\mathcal{M}_n$ and $\mathcal{N}_m(\theta)$, we know that the second term in the log term of (181) can be upper bounded by

$$\begin{aligned}
\frac{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_{(1)}, \theta'_{(-1)})}{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_{(1)}, \theta_{(-1)})} &= e^{-n\left[\widehat{H}_{\mathcal{P},n}(\theta_{(1)},\theta'_{(-1)}) - \widehat{H}_{\mathcal{P},n}(\theta_{(1)},\theta_{(-1)})\right]} \leq e^{-n\left[H_{\mathcal{P}}(\theta_{(1)},\theta'_{(-1)}) - H_{\mathcal{P}}(\theta_{(1)},\theta_{(-1)})\right] + n\xi_1/4} \\
&\leq e^{-n\left[H_{\mathcal{P}}(\theta_{(1)},\theta'_{(-1)}) - H_{\mathcal{P}}(\theta_{0,(1)},\theta_{0,(-1)})\right] + n\xi_1 3/8} \leq e^{n\xi_1 3/8}.
\end{aligned}$$

Thus, we have on the event $\mathcal{M}_n$ and $\mathcal{N}_m(\theta)$

$$\begin{aligned}
&\frac{\int_{\Omega_{(-1)}(\theta,\delta)^{\complement}} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta_{(1)}, \theta'_{(-1)})\pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta)} \\
&\leq e^{n\xi_1 3/8} \frac{\int_{\Omega_{(-1)}(\theta,\delta)^{\complement}} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta_{(1)}, \theta'_{(-1)})\pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta_{(1)}, \theta_{(-1)})} \leq e^{-n\xi_1 5/8}.
\end{aligned}$$

On the other hand, we have on the event $\mathcal{M}_n$ and $\mathcal{N}_m(\theta)$

$$\frac{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_{(1)}, \theta'_{(-1)})}{\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_{(1)}, \theta_{(-1)})} = e^{-n\left[\widehat{H}_{\mathcal{P},n}(\theta_{(1)},\theta'_{(-1)}) - \widehat{H}_{\mathcal{P},n}(\theta_{(1)},\theta_{(-1)})\right]} \geq e^{-n\left[H_{\mathcal{P}}(\theta_{(1)},\theta'_{(-1)}) - H_{\mathcal{P}}(\theta_{(1)},\theta_{(-1)})\right] - n\xi_1/4} \geq e^{-n\xi_1 3/8}.$$

Thus, we have on the event $\mathcal{M}_n$ and $\mathcal{N}_m(\theta)$

$$\begin{aligned}
&\frac{\int_{\Omega_{(-1)}(\theta,\delta)} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta'_{(-1)})\pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta)} \\
&\geq e^{-n\xi_1 3/8}\frac{\int_{\Omega_{(-1)}(\theta,\delta)} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta_{(1)}, \theta'_{(-1)})\pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta_{(1)}, \theta_{(-1)})} \geq e^{-n\xi_1/2}.
\end{aligned}$$

Therefore, on the event $\mathcal{M}_n$ and $\mathcal{N}_m(\theta)$, for the positive constant $\eta > 0$, we know that when n is large enough, it holds that

$$\frac{\int \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta'_{(-1)})\pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta)} \leq (1+\eta)\frac{\int_{\Omega_{(-1)}(\theta,\delta)} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta'_{(-1)})\pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta)}$$

Because $H_{\mathcal{P}}(\theta)$ is continuous on Θ, we can define

$$M_H \equiv \sup_{\theta\in\Theta} H_{\mathcal{P}}(\theta) - \inf_{\theta\in\Theta} H_{\mathcal{P}}(\theta). \tag{182}$$

Then, on the event $\mathcal{M}_n$, we have $|R_{n,3} - R^e_{n,3}|$ is upper bounded by

$$\begin{aligned}
&\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}}) \int_{\mathcal{N}_m(\theta)^{\complement}} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left|\ln \frac{\int_{\Omega_{(-1)}(\theta,\delta)} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta'_{(-1)})\pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta)}\right| \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta \\
&\qquad + \ln(1+\eta) + \pi_{\mathcal{P}}(\Omega(\theta_0,\delta_0)^{\complement}|\mathbf{x}^{\mathbf{n}})(m+n)M_H.
\end{aligned}$$

The first term in the long expression above is upper bounded by

$$\begin{aligned}
&(m+n)M_H \sup_{\theta\in\Omega(\theta_0,\delta_0)} \mathbb{P}_{\theta,m}\mathcal{N}_m(\theta)^{\complement} \\
&\leq (m+n)M_H \left[\sup_{\theta\in\Omega(\theta_0,\delta_0)} \mathbb{P}_{\theta,m}\mathcal{B}_m(\theta,\delta,\xi_1/\varrho)^{\complement} + \sup_{\theta\in\Omega(\theta_0,\delta_0)} \mathbb{P}_{\theta,m}\mathcal{I}_{2,m}(\theta,\delta,\eta)^{\complement} + \sup_{\theta\in\Omega(\theta_0,\delta_0)} \mathbb{P}_{\theta,m}\mathcal{C}_m(\theta,\delta,\xi_1/8)^{\complement}\right] \\
&= o(1).
\end{aligned}$$

On the set $\mathcal{M}_n$, we know that

$$\pi_{\mathcal{P}}(\Omega(\theta_0,\delta_0)^{\complement}|\mathbf{x}^{\mathbf{n}})(m+n)M_H = O\left(ne^{-\xi_0 n}\right). \tag{183}$$

Therefore, by the arbitrariness of positive constant η, we know that the asymptotically essential component of $R_{n,3}$ is $R^e_{n,3}$ with expression

$$\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \ln \frac{\int_{\Omega_{(-1)}(\theta,\delta)} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta'_{(-1)})\pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta)}\mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta$$

Step 3: We first show that the term $R^*_{n,3}$ is upper-bounded, asymptotically, by zero.

Step 3.1: We find the upper bound for the log term in the expression of $R^e_{n,3}$ when $\mathbf{x^n} \in \mathcal{M}_n$, $\theta \in \Omega(\theta_0, \delta_0)$ and $\tilde{\mathbf{x}}^{\mathbf{m}} \in \mathcal{N}_m(\theta)$. By Taylor's expansion, we have

$$\begin{aligned}
&\ln \int_{\Omega_{(-1)}(\theta,\delta)} \frac{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n} | \theta_{(1)}, \theta'_{(-1)})}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n} | \theta_{(1)}, \theta_{(-1)})} \pi_{\mathcal{P}}(\theta'_{(-1)} | \theta_{(1)}) \mathrm{d}\theta'_{(-1)} \\
&= \ln \int_{\Omega_{(-1)}(\theta,\delta)} e^{(\theta'-\theta)^T [n s_{\mathcal{P},n}(\theta) + m s_{\mathcal{P},m}(\theta)] - \frac{1}{2}(\theta'-\theta)^T [n \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}) + m \widehat{\mathbf{I}}_{\mathcal{P},m}(\tilde{\theta})](\theta'-\theta)} \pi_{\mathcal{P}}(\theta'_{(-1)} | \theta_{(1)}) \mathrm{d}\theta'_{(-1)},
\end{aligned}$$

where $\tilde{\theta}$ is between θ' and θ, and

$$\theta' \equiv \begin{pmatrix} \theta_{(1)} \\ \theta'_{(-1)} \end{pmatrix}, \tag{184}$$

and

$$s_{\mathcal{P},n}(\theta) \equiv \frac{1}{n} \sum_{t=1}^{n} \nabla \ln \pi_{\mathcal{P}}(\mathbf{x_t}; \theta), \quad \text{and} \quad s_{\mathcal{P},m}(\theta) \equiv \frac{1}{m} \sum_{t=1}^{m} \nabla \ln \pi_{\mathcal{P}}(\tilde{\mathbf{x}}_{\mathbf{t}}; \theta),$$

and

$$\widehat{\mathbf{I}}_{\mathcal{P},n}(\theta) \equiv -\frac{1}{n} \sum_{t=1}^{n} \nabla^2 \ln \pi_{\mathcal{P}}(\mathbf{x_t}; \theta), \quad \text{and} \quad \widehat{\mathbf{I}}_{\mathcal{P},m}(\theta) \equiv -\frac{1}{m} \sum_{t=1}^{m} \nabla^2 \ln \pi_{\mathcal{P}}(\tilde{\mathbf{x}}_{\mathbf{t}}; \theta).$$

Let's define

$$\rho(\delta) \equiv \sup_{\theta \in \Omega(\theta_0, 2\delta)} \left| \ln \frac{\pi_{\mathcal{P}}(\theta_{(-1)} | \theta_{(1)})}{\pi_{\mathcal{P}}(\theta_{0,(-1)} | \theta_{0,(1)})} \right|. \tag{185}$$

Because $\theta \in \Omega(\theta_0, \delta_0)$ and $\theta'_{(-1)} \in \Omega_{(-1)}(\theta, \delta)$ imply that $\theta' \in \Omega(\theta_0, 2\delta)$, we know that

$$\begin{aligned}
&\ln \int_{\Omega_{(-1)}(\theta,\delta)} \frac{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n} | \theta_{(1)}, \theta'_{(-1)})}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n} | \theta_{(1)}, \theta_{(-1)})} \pi_{\mathcal{P}}(\theta'_{(-1)} | \theta_{(1)}) \mathrm{d}\theta'_{(-1)} \\
&\leq \ln \int_{\Omega_{(-1)}(\theta,\delta)} e^{(\theta'-\theta)^T [n s_{\mathcal{P},n}(\theta) + m s_{\mathcal{P},m}(\theta)] - \frac{1}{2}(\theta'-\theta)^T [n \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}) + m \widehat{\mathbf{I}}_{\mathcal{P},m}(\tilde{\theta})](\theta'-\theta)} \mathrm{d}\theta'_{(-1)} + \ln \pi_{\mathcal{P}}(\theta_{0,(-1)} | \theta_{0,(1)}) + \rho(\delta).
\end{aligned}$$

It is obvious that $\rho(\cdot)$ is increasing a univariate increasing function. Then, $\rho(\delta) \leq \rho(\eta)$ since $\delta < \eta$. Thus, we have

$$\begin{aligned}
&\ln \int_{\Omega_{(-1)}(\theta,\delta)} \frac{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n} | \theta_{(1)}, \theta'_{(-1)})}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n} | \theta_{(1)}, \theta_{(-1)})} \pi_{\mathcal{P}}(\theta'_{(-1)} | \theta_{(1)}) \mathrm{d}\theta'_{(-1)} \\
&\leq \ln \int_{\Omega_{(-1)}(\theta,\delta)} e^{(\theta'-\theta)^T [n s_{\mathcal{P},n}(\theta) + m s_{\mathcal{P},m}(\theta)] - \frac{1}{2}(\theta'-\theta)^T [n \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta}) + m \widehat{\mathbf{I}}_{\mathcal{P},m}(\tilde{\theta})](\theta'-\theta)} \mathrm{d}\theta'_{(-1)} + \ln \pi_{\mathcal{P}}(\theta_{0,(-1)} | \theta_{0,(1)}) + \rho(\eta),
\end{aligned}$$

where the function $\rho(\eta)$ is defined in (185). On the event $\mathcal{I}_{1,n}(\delta, \eta)$, we have for all $\theta \in \Omega(\theta_0, \delta_0)$

$$(\theta' - \theta)^T \widehat{\mathbf{I}}_{\mathcal{P},n}(\theta' - \theta) \geq (1 - \eta)(\theta' - \theta)^T \mathbf{I}_{\mathcal{P}}(\theta)(\theta' - \theta).$$

On the event $\mathcal{I}_{2,m}(\theta,\delta,\eta)$, we have

$$(\theta'-\theta)^T\widehat{\mathbf{I}}_{\mathcal{P},m}(\theta'-\theta) \geq (1-\eta)(\theta'-\theta)^T\mathbf{I}_{\mathcal{P}}(\theta)(\theta'-\theta).$$

Thus, it follows that

$$\begin{aligned}
&\ln\int_{\Omega_{(-1)}(\theta,\delta)}\frac{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x}^{\mathbf{n}}|\theta_{(1)},\theta'_{(-1)})}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x}^{\mathbf{n}}|\theta_{(1)},\theta_{(-1)})}\pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}\\
&\leq \ln\int_{\Omega_{(-1)}(\theta,\delta)} e^{(\theta'-\theta)^T[ns_{\mathcal{P},n}(\theta)+ms_{\mathcal{P},m}(\theta)]-\frac{1}{2}(1-\eta)(\theta'-\theta)^T[n\mathbf{I}_{\mathcal{P}}(\theta)+m\mathbf{I}_{\mathcal{P}}(\theta)](\theta'-\theta)}\mathrm{d}\theta'_{(-1)}+\ln\pi_{\mathcal{P}}(\theta_{0,(-1)}|\theta_{0,(1)})+\rho(\eta),
\end{aligned}$$

where the function $\rho(\eta)$ is defined in (185).

Denote

$$\alpha = \frac{n}{n+m}.$$

Let's consider the following identities

$$\begin{aligned}
&(\theta'-\theta)^T[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)]-\frac{1}{2}(1-\eta)(\theta'-\theta)^T\mathbf{I}_{\mathcal{P}}(\theta)(\theta'-\theta)\\
&=-\frac{1-\eta}{2}(\theta'-u)^T\mathbf{I}_{\mathcal{P}}(\theta)(\theta-u)+\frac{1}{2(1-\eta)}[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)]^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)]
\end{aligned}$$

where

$$u \equiv \theta + \frac{1}{1-\eta}\mathbf{I}_{\mathcal{P}}(\theta)^{-1}[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)]. \tag{186}$$

Therefore, we have

$$\begin{aligned}
&\ln\int_{\Omega_{(-1)}(\theta,\delta)}\frac{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x}^{\mathbf{n}}|\theta_{(1)},\theta'_{(-1)})}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x}^{\mathbf{n}}|\theta_{(1)},\theta_{(-1)})}\pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}+\ln\pi_{\mathcal{P}}(\theta_{0,(-1)}|\theta_{0,(1)})+\rho(\eta)\\
&\leq \ln\int_{\Omega_{(-1)}(\theta,\delta)} e^{-\frac{m+n}{2}(1-\eta)(\theta'-u)^T\mathbf{I}_{\mathcal{P}}(\theta)(\theta'-u)}\mathrm{d}\theta'_{(-1)}+\frac{m+n}{2(1-\eta)}[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)]^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)],
\end{aligned}$$

where the function $\rho(\eta)$ is defined in (185). Further, we have

$$\begin{aligned}
&\int_{\Omega_{(-1)}(\theta,\delta)} e^{-\frac{m+n}{2}(1-\eta)(\theta'-u)^T\mathbf{I}_{\mathcal{P}}(\theta)(\theta'-u)}\mathrm{d}\theta'_{(-1)} \leq \int_{\mathbb{R}^{d-1}} e^{-\frac{m+n}{2}(1-\eta)(\theta'-u)^T\mathbf{I}_{\mathcal{P}}(\theta)(\theta'-u)}\mathrm{d}\theta'_{(-1)}\\
&=\frac{(2\pi)^{D_\Theta/2}|(1-\eta)^{-1}\mathbf{I}_{\mathcal{P}}(\theta)^{-1}|^{1/2}}{(m+n)^{D_\Theta/2}}\int_{\mathbb{R}^{D_\Theta-1}}\frac{(m+n)^{D_\Theta/2}}{(2\pi)^{D_\Theta/2}|(1-\eta)^{-1}\mathbf{I}_{\mathcal{P}}(\theta)^{-1}|^{1/2}}e^{-\frac{m+n}{2}(1-\eta)(\theta'-u)^T\mathbf{I}_{\mathcal{P}}(\theta)(\theta'-u)}\mathrm{d}\theta'_{(-1)}
\end{aligned}$$

Thus, we can obtain

$$\begin{aligned}
&\int_{\Omega_{(-1)}(\theta,\delta)} e^{-\frac{m+n}{2}(1-\eta)(\theta'-u)^T\mathbf{I}_{\mathcal{P}}(\theta)(\theta'-u)}\mathrm{d}\theta'_{(-1)} \\
&\le \frac{(2\pi)^{D_\Theta/2}\left[\det\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\right]^{1/2}}{(1-\eta)^{D_\Theta/2}(m+n)^{D_\Theta/2}} \frac{(m+n)^{1/2}}{(2\pi)^{1/2}|(1-\eta)^{-1}\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}|^{1/2}} e^{-\frac{m+n}{2}(1-\eta)\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}} \\
&= \left(\frac{2\pi}{m+n}\right)^{\frac{D_\Theta-1}{2}} \frac{\left[\det\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\right]^{1/2}}{|\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}|^{1/2}} e^{-\frac{m+n}{2}(1-\eta)\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}}(1-\eta)^{-\frac{D_\Theta-1}{2}}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\ln\int_{\Omega_{(-1)}(\theta,\delta)} \frac{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x}^{\mathbf{n}}|\theta_{(1)},\theta'_{(-1)})}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x}^{\mathbf{n}}|\theta_{(1)},\theta_{(-1)})}\pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)} \\
&\quad\le \frac{D_\Theta-1}{2}\ln\left(\frac{2\pi}{m+n}\right)+\frac{1}{2}\ln\frac{\det\left[\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\right]}{|\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}|}-\frac{m+n}{2}(1-\eta)\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}} \\
&\qquad+\frac{m+n}{2(1-\eta)}[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)]^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)] \qquad (187) \\
&\qquad+\ln\pi_{\mathcal{P}}(\theta_{0,(-1)}|\theta_{0,(1)})+\rho(\eta)-\frac{D_\Theta-1}{2}\ln(1-\eta), \qquad (188)
\end{aligned}$$

where $\rho(\eta)$ is the function defined in (185).

Step 3.2: We find the upper bound for the asymptotically essential term of $R^e_{n,3}$. We take integrations over θ and $\tilde{\mathbf{x}}^{\mathbf{m}}$ in (180) over each term on the right hand side of the inequality (187).

$$\int_{\Omega(\theta_0,\delta_0)}\pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}})\int_{\mathcal{N}_m(\theta)}\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta)\frac{D_\Theta-1}{2}\ln\left(\frac{2\pi}{m+n}\right)\mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta\le\frac{D_\Theta-1}{2}\ln\left(\frac{2\pi}{m+n}\right)$$

and

$$\int_{\Omega(\theta_0,\delta_0)}\pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}})\int_{\mathcal{N}_m(\theta)}\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta)\frac{1}{2}\ln\frac{\det\left[\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\right]}{|\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}|}\mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta\le\frac{1}{2}\ln\frac{\det\left[\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\right]}{|\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}|}+\eta$$

and

$$
\begin{aligned}
&\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2}\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}\right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta \\
&\quad= \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{X}^m} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2}\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}\right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta \\
&\qquad- \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)^{\complement}} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2}\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}\right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta \\
&\quad= \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{X}^m} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2}\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}\right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta + o(1),
\end{aligned}
$$

and recall the definition of θ' in (184) which implies that $\theta'_{(1)} \equiv \theta_{(1)}$ and remember the definition of u in (186), we have

$$
\begin{aligned}
&\int_{\mathcal{X}^m} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2}\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}\right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \\
&= -\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\left\{\int_{\mathcal{X}^m} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta)\left[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)\right]\left[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)\right]^T \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\right\}\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v} \\
&\qquad \times\left[2(m+n)^{-1}(1-\eta)^2\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}\right]^{-1} \\
&= -\frac{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\left\{\alpha^2 s_{\mathcal{P},n}(\theta)s_{\mathcal{P},n}(\theta)^T+(1-\alpha)^2\mathbf{I}_{\mathcal{P}}(\theta)/m\right\}\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}{2(m+n)^{-1}(1-\eta)^2\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}} \\
&= -\frac{\alpha}{(1-\eta)^2}\frac{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}S_{\mathcal{P},n}(\theta)S_{\mathcal{P},n}(\theta)^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}{2\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}} - \frac{1-\alpha}{2(1-\eta)^2},
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2}(1-\eta)\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}\right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta \\
&= -\frac{\alpha}{2(1-\eta)}\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathcal{P}}(\theta|\mathbf{x^n})\frac{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}S_{\mathcal{P},n}(\theta)S_{\mathcal{P},n}(\theta)^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}\mathrm{d}\theta - \frac{1-\alpha}{2(1-\eta)} + o(1) \\
&\leq -\frac{\alpha}{2} - \frac{1-\alpha}{2(1-\eta)} + o(1) \leq -\frac{1}{2} + o(1),
\end{aligned}
$$

and

$$\int_{\Omega(\theta_0,\delta_0)}\pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}})\int_{\mathcal{N}_m(\theta)}\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta)\frac{m+n}{2(1-\eta)}[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)]^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)]\mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta$$

$$=\frac{\alpha}{2(1-\eta)}\int_{\Omega(\theta_0,\delta_0)}\pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}})S_{\mathcal{P},n}(\theta)^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}S_{\mathcal{P},n}(\theta)\mathrm{d}\theta+\frac{1-\alpha}{2(1-\eta)}D_\Theta+o(1)$$

$$\leq\frac{\alpha}{2(1-\eta)}\int_{\Omega(\theta_0,\delta_0)}\pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}})S_{\mathcal{P},n}(\theta)^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}S_{\mathcal{P},n}(\theta)\mathrm{d}\theta+\frac{1-\alpha}{2}D_\Theta+D_\Theta\eta+o(1)$$

$$\leq\frac{\alpha(D_\Theta+\eta)}{2(1-\eta)}+\frac{1-\alpha}{2}D_\Theta+D_\Theta\eta+o(1)$$

$$\leq\frac{D_\Theta}{2}+(\alpha+D_\Theta+\alpha D_\Theta)\eta+o(1).$$

and

$$\int_{\Omega(\theta_0,\delta_0)}\pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}})\int_{\mathcal{N}_m(\theta)}\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta)\left[\ln\pi_{\mathcal{P}}(\theta_{0,(-1)}|\theta_{0,(1)})+\rho(\eta)-\frac{D_\Theta-1}{2}\ln(1-\eta)\right]\mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta$$
$$=\ln\pi_{\mathcal{P}}(\theta_{0,(-1)}|\theta_{0,(1)})+\rho(\eta)-\frac{D_\Theta-1}{2}\ln(1-\eta)+o(1),$$

where $\rho(\eta)$ is defined in (185). Therefore, we know that the asymptotically essential component (180) of the term $R_{n,3}$ is upper bounded by

$$\frac{D_\Theta-1}{2}\ln\left(\frac{2\pi}{m+n}\right)+\frac{1}{2}\ln\frac{\det\left[\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\right]}{|\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}|}+\frac{D_\Theta-1}{2}+\ln\pi_{\mathcal{P}}(\theta_{0,(-1)}|\theta_{0,(1)})$$
$$+\rho(\eta)+(1+\alpha+D_\Theta+\alpha D_\Theta)\eta-\frac{D_\Theta-1}{2}\ln(1-\eta)+o(1),$$

where the function $\rho(\eta)$ is defined in (185). Because of the fact that $\lim_{x\to 0}\rho(x)=0$ and the arbitrariness of η, we know that $R^*_{n,3}$ defined in (173) is asymptotically upper bounded by zero.

Step 4: We then show that the difference $R^*_{n,3}$ is lower-bounded, asymptotically, by zero.

Step 4.1: We find the lower bound for the log term in the expression of $R^e_{n,3}$ when $\mathbf{x}^{\mathbf{n}}\in\mathcal{M}_n(\theta_0)$, $\theta\in\Omega(\theta_0,\delta_0)$ and $\tilde{\mathbf{x}}^{\mathbf{m}}\in\mathcal{N}_m(\theta)$. By Taylor's expansion, we have

$$\ln\int_{\Omega_{(-1)}(\theta,\delta)}\frac{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x}^{\mathbf{n}}|\theta_{(1)},\theta'_{(-1)})}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x}^{\mathbf{n}}|\theta_{(1)},\theta_{(-1)})}\pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}$$
$$=\ln\int_{\Omega_{(-1)}(\theta,\delta)}e^{(\theta'-\theta)^T[ns_{\mathcal{P},n}(\theta)+ms_{\mathcal{P},m}(\theta)]-\frac{1}{2}(\theta'-\theta)^T[n\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})+m\widehat{\mathbf{I}}_{\mathcal{P},m}(\tilde{\theta})](\theta'-\theta)}\pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)},$$

where $\tilde{\theta}$ is between θ' and θ, and

$$\theta' \equiv \left(\begin{array}{c} \theta_{(1)} \\ \theta'_{(-1)} \end{array} \right).$$

Because $\theta \in \Omega(\theta_0, \delta_0)$ and $\theta'_{(-1)} \in \Omega_{(-1)}(\theta, \delta)$ imply that $\theta' \in \Omega(\theta_0, 2\delta)$, we know that

$$\begin{aligned}
&\ln \int_{\Omega_{(-1)}(\theta,\delta)} \frac{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta'_{(-1)})}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta_{(-1)})} \pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)}) \mathrm{d}\theta'_{(-1)} \\
&\geq \ln \int_{\Omega_{(-1)}(\theta,\delta)} e^{(\theta'-\theta)^T[ns_{\mathcal{P},n}(\theta)+ms_{\mathcal{P},m}(\theta)]-\frac{1}{2}(\theta'-\theta)^T[n\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})+m\widehat{\mathbf{I}}_{\mathcal{P},m}(\tilde{\theta})](\theta'-\theta)} \mathrm{d}\theta'_{(-1)} + \ln \pi_{\mathcal{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) - \rho(\delta),
\end{aligned}$$

where the function $\rho(\eta)$ is defined in (185). It is obvious that $\rho(\cdot)$ is increasing a univariate increasing function. Then,

$$\rho(\delta) \leq \rho(\eta) \quad \text{since} \quad \delta < \eta.$$

Thus, we have

$$\begin{aligned}
&\ln \int_{\Omega_{(-1)}(\theta,\delta)} \frac{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta'_{(-1)})}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta_{(-1)})} \pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)}) \mathrm{d}\theta'_{(-1)} \\
&\geq \ln \int_{\Omega_{(-1)}(\theta,\delta)} e^{(\theta'-\theta)^T[ns_{\mathcal{P},n}(\theta)+ms_{\mathcal{P},m}(\theta)]-\frac{1}{2}(\theta'-\theta)^T[n\widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})+m\widehat{\mathbf{I}}_{\mathcal{P},m}(\tilde{\theta})](\theta'-\theta)} \mathrm{d}\theta'_{(-1)} + \ln \pi_{\mathcal{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) - \rho(\eta),
\end{aligned}$$

where the function $\rho(\eta)$ is defined in (185). On the event $\mathcal{I}_{1,n}(\delta, \eta)$, we have for all $\theta \in \Omega(\theta_0, \delta_0)$

$$(\theta' - \theta)^T \widehat{\mathbf{I}}_{\mathcal{P},n}(\tilde{\theta})(\theta' - \theta) \leq (1 + \eta)(\theta' - \theta)^T \mathbf{I}_{\mathcal{P}}(\theta)(\theta' - \theta).$$

On the event $\mathcal{I}_{2,m}(\theta, \delta, \eta)$, we have

$$(\theta' - \theta)^T \widehat{\mathbf{I}}_{\mathcal{P},m}(\tilde{\theta})(\theta' - \theta) \leq (1 + \eta)(\theta' - \theta)^T \mathbf{I}_{\mathcal{P}}(\theta)(\theta' - \theta).$$

Thus, it follows that

$$\begin{aligned}
&\ln \int_{\Omega_{(-1)}(\theta,\delta)} \frac{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta'_{(-1)})}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta_{(-1)})} \pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)}) \mathrm{d}\theta'_{(-1)} \\
&\geq \ln \int_{\Omega_{(-1)}(\theta,\delta)} e^{(\theta'-\theta)^T[ns_{\mathcal{P},n}(\theta)+ms_{\mathcal{P},m}(\theta)]-\frac{1}{2}(1+\eta)(\theta'-\theta)^T[n\mathbf{I}_{\mathcal{P}}(\theta)+m\mathbf{I}_{\mathcal{P}}(\theta)](\theta'-\theta)} \mathrm{d}\theta'_{(-1)} + \ln \pi_{\mathcal{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) - \rho(\eta),
\end{aligned}$$

where the function $\rho(\eta)$ is defined in (185). Denote

$$\alpha = \frac{n}{n+m}.$$

Let's consider the following identities

$$(\theta'-\theta)^T[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)]-\frac{1}{2}(1+\eta)(\theta'-\theta)^T\mathbf{I}_{\mathcal{P}}(\theta)(\theta'-\theta)$$
$$=-\frac{1+\eta}{2}(\theta'-v)^T\mathbf{I}_{\mathcal{P}}(\theta)(\theta-v)+\frac{1}{2(1+\eta)}[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)]^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)]$$

where

$$v\equiv\theta+\frac{1}{1+\eta}\mathbf{I}_{\mathcal{P}}(\theta)^{-1}[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)].$$

Therefore, we have

$$\ln\int_{\Omega_{(-1)}(\theta,\delta)}\frac{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x}^{\mathbf{n}}|\theta_{(1)},\theta'_{(-1)})}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x}^{\mathbf{n}}|\theta_{(1)},\theta_{(-1)})}\pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}$$
$$\geq\ln\int_{\Omega_{(-1)}(\theta,\delta)}e^{-\frac{m+n}{2}(1+\eta)(\theta'-v)^T\mathbf{I}_{\mathcal{P}}(\theta)(\theta'-v)}\mathrm{d}\theta'_{(-1)}$$
$$+\frac{m+n}{2(1+\eta)}[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)]^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)]+\ln\pi_{\mathcal{P}}(\theta_{0,(-1)}|\theta_{0,(1)})-\rho(\eta),$$

where the function $\rho(\eta)$ is defined in (185). Further, we have

$$\int_{\Omega_{(-1)}(\theta,\delta)}e^{-\frac{m+n}{2}(1+\eta)(\theta'-v)^T\mathbf{I}_{\mathcal{P}}(\theta)(\theta'-v)}\mathrm{d}\theta'_{(-1)}$$
$$=\int_{\mathbb{R}^{D_\Theta-1}}e^{-\frac{m+n}{2}(1+\eta)(\theta'-v)^T\mathbf{I}_{\mathcal{P}}(\theta)(\theta'-v)}\mathrm{d}\theta'_{(-1)}-\int_{\Omega_{(-1)}(\theta,\delta)^\complement}e^{-\frac{m+n}{2}(1+\eta)(\theta'-v)^T\mathbf{I}_{\mathcal{P}}(\theta)(\theta'-v)}\mathrm{d}\theta'_{(-1)}$$

and

$$\int_{\mathbb{R}^{D_\Theta-1}}e^{-\frac{m+n}{2}(1+\eta)(\theta'-v)^T\mathbf{I}_{\mathcal{P}}(\theta)(\theta'-v)}\mathrm{d}\theta'_{(-1)}$$
$$=\frac{(2\pi)^{D_\Theta/2}\left[\det\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\right]^{1/2}}{(1+\eta)^{D_\Theta/2}(m+n)^{D_\Theta/2}}\int_{\mathbb{R}^{D_\Theta-1}}\frac{(m+n)^{D_\Theta/2}(1+\eta)^{D_\Theta/2}}{(2\pi)^{D_\Theta/2}\left[\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\right]^{1/2}}e^{-\frac{m+n}{2}(1+\eta)(\theta'-v)^T\mathbf{I}_{\mathcal{P}}(\theta)(\theta'-v)}\mathrm{d}\theta'_{(-1)}$$
$$=\frac{(2\pi)^{D_\Theta/2}\left[\det\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\right]^{1/2}}{(1+\eta)^{D_\Theta/2}(m+n)^{D_\Theta/2}}\frac{(m+n)^{1/2}}{(2\pi)^{1/2}|(1+\eta)^{-1}\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}|^{1/2}}e^{-\frac{m+n}{2}(1+\eta)\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}}$$
$$=\left(\frac{2\pi}{m+n}\right)^{\frac{D_\Theta-1}{2}}\frac{\left[\det\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\right]^{1/2}}{|\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}|^{1/2}}e^{-\frac{m+n}{2}(1+\eta)\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}}(1+\eta)^{-\frac{D_\Theta-1}{2}}.$$

Using the definition of v, we know that for any θ' such that $\theta'_{(-1)} \in \Omega_{(-1)}(\theta, \delta)^{\complement}$

$$
\begin{aligned}
&(\theta' - v)^T \mathbf{I}_{\mathcal{P}}(\theta)(\theta' - v) \\
&= \left[\theta' - \theta - \frac{1}{1+\eta}\mathbf{I}_{\mathcal{P}}(\theta)^{-1} s_{\mathcal{P},n}(\theta)\right]^T \mathbf{I}_{\mathcal{P}}(\theta) \left[\theta' - \theta - \frac{1}{1+\eta}\mathbf{I}_{\mathcal{P}}(\theta)^{-1} s_{\mathcal{P},n}(\theta)\right] \\
&\geq \frac{1}{2}(\theta' - \theta)^T \mathbf{I}_{\mathcal{P}}(\theta)(\theta' - \theta) - \frac{1}{(1+\eta)^2} s_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} s_{\mathcal{P},n}(\theta) \\
&\geq \frac{\underline{\lambda}\delta^2}{2(1+\eta)^2} - \frac{\underline{\lambda}\delta^2\eta}{2(1+\eta)^2} \quad \text{because of } \ \mathcal{L}_n(\theta_0, \delta_0, \frac{1}{2}\underline{\lambda}\delta^2\eta) \\
&= \frac{\underline{\lambda}\delta^2(1-\eta)}{2(1+\eta)^2}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&- \int_{\Omega_{(-1)}(\theta,\delta)^{\complement}} e^{-\frac{m+n}{2}(1+\eta)(\theta'-v)^T \mathbf{I}_{\mathcal{P}}(\theta)(\theta'-v)} \mathrm{d}\theta'_{(-1)} \\
&\geq - \int_{\Omega_{(-1)}(\theta,\delta)^{\complement}} e^{-\frac{(m+n)(1-\eta)}{8(1+\eta)}\underline{\lambda}\delta^2} e^{-\frac{m+n}{4}(1+\eta)(\theta'-v)^T \mathbf{I}_{\mathcal{P}}(\theta)(\theta'-v)} \mathrm{d}\theta'_{(-1)} \\
&\geq -e^{-\frac{(m+n)(1-\eta)}{8(1+\eta)}\underline{\lambda}\delta^2} \int_{\mathbb{R}^{D_\Theta - 1}} e^{-\frac{m+n}{4}(1+\eta)(\theta'-v)^T \mathbf{I}_{\mathcal{P}}(\theta)(\theta'-v)} \mathrm{d}\theta'_{(-1)} \\
&= -e^{-\frac{(m+n)(1-\eta)}{8(1+\eta)}\underline{\lambda}\delta^2} 2^{\frac{D_\Theta-1}{2}} \left(\frac{2\pi}{m+n}\right)^{\frac{D_\Theta-1}{2}} \frac{\left[\det \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\right]^{1/2}}{|\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}|^{1/2}} e^{-\frac{m+n}{4}(1+\eta)\frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}} (1+\eta)^{-\frac{D_\Theta-1}{2}} \\
&\geq -e^{-\frac{(m+n)(1-2\eta)}{8(1+\eta)}\underline{\lambda}\delta^2} 2^{\frac{D_\Theta-1}{2}} \left(\frac{2\pi}{m+n}\right)^{\frac{D_\Theta-1}{2}} \frac{\left[\det \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\right]^{1/2}}{|\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}|^{1/2}} e^{-\frac{m+n}{2}(1+\eta)\frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}} (1+\eta)^{-\frac{D_\Theta-1}{2}}
\end{aligned}
$$

where the last inequality is due to the fact that

$$
\frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}} = \frac{1}{(1+\eta)^2} \frac{\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} s_{\mathcal{P},n}(\theta) s_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}{\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}.
$$

and

$$
\begin{aligned}
&\frac{\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} s_{\mathcal{P},n}(\theta) s_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}{\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}} \leq \sup_{|\mathbf{u}|=1} \frac{\mathbf{u}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} s_{\mathcal{P},n}(\theta) s_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{u}}{\mathbf{u}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{u}} \\
&= \lambda_m \left[\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2} s_{\mathcal{P},n}(\theta) s_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\right] \leq \operatorname{tr}\left[\mathbf{I}_{\mathcal{P}}(\theta)^{-1/2} s_{\mathcal{P},n}(\theta) s_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1/2}\right] \\
&= s_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} s_{\mathcal{P},n}(\theta) < \frac{\underline{\lambda}\delta^2\eta}{2(1+\eta)^2} \quad \text{because of } \ \mathcal{L}_n(\theta_0, \delta_0, \frac{1}{2}\underline{\lambda}\delta^2\eta).
\end{aligned}
$$

and hence

$$\int_{\Omega_{(-1)}(\theta,\delta)} e^{-\frac{m+n}{2}(1+\eta)(\theta'-v)^T\mathbf{I}_{\mathcal{P}}(\theta)(\theta'-v)}\mathrm{d}\theta'_{(-1)}$$

$$\geq \left[1-e^{-\frac{(m+n)(1-2\eta)}{8(1+\eta)}\underline{\lambda}\delta^2}2^{\frac{D_\Theta-1}{2}}\right]\left(\frac{2\pi}{m+n}\right)^{\frac{D_\Theta-1}{2}}\frac{\left[\det\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\right]^{1/2}}{|\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}|^{1/2}}\times e^{-\frac{m+n}{2}(1+\eta)\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}}(1+\eta)^{-\frac{D_\Theta-1}{2}}$$

$$\geq \left(\frac{2\pi}{m+n}\right)^{\frac{D_\Theta-1}{2}}\frac{\left[\det\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\right]^{1/2}}{|\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}|^{1/2}}e^{-\frac{m+n}{2}(1+\eta)\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}}(1+\eta)^{-\frac{D_\Theta}{2}}, \quad \text{for large } n,m.$$

Therefore,

$$\begin{aligned}
&\ln\int_{\Omega_{(-1)}(\theta,\delta)}\frac{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x}^{\mathbf{n}}|\theta_{(1)},\theta'_{(-1)})}{\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x}^{\mathbf{n}}|\theta_{(1)},\theta_{(-1)})}\pi_{\mathcal{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}\\
&\geq \frac{D_\Theta-1}{2}\ln\left(\frac{2\pi}{m+n}\right)+\frac{1}{2}\ln\frac{|\mathbf{I}_{\mathcal{P}}(\theta)^{-1}|}{|\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}|}\\
&\quad -\frac{m+n}{2}(1+\eta)\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}\\
&\quad +\frac{m+n}{2(1+\eta)}[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)]^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}[\alpha s_{\mathcal{P},n}(\theta)+(1-\alpha)s_{\mathcal{P},m}(\theta)]\\
&\quad +\ln\pi_{\mathcal{P}}(\theta_{0,(-1)}|\theta_{0,(1)})-\rho(\eta)-\frac{D_\Theta}{2}\ln(1+\eta),
\end{aligned}\tag{189}$$

where the function $\rho(\eta)$ is defined in (185).

Step 4.2: We find the lower bound for the asymptotically essential term $R^e_{n,3}$. We take integrations over θ and $\tilde{\mathbf{x}}^{\mathbf{m}}$ over each term on the right hand side of the inequality (189).

$$\begin{aligned}
&\int_{\Omega(\theta_0,\delta_0)}\pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}})\int_{\mathcal{N}_m(\theta)}\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta)\frac{D_\Theta-1}{2}\ln\left(\frac{2\pi}{m+n}\right)\mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta\\
&\quad\geq\frac{D_\Theta-1}{2}\ln\left(\frac{2\pi}{m+n}\right)\pi_{\mathcal{P}}\left(\Omega(\theta_0,\delta_0)|\mathbf{x}^{\mathbf{n}}\right)\left[1-\sup_{\theta\in\Omega(\theta_0,\delta_0)}\mathcal{P}^m_\theta\mathcal{N}_m(\theta)\right]\\
&\quad=\frac{D_\Theta-1}{2}\ln\left(\frac{2\pi}{m+n}\right)+o(1),\quad\text{due to }\mathcal{A}_n(\theta_0,\delta_0,\xi_0).
\end{aligned}$$

and

$$\begin{aligned}
&\int_{\Omega(\theta_0,\delta_0)}\pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}})\int_{\mathcal{N}_m(\theta)}\pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta)\frac{1}{2}\ln\frac{\det\left[\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\right]}{|\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}|}\mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta\\
&\geq\left[\frac{1}{2}\ln\frac{\det\left[\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\right]}{|\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}|}-\eta\right]\pi_{\mathcal{P}}\left(\Omega(\theta_0,\delta_0)|\mathbf{x}^{\mathbf{n}}\right)\left[1-\sup_{\theta\in\Omega(\theta_0,\delta_0)}\mathcal{P}^m_\theta\mathcal{N}_m(\theta)\right]\\
&=\frac{1}{2}\ln\frac{\det\left[\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\right]}{|\mathbf{v}^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{v}|}-\eta+o(1),\quad\text{due to }\mathcal{A}_n(\theta_0,\delta_0,\xi_0).
\end{aligned}$$

and

$$\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2} \frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}} \right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta$$

$$= \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}}) \int_{\mathcal{X}^m} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2} \frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}} \right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta$$

$$- \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}}) \int_{\mathcal{N}_m(\theta)^{\complement}} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2} \frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}} \right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta$$

$$= \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}}) \int_{\mathcal{X}^m} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2} \frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}} \right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta + o(1),$$

and

$$\begin{aligned}
&\int_{\mathcal{X}^m} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2} \frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}} \right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \\
&= -\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} \left\{ \int_{\mathcal{X}^m} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[\alpha s_{\mathcal{P},n}(\theta) + (1-\alpha) s_{\mathcal{P},m}(\theta)\right] \left[\alpha s_{\mathcal{P},n}(\theta) + (1-\alpha) s_{\mathcal{P},m}(\theta)\right]^T \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \right\} \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v} \\
&\qquad\qquad \times \left[2(m+n)^{-1}(1+\eta)^2 \mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}\right]^{-1} \\
&= -\frac{\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} \left\{\alpha^2 s_{\mathcal{P},n}(\theta) s_{\mathcal{P},n}(\theta)^T + (1-\alpha)^2 \mathbf{I}_{\mathcal{P}}(\theta)/m\right\} \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}{2(m+n)^{-1}(1+\eta)^2 \mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}} \\
&= -\frac{\alpha}{(1+\eta)^2} \frac{\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} S_{\mathcal{P},n}(\theta) S_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}{2\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}} - \frac{1-\alpha}{2(1+\eta)^2},
\end{aligned}$$

Thus,

$$\begin{aligned}
&\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2}(1-\eta) \frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}} \right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta \\
&= -\frac{\alpha}{2(1+\eta)} \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathcal{P}}(\theta|\mathbf{x}^{\mathbf{n}}) \frac{\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} S_{\mathcal{P},n}(\theta) S_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}}{\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1}\mathbf{v}} \mathrm{d}\theta - \frac{1-\alpha}{2(1+\eta)} + o(1) \\
&\geq -\frac{\alpha}{2} - \frac{1-\alpha}{2(1+\eta)} + o(1) \geq -\frac{1}{2} + o(1).
\end{aligned}$$

and

$$\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \frac{m+n}{2(1+\eta)} [\alpha s_{\mathcal{P},n}(\theta) + (1-\alpha) s_{\mathcal{P},m}(\theta)]^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} [\alpha s_{\mathcal{P},n}(\theta) + (1-\alpha) s_{\mathcal{P},m}(\theta)] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta$$
$$= \frac{\alpha}{2(1+\eta)} \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) S_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} S_{\mathcal{P},n}(\theta) \mathrm{d}\theta + \frac{1-\alpha}{2(1+\eta)} D_\Theta + o(1)$$
$$\geq \frac{\alpha}{2(1+\eta)} \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) S_{\mathcal{P},n}(\theta)^T \mathbf{I}_{\mathcal{P}}(\theta)^{-1} S_{\mathcal{P},n}(\theta) \mathrm{d}\theta + \frac{1-\alpha}{2} D_\Theta - D_\Theta \eta + o(1)$$
$$\geq \frac{\alpha(D_\Theta - \eta)}{2(1+\eta)} + \frac{1-\alpha}{2} D_\Theta - D_\Theta \eta + o(1) \geq \frac{D_\Theta}{2} - (D_\Theta + \alpha D_\Theta)\eta + o(1).$$

and

$$\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathcal{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathcal{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[\ln \pi_{\mathcal{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) - \rho(\eta) - \frac{D_\Theta}{2} \ln(1+\eta) \right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta$$
$$= \ln \pi_{\mathcal{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) - \rho(\eta) - \frac{D_\Theta}{2} \ln(1+\eta) + o(1),$$

where the function $\rho(\eta)$ is defined in (185).

Therefore, we know that the asymptotically essential component (180) of the term $R_{n,3}$ is lower bounded by

$$\frac{D_\Theta - 1}{2} \ln \left(\frac{2\pi}{m+n} \right) + \frac{1}{2} \ln \frac{\det \left[\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1} \right]}{|\mathbf{v}^T \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1} \mathbf{v}|} + \frac{D_\Theta - 1}{2} + \ln \pi_{\mathcal{P}}(\theta_{0,(-1)}|\theta_{0,(1)})$$
$$- \rho(\eta) - (D_\Theta + \alpha D_\Theta)\eta - \frac{D_\Theta}{2} \ln(1+\eta) + o(1),$$

where the function $\rho(\eta)$ is defined in (185). Because of the fact that $\lim_{x\to 0} \rho(x) = 0$ and the arbitrariness of η, we know that $R^*_{n,3}$ defined in (173) is asymptotically lower bounded by zero.

2 Disaster risk model

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

2.1 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 interest gain of risk-free bond 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 risk-free rate and the market equity return are

$$1=\mathbb{E}_t\left[\frac{\Lambda_{t+1}}{\Lambda_t}e^{r_{M,t+1}}\right] \quad \text{and} \quad e^{-r_{f,t}}=\mathbb{E}_t\left[\frac{\Lambda_{t+1}}{\Lambda_t}\right]. \tag{190}$$

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

$$\mathbb{E}_t\left[\frac{\Lambda_{t+1}}{\Lambda_t}\right]=\mathbb{E}_t\left[\frac{\Lambda_{t+1}}{\Lambda_t}e^{r_{t+1}}\right]. \tag{191}$$

The left-hand side of (191) can be computed as

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

and the right-hand side of (191) can be computed as

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

Thus, the Euler equation (191) can be rewritten as

$$(1-p)e^{-\gamma_{\mathrm{D}}\mu+\frac{1}{2}\gamma_{\mathrm{D}}^2\sigma^2}\left[e^{\eta+\frac{1}{2}\tau^2-\gamma_{\mathrm{D}}\rho\sigma\tau}-1\right]=p\Delta(\xi), \tag{192}$$

where

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

Using the Taylor expansion, we have the following approximation:

$$e^{\eta+\frac{1}{2}\tau^2-\gamma_{\mathrm{D}}\rho\sigma\tau}-1\approx\eta+\frac{1}{2}\tau^2-\gamma_{\mathrm{D}}\rho\sigma\tau. \tag{193}$$

Combining (192) and the approximation in (193), we have finished proving the approximated Euler equation in the main text.

2.2 Fisher fragility measure

The joint probability density for rare disasters (z,v) in the baseline model is

$$f_{\mathcal{P}}(z,v|p,\xi)=p^z(1-p)^{1-z}\delta(v)^{1-z}\left[\mathbf{1}\{v>\underline{v}\}\xi\exp\{-\xi(v-\underline{v})\}\right]^z,. \tag{194}$$

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

$$\mathbf{I}_{\mathcal{P}}(p,\xi)=\begin{bmatrix}\frac{1}{p(1-p)} & 0\\ 0 & \frac{p}{\xi^2}\end{bmatrix}. \tag{195}$$

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

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

where

$$\eta(\theta,\phi)\equiv\gamma_{\mathrm{D}}\rho\sigma\tau-\frac{\tau^2}{2}+\ln\left[1+e^{\gamma_{\mathrm{D}}\mu-\frac{\gamma_{\mathrm{D}}^2\sigma^2}{2}}\xi\left(\frac{e^{\gamma_{\mathrm{D}}\underline{v}}}{\xi-\gamma_{\mathrm{D}}}-e^{\frac{1}{2}\varsigma^2}\frac{e^{(\gamma_{\mathrm{D}}-b)\underline{v}}}{\xi+b-\gamma_{\mathrm{D}}}\right)\frac{p}{1-p}\right]. \tag{196}$$

We can derive the simple intuitive closed-form approximation for the fragility measure, if we consider the approximated Euler equation. More precisely, we consider the following approximation:

$$\eta(\theta,\phi)\approx\gamma_{\mathrm{D}}\rho\sigma\tau-\frac{\tau^2}{2}+e^{\gamma_{\mathrm{D}}\mu-\frac{\gamma_{\mathrm{D}}^2\sigma^2}{2}}\xi\left(\frac{e^{\gamma_{\mathrm{D}}\underline{v}}}{\xi-\gamma_{\mathrm{D}}}-e^{\frac{1}{2}\varsigma^2}\frac{e^{(\gamma_{\mathrm{D}}-b)\underline{v}}}{\xi+b-\gamma_{\mathrm{D}}}\right)\frac{p}{1-p}, \tag{197}$$

Then, we can express the Fisher information for (p,ξ) under the full structural model as

$$\mathbf{I}_{\mathcal{Q}}(p,\xi)\approx\begin{bmatrix}\frac{1}{p(1-p)}+\frac{\Delta(\xi)^2}{(1-\rho^2)\tau^2}\frac{e^{2\gamma_{\mathrm{D}}\mu-\gamma_{\mathrm{D}}^2\sigma^2}}{(1-p)^3} & \frac{p}{(1-\rho^2)\tau^2}\frac{e^{2\gamma_{\mathrm{D}}\mu-\gamma_{\mathrm{D}}^2\sigma^2}}{(1-p)^2}\Delta(\xi)\dot{\Delta}(\xi)\\ \frac{p}{(1-\rho^2)\tau^2}\frac{e^{2\gamma_{\mathrm{D}}\mu-\gamma_{\mathrm{D}}^2\sigma^2}}{(1-p)^2}\Delta(\xi)\dot{\Delta}(\xi) & \frac{p}{\xi^2}+\frac{\dot{\Delta}\left(\xi\right)^2}{\left(1-\rho^2\right)\tau^2}e^{2\gamma_{\mathrm{D}}\mu-\gamma_{\mathrm{D}}^2\sigma^2}\frac{p^2}{1-p}\end{bmatrix}. \tag{198}$$

where

$$\Delta(\xi)=\xi\left(\frac{e^{\gamma_{\mathrm{D}}\underline{v}}}{\xi-\gamma_{\mathrm{D}}}-\frac{e^{\frac{\varsigma^2}{2}+(\gamma_{\mathrm{D}}-b)\underline{v}}}{\xi+b-\gamma_{\mathrm{D}}}\right). \tag{199}$$

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

$$\dot{\Delta}(\xi)=-\frac{e^{\gamma_{\mathrm{D}}\underline{v}}\gamma_{\mathrm{D}}}{(\xi-\gamma_{\mathrm{D}})^2}+\frac{e^{(\gamma_{\mathrm{D}}-b)\underline{v}}(\gamma_{\mathrm{D}}-b)}{(\xi-\gamma_{\mathrm{D}}+b)^2}e^{\varsigma^2/2}. \tag{200}$$

The worst-case Fisher fragility is the largest eigenvalue of the matrix $\Pi_0(I_{D_\Theta})\equiv\mathbf{I}_{\mathcal{Q}}(\theta_0)^{1/2}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{I}_{\mathcal{Q}}(\theta_0)^{1/2}$. Important for simplifying the calculation, 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 $\varrho^1(p,\xi)$. The minimum Fisher fragility in this case is 1, which is obtained in the direction along the deterministic cross-equation restriction.

2.3 Posteriors

Next, we construct the posteriors of the parameters $\theta=(p,\xi)$ under the baseline model and the structural model. We appeal to the Jeffreys prior for (p,ξ) under the model without asset pricing constraint as the econometrician's prior. The structural parameter γ_{D} has an independent prior $\pi(\gamma_{\mathrm{D}})$. The prior for γ_{D} can be delta distributions or the uniform priors. Given the likelihood function in (194), the parameters are mutually independent under the Jeffreys prior and their probability density functions (PDFs) are explicitly specified in Table 1.

Table 1: Independent Jeffreys/Reference priors for parameters

Parameters	Prior PDF (up to a constant)
p	$p^{-1/2}(1-p)^{-1/2}$
ξ	$\xi^{-1}\mathbf{1}_{(\xi>0)}$

The constrained likelihood function is "nonstandard" when we impose equality and inequality constraints on the parameters. Given the independent reference priors specified in Table 1 and the "nonstandard" likelihood function, not only the analytical form of the posterior density function becomes inaccessible, but also the traditional Monte Carlo methods designed to draw i.i.d. samples from the posterior become inefficient. For simulations of posterior based on a "nonstandard" likelihood function, one of the general methods is the Approximate Bayesian Computation (ABC) method.[6] One issue concerning with applying the conventional ABC method to our disaster risk model is the lack of efficiency when the priors are flat. Given the specific structure of our problem, we propose a tilted ABC method to boost the speed of our simulation. The details of the procedure are in Appendix 2.4.

2.4 ABC Method and Implementation

Given the special structure of our problem, we propose a tilted ABC method in the hope of boosting the speed of our simulation. The algorithm described here is for the case of joint estimation with the risk aversion coefficient γ_{D}. We illustrate the case where γ_{D} has the prior $\pi(\gamma_{\mathrm{D}})$. The algorithm can be adapted easily for the special case where the value of γ_{D} is fixed (i.e. delta prior for γ_{D}).

The posterior for $(p, \xi, \gamma_{\mathrm{D}})$ under the baseline model satisfies

$$
\begin{aligned}
p, \xi, \gamma_{\mathrm{D}} \mid \mathbf{r}, \mathbf{g}, \mathbf{z} \sim \ & \mathrm{Beta}\left(p|0.5 + n - \kappa_n, 0.5 + \kappa_n\right) \\
& \otimes \mathrm{Gamma}\left(\xi | n - \kappa_n, \sum_{t=1}^{n} z_t(g_t - \underline{v})\right) \\
& \otimes \pi(\gamma_{\mathrm{D}}),
\end{aligned} \tag{201}
$$

[6] For general introduction to the ABC method, see Blum (2010) and Fearnhead and Prangle (2012), among others.

where

$$x_t = (g_t, r_t)^T, \quad \mu_n = \sum_{t=1}^{n}(1-z_t)x_t / \sum_{t=1}^{n}(1-z_t), \quad \kappa_n = \sum_{t=1}^{n}(1-z_t), \quad \nu_n = \kappa_n - 1,$$

$$S_n = \sum_{t=1}^{n}(1-z_t)(x_t-\mu_n)(x_t-\mu_n)^T, \quad s_n = \sum_{t=1}^{n} z_t(r_t - bg_t)^2.$$

Define

$$\overline{g} = \sum_{t=1}^{n}(1-z_t)g_t/\kappa_n \quad \text{and} \quad \overline{r} = \sum_{t=1}^{n}(1-z_t)r_t/\kappa_n.$$

The posterior for $(p, \xi, \gamma_{\mathrm{D}})$ under the structural model satisfies:

$$\begin{aligned}
\pi_{\mathbb{Q}}(p,\xi,\gamma_{\mathrm{D}}|\mathbf{g^n},\mathbf{r^n},\mathbf{z^n}) \propto\ & p^{n-\kappa_n+1/2-1}(1-p)^{\kappa_n+1/2-1} \\
& \times \mathbf{1}_{(\xi>\gamma_{\mathrm{D}})}\xi^{n-\kappa_n-1}\exp\left\{-\xi\sum_{t=1}^{n} z_t(-g_t-\underline{v})\right\} \\
& \times \tau^{-1}(1-\rho^2)^{-1/2}\times\exp\left\{-\frac{\kappa_n}{2(1-\rho^2)\tau^2}\left[\eta(p,\xi,\gamma_{\mathrm{D}})-\overline{r}-\rho\frac{\tau}{\sigma}(\mu-\overline{g})\right]^2\right\} \\
& \times \mathbf{1}_{\{\eta(p,\xi,\gamma_{\mathrm{D}})>\underline{\eta}^*\}}\times\pi(\gamma_{\mathrm{D}}).
\end{aligned} \tag{202}$$

Then, the posterior distribution will not change if we view the model in a different way as follows:

$$\overline{r} \sim \mathrm{N}\left(\eta(p,\xi,\gamma_{\mathrm{D}}) + \rho\frac{\tau}{\sigma}(\overline{g}-\mu), \tau^2(1-\rho^2)\right) \quad \text{where} \quad \eta(p,\xi,\gamma_{\mathrm{D}}) > \underline{\eta}^*,$$

with priors

$$\gamma_{\mathrm{D}} \sim \pi(\gamma_{\mathrm{D}}),$$

$$p \sim \mathrm{Beta}(n-\kappa_n+1/2, \kappa_n+1/2),$$

$$\xi \sim \mathrm{Gamma}\left(\xi|n-\kappa_n, \sum_{t=1}^{n} z_t(g_t-\underline{v}), \xi > \gamma_{\mathrm{D}}\right).$$

The tilted ABC method is implemented as follows.

Algorithm We illustrate the algorithm for simulating samples from the posterior (202) based ABC method. We choose the threshold in ABC algorithm as $\epsilon = \hat{\tau}/n/100$, where $\hat{\tau}$ is the sample standard deviation of the observations $r_1, \cdots, r_n$. Our tilted ABC algorithm can be summarized as follows:

For step $i = 1, \cdots, N$:

Repeat the following simulations and calculations:

(1) simulate $\tilde{\gamma}_{\mathrm{D}} \sim \pi(\gamma_{\mathrm{D}})$,

(2) simulate $\tilde{p} \sim \mathrm{Beta}(n-\kappa_n+1/2, \kappa_n+1/2)$,

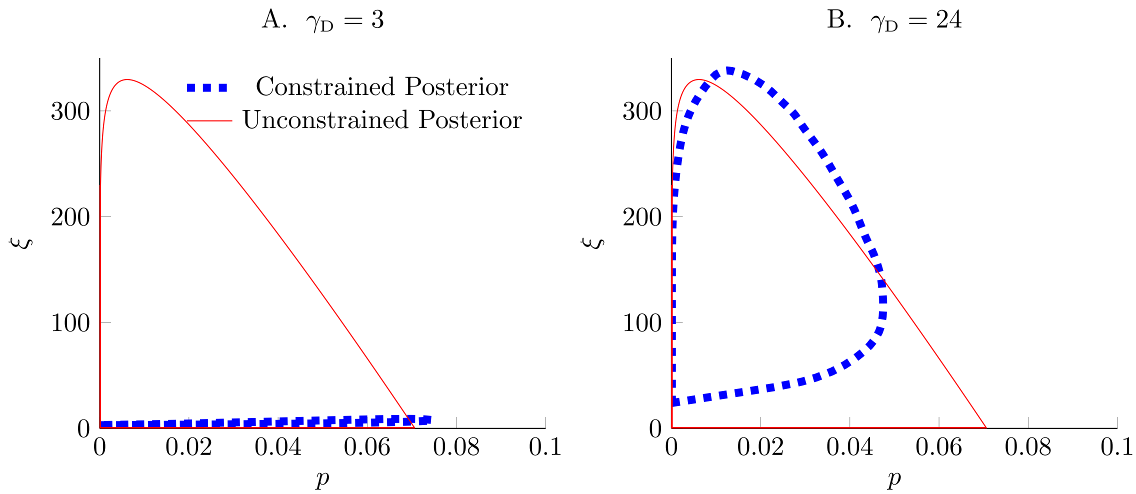

Figure 1: The 95% Bayesian confidence regions for (p, ξ). In the left panel, the posterior under the structural model (i.e. constrained posterior) sets $\gamma_{\mathrm{D}} = 3$. In the right panel, the posterior under the structural model sets $\gamma_{\mathrm{D}} = 24$. Both are compared with the posterior under the baseline model (i.e. unconstrained posterior).

(3) simulate $\tilde{\xi} \sim \text{Gamma}\left(\xi | n - \kappa_n, \sum_{t=1}^{n} z_t(g_t - \underline{v})\right)$,

(4) calculate $\tilde{\eta} = \eta(\tilde{\theta}, \tilde{\psi})$ with

$$\tilde{\theta} = (\tilde{p}, \tilde{\xi}) \quad \text{and} \quad \tilde{\psi} = \tilde{\gamma_{\mathrm{D}}},$$

(5) simulate $\tilde{r} \sim \mathrm{N}\left(\tilde{\eta} + \tilde{\rho}\frac{\tilde{\tau}}{\tilde{\sigma}}(\overline{g} - \tilde{\mu}), \tilde{\tau}^2(1 - \tilde{\rho}^2)\right)$,

Until (i) $|\tilde{r} - \overline{r}| < \epsilon$ and (ii) $\tilde{\eta} > \underline{\eta}^*$, we record

$$\theta^{(i)} = \tilde{\theta}$$
$$\psi^{(i)} = \tilde{\psi}$$

Set $i = i + 1$, if $i < N$; end the loop, if $i = N$.

Using this algorithm, we shall get simulated samples $\theta^{(1)}, \cdots, \theta^{(N)}$ from the posterior (202).

2.5 Results

Now, we show some examples of posteriors for (p, ξ) when γ has delta priors. In Figure 1 we illustrate their differences by plotting the 95% Bayesian confidence regions for (p, ξ) according to the two posteriors. The 95% Bayesian region for (p, ξ) under the baseline posterior distribution is similar to the 95% confidence region for (p, ξ) under the baseline model.

The shape of the 95% Bayesian region for the constrained posterior depends on the coefficient of relative risk aversion γ. When γ is high (e.g, $\gamma = 24$), the constrained posterior is largely similar to the unconstrained posterior (see Panel B), except that it assigns lower weight to the lower right region, because these relatively frequent and large disasters are inconsistent with the equity premium constraint. For a lower level of risk aversion, $\gamma = 3$, the constrained posterior is drastically different. The only parameter configurations consistent with the equity premium constraint are those with large average disaster size, with ξ close to its lower limit γ.

3 Long-run Risk Model: Solutions and Moment Conditions

3.1 The Model Solution

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

$$\Delta c_{t+1} = \mu_c + x_t + \sigma_t \epsilon_{c,t+1} \tag{203a}$$

$$x_{t+1} = \rho x_t + \varphi_x \sigma_t \epsilon_{x,t+1} \tag{203b}$$

$$\widetilde{\sigma}^2_{t+1} = \overline{\sigma}^2 + \nu(\widetilde{\sigma}^2_t - \overline{\sigma}^2) + \sigma_w \epsilon_{\sigma,t+1} \tag{203c}$$

$$\sigma^2_{t+1} = \max(\underline{\sigma}^2, \widetilde{\sigma}^2_{t+1}), \tag{203d}$$

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

The preference of the representative agent is assumed to be Epstein-Zin-Weil preference:

$$V_t = \left[(1-\delta_{\mathrm{L}}) C_t^{\frac{1-\gamma_{\mathrm{L}}}{\vartheta}} + \delta_{\mathrm{L}} \left(\mathbb{E}_t \left[V_{t+1}^{1-\gamma_{\mathrm{L}}} \right] \right)^{\frac{1}{\vartheta}} \right]^{\frac{\vartheta}{1-\gamma_{\mathrm{L}}}} \tag{204}$$

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

$$W_{t+1} = (W_t - C_t) R_{c,t+1}. \tag{205}$$

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

$$M_{t+1} = \delta_{\mathrm{L}}^{\vartheta} \left(\frac{C_{t+1}}{C_t} \right)^{-\vartheta/\psi_{\mathrm{L}}} R_{c,t+1}^{\vartheta-1}. \tag{206}$$

The log SDF can be written as

$$m_{t+1} = \vartheta \log \delta_{\mathrm{L}} - \frac{\vartheta}{\psi_{\mathrm{L}}} \Delta c_{t+1} + (\vartheta - 1) r_{c,t+1}. \tag{207}$$

The state variables in long-run risk models are (x_t, σ^2_t). The log consumption growth rate Δc_{t+1} can be expressed in terms of x_t and σ^2_t. In contrast, the dependence of $r_{c,t+1}$ on the state variables are endogenous. To turn the system into an affine model, we first exploit the Campbell-Shiller log-linearization approximation:

$$r_{c,t+1} = \kappa_0 + \kappa_1 z_{t+1} + \Delta c_{t+1} - z_t, \tag{208}$$

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

$$\kappa_0 = \log(1 + e^{\overline{z}}) - \kappa_1 \overline{z} \tag{209}$$

$$\kappa_1 = \frac{e^{\overline{z}}}{1 + e^{\overline{z}}}, \tag{210}$$

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

Given the log-linearization approximation (208 – 210), we can search the equilibrium characterized by

$$z_t = A_0 + A_1 x_t + A_2 \sigma_t^2, \tag{211}$$

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

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

$$r_{c,t+1} = \kappa_0 + \kappa_1 \left(A_0 + A_1 x_{t+1} + A_2 \sigma_{t+1}^2\right) + \Delta c_{t+1} - \left(A_0 + A_1 x_t + A_2 \sigma_t^2\right). \tag{212}$$

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

$$m_{t+1} = \Gamma_0 + \Gamma_1 x_t + \Gamma_2 \sigma_t^2 - \lambda_c \sigma_t \epsilon_{c,t+1} - \lambda_x \sigma_t \varphi_x \epsilon_{x,t+1} - \lambda_\sigma \sigma_w \epsilon_{\sigma,t+1}, \tag{213}$$

where predictive coefficients are

$$\Gamma_0 = \log \delta_{\mathrm{L}} - \psi_{\mathrm{L}}^{-1} \mu_c - \frac{1}{2} \vartheta(\vartheta - 1) \left(\kappa_1 A_2 \sigma_w\right)^2 \tag{214}$$

$$\Gamma_1 = -\psi_{\mathrm{L}}^{-1} \tag{215}$$

$$\Gamma_2 = (\vartheta - 1)(\kappa_1 \nu - 1) A_2 = \frac{1}{2}(\gamma_{\mathrm{L}} - 1)(\psi_{\mathrm{L}}^{-1} - \gamma_{\mathrm{L}}) \left[1 + \left(\frac{\kappa_1 \varphi_x}{1 - \kappa_1 \rho}\right)^2\right] \tag{216}$$

and the market price of risk coefficients are

$$\lambda_c = \gamma_{\mathrm{L}} \tag{217}$$

$$\lambda_x = \left(\gamma_{\mathrm{L}} - \psi_{\mathrm{L}}^{-1}\right) \frac{\kappa_1 \varphi_x}{1 - \kappa_1 \rho} \tag{218}$$

$$\lambda_\lambda = -(\gamma_{\mathrm{L}} - 1) \left(\gamma_{\mathrm{L}} - \psi_{\mathrm{L}}^{-1}\right) \frac{\kappa_1}{2(1 - \kappa_1 \nu)} \left[1 + \left(\frac{\kappa_1 \varphi_x}{1 - \kappa_1 \rho}\right)^2\right] \tag{219}$$

It can be seen that as ρ or ν approaches to unit, the risk premium goes to infinity. The coefficients

A_j's are determined by equilibrium condition (i.e. Euler Equation for price of consumption claim – pure intertemporal first-order condition of consumption decision), which is

$$1 = \mathbb{E}_t\left[M_{t+1}R_{c,t+1}\right] = \mathbb{E}_t\left[e^{m_{t+1}+r_{c,t+1}}\right] \tag{220}$$

It leads to the equilibrium conditions:

$$A_0 = \frac{1}{1-\kappa_1}\left[\log\delta_{\text{L}} + \kappa_0 + \left(1-\psi_{\text{L}}^{-1}\right)\mu_c + \kappa_1 A_2(1-\nu)\overline{\sigma}^2 + \frac{\vartheta}{2}\left(\kappa_1 A_2\sigma_w\right)^2\right] \tag{221}$$

$$A_1 = \frac{1-\psi_{\text{L}}^{-1}}{1-\kappa_1\rho} \tag{222}$$

$$A_2 = -\frac{(\gamma_{\text{L}}-1)(1-\psi_{\text{L}}^{-1})}{2(1-\kappa_1\nu)}\left[1+\left(\frac{\kappa_1\varphi_x}{1-\kappa_1\rho}\right)^2\right] \tag{223}$$

The long-run mean $\overline{z}$ is also determined endogenously in the equilibrium. More precisely, given all parameters fixed, we have $A_j = A_j(\overline{z})$ in Equations (221 – 223) because κ_0 and κ_1 are functions of $\overline{z}$. In the long-run steady state, we have

$$\overline{z} = A_0(\overline{z}) + A_2(\overline{z})\overline{\sigma}^2. \tag{224}$$

Thus, in the equilibrium, the long-run mean $\overline{z}$ is a function of all parameters in the model, according to (224) and Implicit Function Theorem,

$$\overline{z} = \overline{z}\left(\mu_c, \rho, \varphi_x, \overline{\sigma}^2, \nu, \sigma_w, \cdots\right). \tag{225}$$

And hence, we can also solve out κ_0 and κ_1 based on (225) as follows, whose explicit forms are usually not available

$$\kappa_0 = \kappa_0(\mu_c, \rho, \varphi_x, \overline{\sigma}^2, \nu, \sigma_w, \cdots) \quad \text{and} \quad \kappa_1 = \kappa_1(\mu_c, \rho, \varphi_x, \overline{\sigma}^2, \nu, \sigma_w, \cdots). \tag{226}$$

The gradients κ_0 and κ_1 with respect to the parameters, such as ρ and ν, can be calculated using Implicit Function Theorem in (224).

Given the pricing kernel in the equilibrium, we can price assets. We specify the joint distribution of the exogenous state variables and the log dividend growth Δd_t, these joint distributional assumptions are part of the structural component of the model. More precisely, we assume that the log dividend growth process is

$$\Delta d_{t+1} = \mu_d + \phi_d x_t + \varphi_{d,c}\sigma_t\epsilon_{c,t+1} + \varphi_{d,d}\sigma_t\epsilon_{d,t+1}. \tag{227}$$

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

$$r_{m,t+1} = \kappa_{m,0} + \kappa_{m,1}z_{m,t+1} + \Delta d_{t+1} - z_{m,t}, \tag{228}$$

where

$$\kappa_{m,0} = \log(1 + e^{\overline{z}_m}) - \kappa_{m,1}\overline{z}_m \tag{229}$$

and

$$\kappa_{m,1} = \frac{e^{\overline{z}_m}}{1 + e^{\overline{z}_m}} \tag{230}$$

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

$$z_{m,t} = A_{m,0} + A_{m,1}x_t + A_{m,2}\sigma_t^2, \tag{231}$$

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

$$\begin{aligned} r_{m,t+1} - \mathbb{E}_t\left[r_{m,t+1}\right] = &\ \varphi_{d,c}\sigma_t\epsilon_{c,t+1} + \kappa_{m,1}A_{m,1}\varphi_x\sigma_t\epsilon_{x,t+1} \\ &+ \kappa_{m,1}A_{m,2}\sigma_w\epsilon_{\sigma,t+1} + \varphi_{d,d}\sigma_t\epsilon_{d,t+1}, \end{aligned} \tag{232}$$

where

$$\mathbb{E}_t\left[r_{m,t+1}\right] = \mu_d + \kappa_{m,0} + (\kappa_{m,1} - 1)A_{m,0} + \kappa_{m,1}A_{m,2}(1-\nu)\overline{\sigma}^2 \tag{233}$$

$$+ \left[\phi_d + (\kappa_{m,1}\rho - 1)A_{m,1}\right]x_t + (\kappa_{m,1}\nu - 1)A_{m,2}\sigma_t^2. \tag{234}$$

Plugging the equation above into the Euler Equation

$$1 = \mathbb{E}_t\left[e^{m_{t+1}+r_{m,t+1}}\right], \tag{235}$$

we can derive the coefficients

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

$$A_{m,1} = \frac{\phi_d - \psi_{\mathrm{L}}^{-1}}{1 - \kappa_{m,1}\rho} \tag{236}$$

and

$$A_{m,2} = \frac{1}{1-\kappa_{m,1}\nu}\left[\Gamma_2 + \frac{1}{2}\left(\varphi_{d,d}^2 + (\varphi_{d,c} - \lambda_c)^2 + (\kappa_{m,1}A_{m,1}\varphi_x - \lambda_x)^2\right)\right] \tag{237}$$

In sum, according to (232), the market return can be re-written as the following beta representation for the priced aggregate shocks:

$$r_{m,t+1} - \mathbb{E}_t\left[r_{m,t+1}\right] = \beta_c\sigma_t\epsilon_{c,t+1} + \beta_x\sigma_t\epsilon_{x,t+1} + \beta_\sigma\sigma_w\epsilon_{\epsilon,t+1} + \varphi_{d,d}\sigma_t\epsilon_{d,t+1}. \tag{238}$$

where the betas are

$$\beta_c = \varphi_{d,c}, \quad \beta_x = \kappa_{m,1} A_{m,1} \varphi_x, \quad \text{and} \quad \beta_\sigma = \kappa_{m,1} A_{m,2} \tag{239}$$

Excess Market Return and Equity Premium. The Euler Equations for market return and riskfree rate can be written in one equation

$$\mathbb{E}_t \left[e^{m_{t+1}} \right] = \mathbb{E}_t \left[e^{m_{t+1} + r^e_{m,t+1}} \right]. \tag{240}$$

The risk premium is given by the beta pricing rule:

$$\mathbb{E}_t \left[r^e_{m,t+1} \right] = \lambda_c \sigma_t^2 \beta_c + \lambda_x \sigma_t^2 \beta_x + \lambda_\sigma \sigma_w^2 \beta_\sigma - \frac{1}{2} \sigma^2_{r_m,t}, \tag{241}$$

$$\text{where} \quad \sigma^2_{r_m,t} = \beta_c^2 \sigma_t^2 + \beta_x^2 \sigma_t^2 + \beta_\sigma^2 \sigma_w^2 + \varphi_{d,d}^2 \sigma_t^2. \tag{242}$$

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

$$\overline{z}_m = A_{m,0}(\overline{z}_m) + A_{m,2}(\overline{z}_m) \overline{\sigma}^2. \tag{243}$$

Based on (238), the excess log return of market portfolio $r^e_{m,t+1} = r_{m,t+1} - r_{f,t}$ has the following expression:

$$r^e_{m,t+1} - \mathbb{E}_t \left[r^e_{m,t+1} \right] = \beta_c \sigma_t \epsilon_{c,t+1} + \beta_x \sigma_t \epsilon_{x,t+1} + \beta_\sigma \sigma_w \epsilon_{\epsilon,t+1} + \varphi_{d,d} \sigma_t \epsilon_{d,t+1}. \tag{244}$$

In sum, the equilibrium excess return follows the dynamics:

$$r^e_{m,t+1} = \mu^e_{r,t} + \beta_c \sigma_t \epsilon_{c,t+1} + \beta_x \sigma_t \epsilon_{x,t+1} + \beta_\sigma \sigma_w \epsilon_{\sigma,t+1} + \varphi_{d,d} \sigma_t \epsilon_{d,t+1}, \tag{245}$$

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

$$r^e_{m,t+1} = \mu^e_{r,t} + \beta_c \sigma_t \epsilon_{c,t+1} + \beta_x \sigma_t \epsilon_{x,t+1} + \beta_\sigma \sigma_w \epsilon_{\sigma,t+1} + \varphi_{d,d} \sigma_t \epsilon_{d,t+1} + \varphi_r \sigma_t \epsilon_{r,t+1}, \tag{246}$$

which augments (245) by a normal shock $\varphi_r \sigma_t \epsilon_{r,t+1}$.

3.2 Calibrations: Simulated and Empirical Moments

The benchmark parametrization (Model 1) follows Bansal, Kiku, and Yaron (2012) and is summarized in Table 2. As Bansal, Kiku, and Yaron (2012) (Table 2, p. 194) show, the simulated first and second moments match the set of key asset pricing moments in the data reasonably well. The same is true for the alternative parametrization (Model 2) in Table 2. The alternative parametrization (Model 2) has $\nu = 0.98$ and $\gamma_{\mathrm{L}} = 27$ with other parameters unchanged. The simulated moments and sample moments are listed in Table 3. The sample moments are based annual data from 1930 to 2008, and the simulated moments are 80-year annual data aggregated from monthly simulated data.

Table 2: Parameters of the Long-Run Risk Models

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: The long-run model 2 (Model 2) has $\nu = 0.98$ and $\gamma_{\mathrm{L}} = 27$, with other parameters unchanged relative to the long-run risk model 1 with benchmark parametrization (Model 1) above.

Table 3: Simulated and Sample Moments.

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

3.3 Generalized Methods of Moments

The likelihood function of the baseline statistical model $\mathbb{P}_{\theta,n}$ can be seen clearly when re-arrange the terms

$$\frac{\Delta c_{t+1} - \mu_c - x_t}{\sigma_t} = \epsilon_{c,t+1} \tag{247a}$$

$$\frac{x_{t+1} - \rho x_t}{\varphi_x \sigma_t} = \epsilon_{x,t+1} \tag{247b}$$

$$\frac{(\sigma_{t+1}^2 - \overline{\sigma}^2) - \nu(\sigma_t^2 - \overline{\sigma}^2)}{\sigma_w} = \epsilon_{\sigma,t+1} \tag{247c}$$

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

$$\Delta d_{t+1} = \mu_d + \phi_d x_t + \varphi_{d,c} \left(\Delta c_{t+1} - \mu_c - x_t\right) + \varphi_{d,d} \sigma_t \epsilon_{d,t+1}. \tag{248}$$

We consider the GMM where the moments functions are identical to the score functions of the likelihood function. Denote the set of moment functions to be $g_{\mathcal{P}}(\Delta c_{t+1}, x_{t+1}, x_t, \sigma_{t+1}^2, \sigma_t^2, \Delta d_{t+1}; \theta)$. More precisely, $g_{\mathcal{P}}(\Delta c_{t+1}, x_{t+1}, x_t, \sigma_{t+1}^2, \sigma_t^2, \Delta d_{t+1}; \theta)$ includes ten moment conditions. The moment conditions that only involve Δc_{t+1}, x_t, and σ_t^2 are the following six moment conditions:

$$0 = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{\Delta c_{t+1} - \mu_c - x_t}{\sigma_t^2}$$

$$0 = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{(x_{t+1} - \rho x_t) x_t}{\varphi_x^2 \sigma_t^2}$$

$$1 = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{(x_{t+1} - \rho x_t)^2}{\varphi_x^2 \sigma_t^2}$$

$$0 = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{\left[(\sigma_{t+1}^2 - \overline{\sigma}^2) - \nu(\sigma_t^2 - \overline{\sigma}^2)\right] (\sigma_t^2 - \overline{\sigma}^2)}{\sigma_w^2}$$

$$1 = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{\left[(\sigma_{t+1}^2 - \overline{\sigma}^2) - \nu(\sigma_t^2 - \overline{\sigma}^2)\right]^2}{\sigma_w^2}$$

$$0 = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{(\sigma_{t+1}^2 - \overline{\sigma}^2) - \nu(\sigma_t^2 - \overline{\sigma}^2)}{\sigma_w^2}$$

The six moment conditions above captures the distribution characterized by (247a – 247c). The joint distribution of fundamental variables $(\Delta c_{t+1}, x_t, \sigma_t^2)$ and dividend growth Δd_{t+1} is captured by the following four additional moment conditions:

$$
\begin{aligned}
0 &= \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{\Delta d_{t+1} - \mu_d - \phi_d x_t - \varphi_{d,c}(\Delta c_{t+1} - \mu_c - x_t)}{\varphi_{d,d}^2 \sigma_t^2} \\
0 &= \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{x_t \left[\Delta d_{t+1} - \mu_d - \phi_d x_t - \varphi_{d,c}(\Delta c_{t+1} - \mu_c - x_t)\right]}{\varphi_{d,d}^2 \sigma_t^2} \\
0 &= \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{(\Delta c_{t+1} - \mu_c - x_t) \left[\Delta d_{t+1} - \mu_d - \phi_d x_t - \varphi_{d,c}(\Delta c_{t+1} - \mu_c - x_t)\right]}{\varphi_{d,d}^2 \sigma_t^2} \\
1 &= \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{\left[\Delta d_{t+1} - \mu_d - \phi_d x_t - \varphi_{d,c}(\Delta c_{t+1} - \mu_c - x_t)\right]^2}{\varphi_{d,d}^2 \sigma_t^2}
\end{aligned}
$$

In the long-run risk model, the major focus is to understand the stock excess return's dynamics explained by the consumption process and dividend process specified in (247a – 247c) and (248). The joint distribution of the excess log return $r_{m,t+1}^e$ and the consumption and dividend variables can be seen clearly from the following formula:

$$
\begin{aligned}
\varphi_r \sigma_t \epsilon_{r,t+1} = r_{m,t+1}^e - \mu_{r,t}^e - (\beta_c - \varphi_{d,c})(\Delta c_{t+1} - \mu_c - x_t) - \beta_x \frac{x_{t+1} - \rho x_t}{\varphi_x} \\
- \beta_\sigma \left[\widehat{\sigma}_{t+1}^2 - \nu \widehat{\sigma}_t^2\right] - (\Delta d_{t+1} - \mu_d - \phi_d x_t)
\end{aligned} \tag{249}
$$

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

$$
\mu_{r,t}^e = \lambda_\eta \beta_\eta \sigma_t^2 + \lambda_e \beta_e \sigma_t^2 + \lambda_w \beta_w \sigma_w^2 - \frac{1}{2}\left(\beta_\eta^2 \sigma_t^2 + \beta_e^2 \sigma_t^2 + \beta_w^2 \sigma_w^2 + \varphi_{d,d}^2 \sigma_t^2\right). \tag{250}
$$

Because $\beta_c = \varphi_{d,c}$, (249) can be rewritten as

$$
\varphi_r \sigma_t \epsilon_{r,t+1} = r_{m,t+1}^e - \mu_{r,t}^e - \beta_x \frac{x_{t+1} - \rho x_t}{\varphi_x} - \beta_\sigma \left[\widehat{\sigma}_{t+1}^2 - \nu \widehat{\sigma}_t^2\right] - (\Delta d_{t+1} - \mu_d - \phi_d x_t). \tag{251}
$$

We choose the over-identification moment constraints $\bar{g}_{\mathbb{Q}}(\Delta c_{t+1}, x_{t+1}, x_t, \sigma_{t+1}^2, \sigma_t^2, \Delta d_{t+1}, r_{m,t+1}^e; \theta, \psi)$ to include the score functions of the conditional likelihood of $r_{m,t+1}^e$ above. Thus, the moment conditions for the optimal GMM setup to assess the fragility of the benchmark version of long-run risk model are

$$
g_{\mathbb{Q}}(\Delta c_{t+1}, x_{t+1}, x_t, \sigma_{t+1}^2, \sigma_t^2, r_{t+1}^e, \Delta d_{t+1}; \theta, \psi) \equiv \begin{bmatrix} g_{\mathbb{P}}(\Delta c_{t+1}, x_{t+1}, x_t, \sigma_{t+1}^2, \sigma_t^2, \Delta d_{t+1}; \theta) \\ \bar{g}_{\mathbb{Q}}(\Delta c_{t+1}, x_{t+1}, x_t, \sigma_{t+1}^2, \sigma_t^2, \Delta d_{t+1}, r_{m,t+1}^e; \theta, \psi) \end{bmatrix}.
$$

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

The analytical formulas of the over-identification moment conditions are quite complicated, since how the beta coefficients and market price of risk coefficients depend on model parameters in equilibrium is extremely complicated for the long-run risk model. We ignore the formulas here and, in fact, we calculate them numerically in obtaining the fragility measures. Moreover, we compute the Fisher Information matrices for the moments in $g_{\mathcal{P}}$ and $g_{\mathcal{Q}}$ based on simulated stationary time series using the Monte Carlo method.

4 Information-Theoretic Interpretation for Model Fragility Based on Chernoff Rates

Chernoff Information. Our Fisher fragility measures are based on the information matrices from the baseline model and the full structural model. We show that by comparing the information matrices, the Fisher fragility measure quantifies the informativeness of the cross-equation restrictions in the structural model under the Chernoff information metric.

We start by introducing the concept of Chernoff information. Chernoff information gives the asymptotic geometric rate (Chernoff rate) at which the detection error probability (the weighted average of the error probabilities in selection between two alternative models) decays as the sample size increases. Intuitively, Chernoff information measures the difficulty of discriminating among alternative models.[7]

Consider a model with density $p(x|\theta_0)$ and an alternative model with density $p(x|\theta)$. Assume the densities are absolutely continuous relative to each other. The Chernoff information between the two models is defined as (see, e.g., Cover and Thomas (1991)):

$$C^*(p(x|\theta) \ : \ p(x|\theta_0)) \equiv -\ln \min_{\alpha\in[0,1]} \int_{\mathcal{X}} p(x|\theta_0)^{\alpha} p(x|\theta)^{1-\alpha} \mathrm{d}x. \tag{252}$$

The cross-equation restrictions imposed by the structural model increase efficiency of parameter estimation, which makes it is easier to distinguish model $\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_0)$ from local alternatives, $\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_0 + n^{-\frac{1}{2}}u)$ (u is a vector), compared to distinguishing $\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_0)$ from $\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_0 + n^{-\frac{1}{2}}u)$. Informativeness of cross-equation restrictions for discrimination between alternative models can be captured asymptotically by the ratio of two Chernoff rates, computed with and without imposing the cross-equation restrictions. The following proposition connects such ratio to the asymptotic fragility measure $\varrho_a(\theta_0)$, and then the one-dimensional case as a corollary will be presented afterwards.

[7] Anderson, Hansen, and Sargent (2003) use Chernoff rate to motivate a measure of model mis-specification in their analysis of robust decision making.

Proposition 23. *Assume the regularity conditions in Section 1.4 hold. Then, there exist D_Θ linearly independent $D_\Theta-$dimensional vectors $u_1, \cdots, u_{D_\Theta}$ such that*

$$\varrho(\theta_0) = \lim_{n\to\infty} \sum_{i=1}^{D_\Theta} \frac{C^*(\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_{u_i}) \ : \ \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_0))}{C^*(\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_{u_i}) \ : \ \pi_{\mathcal{P}}(\mathbf{x^n}|\theta_0))}, \tag{253}$$

where $\theta_{u_i} = \theta_0 + n^{-\frac{1}{2}} u_i$ and n is the sample size.

Proof of Proposition 23. In the proof, we consider a mathematically more general case where the matrix $\mathbf{v}$ for $\varrho^{\mathbf{v}}(\theta_0)$ is not necessarily a $D_\Theta \times D_\Theta$ identity matrix. We assume that $\mathbf{v}$ is a full-rank $D_{\mathbf{v}} \times D_\Theta$ matrix with $1 \le D_{\mathbf{v}} \le D_\Theta$. We can show that there exists a system of orthonormal basis $[\tilde{v}_1, \cdots, \tilde{v}_{D_{\mathbf{v}}}]$ of the linear space spanned by the column vectors of $\mathbf{I}_{\mathcal{Q}}(\theta_0)^{-1/2}\mathbf{v}$ such that

$$\varrho^{\mathbf{v}}(\theta_0) = \sum_{i=1}^{D_{\mathbf{v}}} \tilde{v}_i^T \mathbf{I}_{\mathcal{Q}}(\theta_0)^{1/2} \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1} \mathbf{I}_{\mathcal{Q}}(\theta_0)^{1/2} \tilde{v}_i = \sum_{i=1}^{D_{\mathbf{v}}} \frac{\hat{v}_i^T \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1} \hat{v}_i}{\hat{v}_i^T \mathbf{I}_{\mathcal{Q}}(\theta_0)^{-1} \hat{v}_i}, \tag{254}$$

where $\hat{v}_i = \mathbf{I}_{\mathcal{Q}}(\theta_0)^{1/2}\tilde{v}_i$. Because $\mathbf{I}_{\mathcal{Q}}(\theta_0)^{1/2}\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}\mathbf{I}_{\mathcal{Q}}(\theta_0)^{1/2}$ has exactly the same eigenvalues as $\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1/2}\mathbf{I}_{\mathcal{Q}}(\theta_0)\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1/2}$, there exist unit vectors $\tilde{u}_1, \cdots, \tilde{u}_{D_{\mathbf{v}}}$ such that

$$\tilde{v}_i^T \mathbf{I}_{\mathcal{Q}}(\theta_0)^{1/2} \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1} \mathbf{I}_{\mathcal{Q}}(\theta_0)^{1/2} \tilde{v}_i = \tilde{u}_i^T \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1/2} \mathbf{I}_{\mathcal{Q}}(\theta_0) \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1/2} \tilde{u}_i. \tag{255}$$

Define $u_i = \mathbf{I}_{\mathcal{P}}(\theta_0)^{-1/2}\tilde{u}_i / |\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1/2}\tilde{u}_i|$, we have

$$\varrho^{\mathbf{v}}(\theta_0) = \sum_{i=1}^{D_{\mathbf{v}}} \frac{u_i^T \mathbf{I}_{\mathcal{Q}}(\theta_0) u_i}{u_i^T \mathbf{I}_{\mathcal{P}}(\theta_0) u_i}. \tag{256}$$

Now, let's consider the Chernoff rates for the "perturbed" parameters

$$\theta_{u_i} \equiv \theta_0 + n^{-1/2} u_i, \quad \text{for } i = 1, \cdots, D_{\mathbf{v}}.$$

First, we have the following identity

$$\int [\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_0)]^{1-\alpha} [\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_{u_i})]^{\alpha} \mathrm{d}\mathbf{x^n} = \int \pi_{\mathcal{P}}(\mathbf{x^n}|\theta_0) e^{\alpha\left[\ln \pi_{\mathcal{P}}(\mathbf{x^n}|\theta_{u_i}) - \ln \pi_{\mathcal{P}}(\mathbf{x^n}|\theta_0)\right]} \mathrm{d}\mathbf{x^n}. \tag{257}$$

According to Lemma 7.6 in van der Vaart (1998), we know that the condition of differentiability in quadratic mean holds for density functions in our case. Then, the Local Asymptotic Normality (LAN) condition holds, i.e.,

$$\begin{aligned} &\ln \pi_{\mathcal{P}}(\mathbf{x^n}|\theta_{u_i}) - \ln \pi_{\mathcal{P}}(\mathbf{x^n}|\theta_0) \\ &\quad = u_i^T \left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial}{\partial \theta} \ln \pi_{\mathcal{P}}(\mathbf{x_t}; \theta_0) \right] - \frac{1}{2} u_i^T \mathbf{I}_{\mathcal{P}}(\theta_0) u_i + R_n, \end{aligned} \tag{258}$$

where $\mathbb{E}_{\mathcal{P}_0}|R_n|^2 \to 0$ because of the **Assumption F**. Under the regularity conditions, the following CLT holds, according to Theorem 2.4 of White and Domowitz (1984),

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\ln\pi_{\mathcal{P}}(\mathbf{x_t};\theta_0) \rightsquigarrow N\left(0, \mathbf{I}_{\mathcal{P}}(\theta_0)\right). \tag{259}$$

Define the Moment Generating Function (MGF) of $\ln\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_{u_i}) - \ln\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_0)$:

$$\begin{aligned} M_n(\alpha) &\equiv \mathbb{E}_{\mathcal{P}_0}\left\{e^{\alpha\left[\ln\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_{u_i}) - \ln\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_0)\right]}\right\} \\ &= e^{-\alpha\frac{1}{2}u_i^T\mathbf{I}_{\mathcal{P}}(\theta_0)u_i}\mathbb{E}_{\mathcal{P}_0}\left\{e^{\alpha u_i^T\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\ln\pi_{\mathcal{P}}(\mathbf{x_t};\theta_0)\right]}\right\} + \epsilon_n(\alpha), \end{aligned}$$

where $\epsilon_n(\alpha) = o(1)$ for each $\alpha \in [0,1]$. Therefore, as n goes large,

$$M_n(\alpha, u_i) \to e^{-\frac{1}{2}\alpha(1-\alpha)u_i^T\mathbf{I}_{\mathcal{P}}(\theta_0)u_i}, \quad \forall\ \alpha \in [0,1]. \tag{260}$$

We denote the Cumulant Generating Function (CGF) as

$$\Lambda_n(\alpha, u_i) = \ln M_n(\alpha, u_i).$$

The CGF $\Lambda_n(\alpha, u_i)$ is convex in α. Because the pointwise convergence for a sequence of convex functions implies their uniform convergence to a convex function(see e.g., Rockafellar, 1970), we know that

$$\Lambda_n(\alpha, u_i) \to -\frac{1}{2}\alpha(1-\alpha)u_i^T\mathbf{I}_{\mathcal{P}}(\theta_0)u_i. \tag{261}$$

Based on the definition of Chernoff information in (252) and the identity in (257), we know that

$$\begin{aligned} & C^*(\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_{u_i})\ :\ \pi_{\mathcal{P}}(\mathbf{x^n}|\theta_0)) \\ &\equiv \max_{\alpha\in[0,1]} -\ln\int\left[\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_0)\right]^{\alpha}\left[\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_{u_i})\right]^{1-\alpha}\mathrm{d}\mathbf{x^n} \\ &= \max_{\alpha\in[0,1]} -\Lambda_n(\alpha, u_i) \to \max_{\alpha\in[0,1]}\frac{1}{2}\alpha(1-\alpha)u_i^T\mathbf{I}_{\mathcal{P}}(\theta_0)u_i = \frac{1}{8}u_i^T\mathbf{I}_{\mathcal{P}}(\theta_0)u_i. \end{aligned} \tag{262}$$

In Equation (262) above, the convergence of maxima of $-\Lambda_n(\alpha, u_i)$ to the maximum of $\frac{1}{2}\alpha(1-\alpha)u_i^T\mathbf{I}(\theta_0)u_i$ is guaranteed by the uniform convergence of $\Lambda_n(\alpha, u_i)$.

Similarly, we can show that

$$C^*(\pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_{u_i})\ :\ \pi_{\mathcal{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_0)) = \frac{1}{8}u_i^T\mathbf{I}_{\mathcal{Q}}(\theta_0)u_i + o(1). \tag{263}$$

Therefore, we have

$$\lim_{n\to\infty}\frac{C^*(\pi_{\mathcal{Q}}(\mathbf{x^n},\mathbf{y^n}|\theta_{u_i})\ :\ \pi_{\mathcal{Q}}(\mathbf{x^n},\mathbf{y^n}|\theta_0))}{C^*(\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_{u_i})\ :\ \pi_{\mathcal{P}}(\mathbf{x^n}|\theta_0))}=\frac{\frac{1}{8}u_i^T\mathbf{I}_{\mathcal{Q}}(\theta_0)u_i}{\frac{1}{8}u_i^T\mathbf{I}_{\mathcal{P}}(\theta_0)u_i}=\frac{u_i^T\mathbf{I}_{\mathcal{Q}}(\theta_0)u_i}{u_i^T\mathbf{I}_{\mathcal{P}}(\theta_0)u_i}$$

Combining with (256), we obtain

$$\varrho^{\mathbf{v}}(\theta_0)=\lim_{n\to\infty}\sum_{i=1}^{D_{\mathbf{v}}}\frac{C^*(\pi_{\mathcal{Q}}(\mathbf{x^n},\mathbf{y^n}|\theta_{u_i})\ :\ \pi_{\mathcal{Q}}(\mathbf{x^n},\mathbf{y^n}|\theta_0))}{C^*(\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_{u_i})\ :\ \pi_{\mathcal{P}}(\mathbf{x^n}|\theta_0))}. \tag{264}$$

Corollary 6. *Assume the regularity conditions in Section 1.4 hold. Suppose $D_\Theta=1$, then for any $v\in\mathbb{R}$ it holds that*

$$\varrho(\theta_0)=\lim_{n\to\infty}\frac{C^*(\pi_{\mathcal{Q}}(\mathbf{x^n},\mathbf{y^n}|\theta_v)\ :\ \pi_{\mathcal{Q}}(\mathbf{x^n},\mathbf{y^n}|\theta_0))}{C^*(\pi_{\mathcal{P}}(\mathbf{x^n}|\theta_v)\ :\ \pi_{\mathcal{P}}(\mathbf{x^n}|\theta_0))}, \tag{265}$$

where $\theta_v=\theta_0+n^{-\frac{1}{2}}v$ and n is the sample size.

Proof of Proposition 6. For the scalar case, it has the following convenient equality:

$$\varrho(\theta_0)=\frac{v^T\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}v}{v^T\mathbf{I}_{\mathcal{Q}}(\theta_0)^{-1}v}=\frac{\mathbf{I}_{\mathcal{P}}(\theta_0)^{-1}}{\mathbf{I}_{\mathcal{Q}}(\theta_0)^{-1}}=\frac{\mathbf{I}_{\mathcal{Q}}(\theta_0)}{\mathbf{I}_{\mathcal{P}}(\theta_0)}=\frac{v^T\mathbf{I}_{\mathcal{Q}}(\theta_0)v}{v^T\mathbf{I}_{\mathcal{P}}(\theta_0)v}. \tag{266}$$

The the rest of the derivations are the same as the proof of Proposition 23.

Detection Error Probability. This subsection is mainly based on Section 12.9 in Cover and Thomas (1991). Assume $X_1,\cdots,X_n$ i.i.d. $\sim Q$. We have two hypothesis or classes: $Q=P_1$ with prior π_1 and $Q=P_2$ with prior π_2. The overall probability of error (detection error probability) is

$$P_e^n=\pi_1E_1^{(n)}+\pi_2E_2^{(n)},$$

where $E_1^{(n)}$ is the error probability when $Q=P_1$ and $E_2^{(n)}$ is the error probability when $Q=P_2$. Define the best achievable exponent in the detection error probability is

$$D^*=\lim_{n\to\infty}\min_{A_n\in\mathcal{X}^n}-\frac{1}{n}\log_2P_e^{(n)},\quad\text{where } A_n \text{ is the acceptance region.}$$

The Chernoff's Theorem shows that $D^*=C^*(P_1:P_2)$. More precisely, Chernoff's Theorem states that the best achievable exponent in the detection error probability is D^*, where

$$D^*=\mathbf{D}_{KL}(P_{\alpha^*}||P_1)=\mathbf{D}_{KL}(P_{\alpha^*}||P_2),$$

with

$$P_\alpha=\frac{P_1^\alpha(x)P_2^{1-\alpha}(x)}{\int_{\mathcal{X}}P_1^\alpha(x)P_2^{1-\alpha}(x)\mathrm{d}x}$$

and α^* is the value of α such that

$$\mathbf{D}_{KL}(P_{\alpha^*}||P_1) = \mathbf{D}_{KL}(P_{\alpha^*}||P_2) = C^*(P_1 : P_2).$$

According to the Chernoff's Theorem, intuitively, the best achievable exponent in the detection error probability is

$$P_e^{(n)} \doteq \pi_1 2^{-n\mathbf{D}_{KL}(P_{\alpha^*}||P_1)} + \pi_2 2^{-n\mathbf{D}_{KL}(P_{\alpha^*}||P_2)} = 2^{-nC^*(P_1:P_2)}. \tag{267}$$

References

Anderson, E. W., L. P. Hansen, and T. J. Sargent, 2003, "A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection," *Journal of the European Economic Association*, 1, 68–123.

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

Barron, A. R., 1989, "Uniformly powerful goodness of fit tests," *Ann. Statist.*, 17, 107–124.

Berger, J. O., 1994, "An overview of robust Bayesian analysis," *Test*, 3, 5–124, With comments and a rejoinder by the author.

Berk, R. H., 1972, "Consistency and asymptotic normality of MLE's for exponential models," *Ann. Mat. Statist.*, 43, 193–204.

Blum, M. G. B., 2010, "Approximate Bayesian computation: a nonparametric perspective," *J. Amer. Statist. Assoc.*, 105, 1178–1187, With supplementary material available online.

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.

Chen, H., W. W. Dou, and L. Kogan, 2017, "Measuring the 'Dark Matter' in Asset Pricing Models," Mit sloan and upenn wharton working papers.

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.

Clarke, B. S., and A. R. Barron, 1990, "Information-theoretic asymptotics of Bayes methods," *IEEE Trans. Inform. Theory*, 36, 453–471.

Clarke, B. S., and A. R. Barron, 1994, "Jeffreys' prior is asymptotically least favorable under entropy risk," *J. Statist. Plann. Inference*, 41, 37–60.

Cover, T. M., and J. A. Thomas, 1991, *Elements of information theory* . Wiley Series in Telecommunications, John Wiley & Sons Inc., New York, A Wiley-Interscience Publication.

Csiszár, I., 1975, "I-divergence geometry of probability distributions and minimization problems," *Ann. Probability*, 3, 146–158.

Doob, J. L., 1950, "Recent Publications: On the Theory of Stochastic Processes and Their Application to the Theory of Cosmic Radiation," *Amer. Math. Monthly*, 57, 497–498.

Dou, W. W., D. Pollard, and H. H. Zhou, 2012, "Estimation in functional regression for general exponential families," *Ann. Statist.*, 40, 2421–2451.

Fearnhead, P., and D. Prangle, 2012, "Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation," *Journal of the Royal Statistical Society Series B*, 74, 419–474.

Gray, R. M., 1988, *Probability, random processes, and ergodic properties*, Springer-Verlag, New York.

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

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

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.

Ibragimov, I. A., 1962, "Some limit theorems for stationary processes," *Teor. Verojatnost. i Primenen.*, 7, 361–392.

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

Kim, J.-Y., 2002, "Limited information likelihood and Bayesian analysis," *Journal of Econometrics*, 107, 175–193.

Kitamura, Y., and M. Stutzer, 1997, "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65, 861–874.

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.

Kullback, S., 1967, "A lower bound for discrimination information in terms of variation," *IEEE Transactions on Information Theory*, 13, 126–127.

Lazar, N. A., 2003, "Bayesian empirical likelihood," *Biometrika*, 90, 319–326.

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

Lin, X., J. Pittman, and B. Clarke, 2007, "Information conversion, effective samples, and parameter size," *IEEE Trans. Inform. Theory*, 53, 4438–4456.

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, 1987, "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–08.

Owen, A., 1990, "Empirical likelihood ratio confidence regions," *Ann. Statist.*, 18, 90–120.

Owen, A., 1991, "Empirical likelihood for linear models," *Ann. Statist.*, 19, 1725–1747.

Owen, A. B., 1988, "Empirical likelihood ratio confidence intervals for a single functional," *Biometrika*, 75, 237–249.

Pericchi, L. R., and M. E. Pérez, 1994, "Posterior robustness with more than one sampling model," *J. Statist. Plann. Inference*, 40, 279–294, With a discussion by J. Mortera and a reply by the authors.

Polson, N. G., 1992, "On the expected amount of information from a nonlinear model," *J. Roy. Statist. Soc. Ser. B*, 54, 889–895.

Rockafellar, R., 1970, *Convex Analysis* . Princeton mathematical series, Princeton University Press.

Roussas, G. G., 1996, "Exponential probability inequalities with some applications," in *Statistics, probability and game theory*, vol. 30 of *IMS Lecture Notes Monogr. Ser.*, . pp. 303–319, Inst. Math. Statist., Hayward, CA.

Samson, P.-M., 2000, "Concentration of measure inequalities for Markov chains and Φ-mixing processes," *Ann. Probab.*, 28, 416–461.

Schennach, S. M., 2005, "Bayesian exponentially tilted empirical likelihood," *Biometrika*, 92, 31–46.

Schennach, S. M., 2007, "Point estimation with exponentially tilted empirical likelihood," *Ann. Statist.*, 35, 634–672.

Schwartz, L., 1965, "On Bayes procedures," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 4, 10–26.

Taniguchi, M., and Y. Kakizawa, 2000, *Asymptotic theory of statistical inference for time series* . Springer Series in Statistics, Springer-Verlag, New York.

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

Wald, A., 1949, "Note on the consistency of the maximum likelihood estimate," *Ann. Math. Statistics*, 20, 595–601.

Wasserman, L., 1992, "Recent methodological advances in robust Bayesian inference," in *Bayesian statistics, 4 (Peñíscola, 1991)* . pp. 483–502, Oxford Univ. Press, New York.

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

Wolfowitz, J., 1949, "On Wald's proof of the consistency of the maximum likelihood estimate," *Ann. Math. Statistics*, 20, 601–602.