The Generalized Method of Moments in the Bayesian Framework and a Model and Moment Selection Criterion

Jae-Young Kim[1]
Department of Economics
SUNY-Albany

January 2000

Abstract

While the classical framework has a rich set of limited information procedures such as GMM and other related methods, the situation is not so in the Bayesian framework. We develop a limited information procedure in the Bayesian framework that does not require the knowledge of the full likelihood. The developed procedure is a Bayesian counterpart of the classical GMM but has advantages over the classical GMM for practical applications. The necessary limited information for our approach is a set of moment conditions, instead of the likelihood function, which has a counterpart in the classical GMM. Such moment conditions in the Bayesian framework are obtained from the equality condition of the Bayes' estimator and the GMM estimator. From such moment conditions, a posterior probability measure is derived that forms the basis of our limited information Bayesian procedure. This limited information posterior has some desirable properties for small and large sample analyses. An alternative approach is also provided in this paper for deriving a limited information posterior based on a variant of the empirical likelihood method where an empirical likelihood is obtained from the moment conditions of the classical GMM. This alternative approach yields asymptotically the same result as the approach explained above. Based on our limited information method, we develop a procedure for selecting the moment for GMM. This moment selection procedure is an extension of the Bayesian information criterion to the Bayesian semi-parametric, limited information framework. It is shown that under some conditions the proposed moment selection procedure is a consistent decision rule.

Keywords: Bayesian limited information procedure, Bayesian GMM, I-projection, entropy distance, empirical likelihoods, selection of moments.

JEL Classification: C11, C14, C2, C3, C5.

[1] Correspondence: Department of Economics, State University of New York - Albany, Albany, NY 12222. (518)437-4418. e-mail: jykim@csc.albany.edu

1 Introduction

In most cases of econometric practice, the researcher has only limited information on the data generating mechanism. While the classical framework has a rich set of inference methods based on limited information such as GMM and other related methods, the situation is not so in the Bayesian framework. The traditional Bayesian approach requires the knowledge of the full likelihood or the full information on the data generating mechanism. This aspect of the Bayesian approach is an important drawback for practical applications. Also, sometimes the full model or the full likelihood may involve nuisance parts that are not of interest. In this case, some type of semi-parametric procedure might be more appropriate for practical applications.

In this paper we develop a semi-parametric, limited information procedure in the Bayesian framework that does not require the knowledge of the likelihood function. The developed procedure is a Bayesian counterpart of the classical GMM but has advantages over the classical GMM for practical applications. We also develop a moment selection method in GMM based on our limited information procedure. This moment selection method is a generalization of the traditional Bayesian information criterion to the Bayesian semi-parametric, limited information framework.

The literature on the semi-parametric, limited information Bayesian procedure is relatively small. There was an earlier literature on the limited information Bayesian analysis of simultanaous equations systems. See Zellner (1971), Zellner et al. (1988), and Dreze and Richard (1983). In this literature, `limited information' refers to the classical LIML situation. However, the analysis is based on the traditional Bayesian approach with a known likelihood function. Innovative work is done by Zellner (1996,1997,1998) who developed a Bayesian method of moments based on the principle of maximum entropy that is discussed by Jaynes (1982a,b), Shore and Johnson (1980), Zellner and Highfield (1988), Cover and Thomas (1991), Zellner (1993), and Soofi (1994). In a linear regression (Zellner (1996)) or in a linear simultaneous equations system (Zellner (1995)) the procedure goes by making assumptions on the realized error terms from which posterior moments of parameters are derived. The principle of maximum entropy is applied to the given moment conditions to get a posterior density. Although Zellner's work is an important input in the literature, the analysis is limited to the case when parameters are linear. Even for a linear model we have nonlinearity in parameters in some inference methods such as two-stage least square in the Bayesian framework. Also, Zellner's analysis is based on orthonality assumptions on the expected realized errors and the regressors (predetermined variables), which is restrictive in many cases of practice. On the other hand, Kwan (1999) shows that under certain regularity conditions sampling distribution of an estimator can be reversed to the distribution of parameters conditional on the estimator. Based on this result Kwan (1999) argues that a classical limited information estimator can be given a Bayesian interpretation. However, Kwan's (1999) analysis is limited in that it requires a condition of uniform convergence in ditribution or uniform asymptotic normality of an estimator. This requirement obviously rules out the case of possible nonstationarity in time series models which has been one of the hottest issues in econometrics literature in recent years. Also, it is unclear in Kwan (1999) in what sense the reversion of the distribution of an estimator conditioned on the estimator is conceivable as a posterior.[2]

In this paper we study Bayesian limited information procedures for a general situation of GMM with a possibly dynamic, nonlinear, full simultaneous equations system. We provide two separate approaches in this paper for developing Bayesian limited information procedures. The two approaches are different in nature, utilizing given limited information in different ways. However, the two approaches yield the same result asymptotically whenever both are feasible.

The first approach studied in Section 3 is based on a set of moment conditions in the Bayesian framework instead of likelihood functions, prior densities and Bayes'

[2] It is stated in Kwan (1999), without justification, that this `reversion' is a posterior. The common notion of a posterior density, on the other hand, is a probability density in the parameter space conditional on data. Therefore, in order for the `reversion' conditioned on an estimator to be a posterior the estimator should be a sufficient statistic for the posterior. However, it is not possible to show the sufficiency before the posterior is known. Kwan's (1999) result of the asymptotic normality of the reversed distribution implies that the estimator and its second moment are asymptotically sufficient for the reversed distribution. However, this result does not imply the sufficiency of the statistics for a posterior.

Theorem. Based on a formal design of a Bayesian framework where a Bayes' estimator is defined, we obtain a set of moments from their counterparts in the classical GMM with the equality condition of the GMM estimator and the Bayes' estimator.[3] The moment conditions are described with respect to the (unknown) true posterior probability measure. We derive a limited information posterior that satisfies the same moment conditions as the true posterior by the principle of maximum entropy.[4] By its nature, our limited information posterior is the closest to the true posterior in the entropy distance in a set of posteriors that satisfy the same posterior moment conditions as the true posterior. Also, since the derived posterior probability is defined in the parameter space, nonstationarity in the sampling process does not matter for Bayesian inference. This fact implies that Sims' (1988) point applies to our limited information framework as well as to the traditional Bayesian framework that Bayesian approach is more sensible and easier to handle analytically than the classical confidence statements in the presence of possible nonstationarity.

We study asymptotic properties of the posterior derived in Section 3. The obtained asymptotic results imply an important fact on the relationship between the Bayesian approach and the classical approach. The obtained asymptotic results also provide a basis for Bayesian analysis in the case when a closed form posterior in a finite-sample is not available. Under some regularity conditions, it is shown that the derived posterior is asymptotically normal with the first and the second moments, respectively, equal to the GMM estimator and its second moment. This result implies that the GMM estimator and its second moment are asymptotically sufficient for the derived posterior. This result also implies that the derived posterior is asymptoti-

[3] An estimator obtained in the classical framework can be also obtained within the Bayesian framework. That is, under mild conditions the same estimator is obtained from minimization of average risk and from minimization of expected posterior loss. See, for example, Judge et al. (1985)

[4] In this sense, the approach in Section 3 is a generalization of Zellner (1996,1997) to the situation of GMM and other related methods. However, our analysis in Section 3 goes farther than the generalization in the dimension of models: We provide a formal approach for deriving moments for Bayesian GMM; We provide a method of obtaining a limited information posterior for a nonlinear models; Also, we study asymptotic properties of the posterior for the case when a closed-form posterior is not available in a finite sample.

cally equivalent to the true posterior if the true posterior is asymptotically quadratic in the parameter. The regularity conditions require equicontinuity of some statistic plus some relatively minor properties of the domain of the posterior. The conditions are general enough to cover a wide variety of models. The regularity conditions do not require the uniform convergence in Kwan (1999) and, therefore, allow the case of possible nonstationarity. Kwan's (1999) counter examples that violate the uniform convergence condition, in fact, violate our equicontinuity condition. As can be easily recognized, those examples in Kwan (1999) are of little importance in practice.

The second approach studied in Section 4 for developing a limited information procedure is based on a limited information likelihood that is derived from some moments of sampling characteristics. The idea of the limited information likelihood is similar to that of the empirical likelihood method studied in Owen (1988,1991), Chen (1993,1994), and Kitamura (1997), among others. That is, given some moments of sampling characteristics it is to derive a probability density of the sampling process satisfying the given moments by the principle of maximum entropy. However, while the existing empirical likelihood method considers only the first order moment to derive an empirical likelihood, our approach utilizes an (implicitly given) second order moment as well. The posterior is, then, obtained from the limited information likelihood and a prior by the Bayes' Theorem.[5] The approach in Section 4, however, would be applicable only under some sufficient stationarity because the given moments of sampling characteristics might not be valid otherwise.

We develop a moment selection method in GMM by applying the limited information Bayesian method studied in this paper. The moment selection procedure derived in this paper is a generalization of the traditional Bayesian information criterion to the Bayesian semi-parametric, limited information framework. This moment selection rule can also be used for determining an econometric model since different moments in GMM imply different models. It is shown that under some conditions the proposed moment selection procedure is a consistent decision rule. In the classical

[5] It is shown in Kim (2000) that the maximum likelihood estimator for the likelihood obtained from our approach matches the mean of a posterior from a flat prior and the likelihood while that obtained from the empirical likelihood method matches the median of the posterior.

GMM the χ^2-test for testing the validity of moments often fails to detect a misspecified model (low power) as pointed by Newey (1985). We compare our method to the χ^2-test theoretically and by Monte Carlo simulation. It is shown that the power of our method after size adjustment is higher than that of the χ^2-test. Monte Carlo study confirms this finding. On the other hand, Andrews and Lu (1998) have proposed a moment selection criterion in GMM by some ad hod manner. Our analysis provides a formal basis of building the functional form of the criterion. We compare our procedure with that of Andrews and Lu (1998).

The discussion of the paper goes as follows. Section 2 provides a summary of some key elements of GMM as a preliminary step for our analysis. In Sections 3 and 4, respectively, the first and the second approaches for developing the limited information Bayesian procedure are studied. Section 5 develops a moment selection method in GMM.

2 Preliminaries

Let x_t be an $n \times 1$ vector of stochastic processes defined on a probability space $(\Omega; \mathcal{F}; P)$. Denote by $x_T(\omega) = (x_1(\omega); \ldots; x_T(\omega))$, for $\omega \in \Omega$, a T-segment of a particular realization of $\{x_t\}$. Let θ be a $q \times 1$ vector of parameters from $\Theta \subset \mathbf{R}^q$. Let $\mathcal{G}$ be the Borel σ-algebra of Θ. Notice that $(\Theta; \mathcal{G})$ is a measurable space. In this paper Θ is a 'grand' parameter space in which all the likelihoods, priors and posteriors under consideration are defined.

Let $h(x_t; \theta)$ be an $r \times 1$ vector-valued function, $h : (\mathbf{R}^n \times \mathbf{R}^q) \to \mathbf{R}^r$. Suppose that the following r moment conditions are satisfied at $\theta_0 \in \Theta$

$$E_P[h(x_t; \theta_0)] = 0: \tag{2:1}$$

Let $g_T(x_T; \theta)$ be the sample average of $h(x_t; \theta)$:

$$g_T(x_T; \theta) \equiv \frac{1}{T} \sum_{t=1}^{T} h(x_t; \theta):$$

Assumption 1 (a) $h(x; \cdot)$ is continuously differentiable in Θ for each $x \in \mathbf{R}^n$.
(b) $h(\cdot; \theta)$ and $\partial h(\cdot; \theta)/\partial\theta$ are Borel measurable for each $\theta \in \Theta$.

Definition 1 The GMM estimator $\{\hat{\theta}_{G,T}(\omega) : T \geq 1\}$, for some $\omega \in \Omega$ is the value of θ that minimizes the objective function

$$g_T(x_T;\theta)' W_T^G g_T(x_T;\theta); \tag{2.2}$$

where $\{W_T^G\}_{T=1}^{\infty}$ is a sequence of $(r \times r)$ positive definite weighting matrices that may be a function of the data x_T.

Assuming an interior optimum, the GMM estimate $\hat{\theta}_G$ is a solution to the following system of nonlinear equations:

$$\left\{ \frac{\partial g_T(x_T;\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}} \right\}' \times W_T^G \times [g(x_T;\hat{\theta})] = 0. \tag{2.3}$$

Let $w_t = h(x_t;\theta_0)$. Denote by S the long-run variance of $w_t = h(x_t;\theta_0)$:

$$S \equiv \sum_{\tau=-\infty}^{\infty} E_P[h(x_t;\theta_0)h(x_{t-\tau};\theta_0)']. \tag{2.4}$$

Notice that the conditions (2.1) and (2.4) form conditions on the first and second moments of $h(x_t;\theta_0)$.

Consistent estimators of S are discussed in Newey and West (1987), Gallant (1987), Andrews (1991), and Andrews and Monahan (1992). Let $\hat{S}_T$ be a consistent estimator of S based on a sample of size T. An optimal GMM estimator is obtained with $W_T^G = \hat{S}_T^{-1}$:

$$g_T(x_T;\theta)' \hat{S}_T^{-1} g_T(x_T;\theta). \tag{2.5}$$

Alternatively, the long-run variance S can be expressed in the following way:

$$S = \lim_{T \to \infty} E_P[T g_T(x;\theta_0) g_T(x;\theta_0)'] \tag{2.6}$$

It is natural to have the following estimator for an estimator of S in (2.6):

$$\hat{S}_T = T g_T(x;\theta_0) g_T(x;\theta_0)'. \tag{2.7}$$

Remark 1 (Nonstationarity and the Moment Conditions in GMM)
(a) The second moment or long-run variance of S in (2.4) can be defined only

when x_t satisfies certain sufficient stationarity conditions. In the case of x_t being nonstationary, however, the expression in (2.4) is invalid because the covariance $E_P[h(x_t;\theta_0)h(x_{t-\gamma};\theta_0)']$ depends on t. On the other hand, the moment condition (2.1), given from an econometric relation or from economic theory, is assumed to hold regardless of the existence of nonstationarity.

(b) GMM in the Bayesian framework studied in this paper is also based on a set of moment conditions (3.6) and (3.8) in Section 3.1, a counterpart of (2.1) and (2.4). Not only the (first order) moment condition (3.6) but also the second order moment condition (3.8) are robust to the existence of the nonstationarity, contrary to the condition (2.4) in the classical framework. It is because the conditions (3.6) (and (3.8)) are made with respect to a probability measure in the σ-field $\mathcal{G}$ of the parameter space Θ while the conditions (2.4) (and (2.1)) are made with respect to a probability measure in the σ-field $\mathcal{F}$ of the sample space Ω.

3 The Bayes' Estimator, GMM and a Limited Information Bayesian Framework

The GMM is a limited information procedure in the classical framework. That is, the GMM estimate in Definition 1 or in (2.3) is based on the moment condition (2.1) - a set of limited information on the data generation process (DGP), not based on the full information on DGP. The main objective of this paper is to build a Bayesian conterpart of the classical GMM. Two different approaches are adopted in this paper for this purpose.

The first approach, which is studied in this section, is based on a set of moment conditions in the Bayesian framework instead of likelihood functions, prior densities and Bayes' Theorem. Based on a formal design of a Bayesian framework where a Bayes' estimator is defined, we obtain a set of moments from their counterparts in the classical GMM with the equality condition of the GMM estimator and the Bayes' estimator. The moment conditions are described with respect to the (unknown) true posterior probability measure. We derive a limited information posterior that satisfies the same moment conditions as the true posterior by the principle of maximum entropy.

The second approach, which is studied in the next section, is based on a limited information likelihood that is derived from the moment condition (2.1) of the classical GMM. A limited information posterior is obtained from the derived limited information likelihood through the Bayes' rule. As is shown in Section 4.2 the two approaches of Sections 3 and 4 yield asymptotically the same result whenever both approaches are feasible.

3.1 The Bayes' Estimator, GMM and Posterior Densities

A Bayesian framework is identified by a posterior probability measure or density defined in the measurable space $(\Theta; \mathcal{G})$ while a classical econometrics framework is identified by a probability measure in $(\Omega; \mathcal{F})$. In this paper we study how to get a posterior density in the measurable space $(\Theta; \mathcal{G})$ based on some limited information on the nature of the world. In general, some characteristics of the true posterior are revealed in the set of such limited information, if the true posterior is unknown. We discuss how to get a posterior probability density in $(\Theta; \mathcal{G})$ that is as close as possible to the true posterior from the given limited information.

Let $\pi_T(\theta|x_T(\omega))$ be the 'true' posterior of θ that may be unknown.[6] Assume that the posterior $\pi_T(\cdot|x_T(\cdot))$ is jointly measurable $\mathcal{G} \times \mathcal{F}$. Defining

$$P_T(G; \omega) = \int_G \pi_T(\theta|x_T(\omega))d\theta$$

for any $G \in \mathcal{G}$ and $\omega \in \Omega$, $P_T(\cdot; \omega)$ is a probability measure on Θ for every $\omega \in \Omega$, and $P_T(G; \cdot)$ is a random variable for each $G \in \mathcal{G}$.

Let $\ell(\theta; \delta)$ be a loss function that reflects the consequences of choosing δ when θ is the real parameter value. The Bayes' estimator is an estimator that minimizes the expected posterior loss:

$$(3{:}1) \qquad \hat{\theta}_B = \delta^*(x_T) = \operatorname{argmin}_\delta E^\pi[\ell(\theta; \delta)]$$

[6] We can think of $\pi_T(\theta|x_T)$ as the posterior of θ obtained from the true likelihood of θ, if any. Or, it is a posterior of θ containing a richer set of information on the true model than that in the limited information posterior studied in this paper.

where

$$(3.2) \qquad E^{\pi}[\ell(\theta,\delta)] = \int_{\Theta} \ell(\theta,\delta)\pi(\theta|x_T)d\theta.$$

We are interested in a loss function that yields an estimator equivalent to the GMM estimator. Since our objective is to study a Bayesian conterpart of the classical GMM, it is natural to adopt a loss function with this property. Thus, consider the following loss function that is quadratic in g_T:

$$(3.3) \qquad \ell(\theta,\delta) = L(g_T(\theta), g_T(\delta)) = [g_T(\theta) - g_T(\delta)]' W_T [g_T(\theta) - g_T(\delta)]$$

where $\{W_T\}_{T=1}^{\infty}$ is a sequence of positive definite weighting matrices. The loss function (3.3) can be transformed into a loss function quadratic in θ:

$$(3.4) \qquad \ell(\theta,\delta) = [\theta - \delta]' \tilde{W}_T [\theta - \delta]$$

where $\tilde{W}_T = \{\partial g(\bar{\theta})/\partial\theta\}' W_T \{\partial g(\bar{\theta})/\partial\theta\}$, where $\bar{\theta} \in (\theta,\delta)$.

The loss function (3.3) or (3.4) is such that it yields an estimator that is the same as the GMM estimator under some conditions. (See Lemma 1 below.) The results of this section would be robust to the choice of the loss function so far as the chosen loss function has this property.

The first order condition for the minimization problem (3.1) with the loss function (3.3) yields the following equation:

$$(3.5) \qquad \left\{\frac{\partial g_T(\theta)}{\partial\theta}\bigg|_{\theta=\hat{\theta}}\right\}' W_T g_T(\hat{\theta}) = E^{\pi}\left[\left\{\frac{\partial g_T(\theta)}{\partial\theta}\bigg|_{\theta=\hat{\theta}}\right\}' W_T g_T(\theta)\right].$$

Then, the equality of the Bayes' estimator and the GMM estimator entails a moment condition for $g_T(\theta)$:

Lemma 1 Assume the second order conditions hold for the GMM estimate in Definition 1 and for the Bayes' estimate in (3.1). Then, under Assumption 1 the Bayes' estimator $\hat{\theta}_B$ is equal to the GMM estimator $\hat{\theta}_G$ if and only if

$$(3.6) \qquad E^{\pi}[\Xi_T g_T(\theta)] = 0$$

where

$$\Xi_T = \left\{\frac{\partial g_T(\theta)}{\partial\theta}\bigg|_{\theta=\hat{\theta}}\right\}' \times W_T$$

with $\{W_T\} = \{W_T^G\}$.

For notational convenience, let

$$\eta_T(\theta) = \Xi_T g_T(\theta). \tag{3.7}$$

The equation (3.6) describes a moment condition on η_T. We assume that $\eta_T(\theta)$ has a second moment:

Assumption 2 Assume that there exists a sequence of $q \times q$ matrices $\{A_T\}_{T=1}^{\infty}$ such that for $T \geq 1$

$$E^{\pi}[T\eta_T(\theta)\eta_T(\theta)'] = A_T. \tag{3.8}$$

Conditions (3.6) and (3.8) are about the first and the second moments of $\eta_T(x_T; \theta)$, forming a counterpart of the conditions (2.1) and (2.4) or (2.6) for $h(x_t; \theta_0)$ in the classical framework. Notice that we only assume the existence of the second moment in Assumption 2 while we have a specific value for the first moment in (3.6). This feature is similar to that in the classical GMM in the conditions (2.1) and (2.4).

Remark 2 (Nonstationarity and the Moment Conditions (3.6) and (3.8)) The conditions (3.6) and (3.8) are made with respect to a probability measure defined in the σ-field $\mathcal{G}$ of Θ, not Ω. Therefore, nonstationarity in $x_t(\omega)$ for $\omega \in \Omega$ does not matter for the conditions (3.6) and (3.8).

Our objective is to construct a limited information posterior (LIP) of θ where the `true' posterior is not available. We are interested in an LIP of θ having the following properties: (1) It is consistent with the properties of the true posterior described in (3.6) and (3.8), (2) It is the closest to the true posterior in the entropy distance or the Kullback-Leibler information distance in a set of posteriors satisfying (1). In addition, the limited information posterior has the following two properties as is studied in Section 3.2: (1) It is asymptotically equivalent to the true posterior as far as the true posterior is asymptotically quadratic in θ. (2) The classical GMM estimator with its second moment is asymptotically sufficient for the derived posterior density.

Let $\dagger$ be the family of posterior densities satisfying the same moment conditions (3.6) and (3.8) as π:

$$\dagger = \left\{\pi : E^{\pi}[\eta_T(\theta)] = 0\right\} \cap \left\{\pi : E^{\pi}[T\eta_T(\theta)\eta_T(\theta)'] = A_T\right\} \tag{3.9}$$

where

$$E^{\pi}[(\cdot)] = \int (\cdot)\pi d\theta.$$

For $\pi \in \Gamma$ we are interested in the one that is the closest to the true posterior π in the entropy distance or the Kullback-Leibler information distance:

$$\pi^{*} = \mathrm{argmin}_{\pi \in \Gamma} \int \ln(\pi/\pi)\pi d\theta. \tag{3.10}$$

The density π^{*} is interpreted as the I-projection of π on Γ. See Csiszar (1975). The intuition is that π^{*} is the projection of π on an information set `spanning' Γ which has the closest distance from π to the set Γ. We call π^{*} an I-projection posterior or a limited information posterior. Following Csiszar (1975), we can show that the density π^{*} is as in the following:

Theorem 1 Let π^{*} be the I-projection of π on Γ. Then, π^{*} is of form

$$\pi^{*}_{T}(\theta|x_T) = C_T \exp[c \cdot T\eta_T(\theta)' A_T^{-1} \eta_T(\theta)] \tag{3.11}$$

where c is a constant, and C_T is a normalizing constant.

Although the posterior $\pi^{*}_{T}(\theta|x_T)$ in (3.11) has desirable properties explained above, it is sometimes not useful in practice since there is a function $\eta_T(\cdot)$ `between' θ and $\pi^{*}_{T}(\cdot|x_T)$. For example, computation of a posterior probability $\bar{P}_T(G;\omega) = \int_G \pi^{*}_{T}(\theta|x_T(\omega))d\theta$ for a $G \in \mathcal{G}$ analytically or by some numerical method such as the Gibbs sampler would not be easy unless $g_T(\theta)$ or $\eta_T(\theta)$ is of a simple form such as a linear function of θ.

The posterior $\pi^{*}_{T}(\theta|x_T)$ in (3.11) can be transformed into an alternative that is a direct function of θ. To get an alternative form of $\pi^{*}_{T}(\theta|x_T)$, notice that by the mean value theorem

$$\eta_T(\theta) = \eta_T(\hat{\theta}) + \left(\frac{\partial \eta_T(\theta)}{\partial \theta}|_{\theta=\bar{\theta}}\right)'(\theta - \hat{\theta})$$

where $\bar{\theta} \in (\hat{\theta};\theta)$ for $\hat{\theta} = \hat{\theta}_B = \hat{\theta}_G$. But, since $\eta_T(\hat{\theta}) = 0$ from the first order condition for the minimization problem of the Bayes' estimate we have

$$(\theta - \hat{\theta}) = B_T(\bar{\theta})^{-1}\eta_T(\theta) \tag{3.12}$$

where

$$B_T(\theta) = \frac{\partial \eta_T(\theta)}{\partial \theta}: \tag{3:13}$$

Then, as a corollary of Theorem 1 we get the following result:

Corollary 1 Let π^* be the I-projection of π on $\dagger_{-}$. Then, π^* is such that

$$\pi_T^*(\theta|x_T) \propto \exp[\gamma \cdot T(\theta - \hat{\theta})' B_T(\bar{\theta})' A_T^{-1} B_T(\bar{\theta})(\theta - \hat{\theta})] \tag{3:14}$$

where $\bar{\theta} \in (\theta; \hat{\theta})$, and γ is a constant.

It is appranent in (3.14) that θ is approximately normal with the first and second `moments', respectively, equal to $\hat{\theta}$ and $\{B_T(\bar{\theta})' A_T^{-1} B_T(\bar{\theta})\}^{-1}$. If the function $h(x_t; \theta)$ or $g_T(x_T; \theta)$ is linear in θ, then $B_T(\cdot)$ does not depend on θ. In this case the I-projection posterior of θ is normal with the first and second `moments', respectively, equal to $\hat{\theta}$ and $\{B_T' A_T^{-1} B_T\}^{-1}$, according to (3.14). However, if $B_T(\cdot)$ depends on θ as in most cases of $g_T(x_T; \theta)$ being nonlinear in θ, this is not true. In fact, in this latter case the form of the posterior $\pi_T^*(\theta|x_T)$ in (3.14) is of no use in practice since $\bar{\theta}$ and thus the second `moment' $\{B_T(\bar{\theta})' A_T^{-1} B_T(\bar{\theta})\}^{-1}$ are not uniquely determined. In this case, however, the second moment is uniquely determined in asymptotics, so that asymptotically the I-projection posterior is normal with unique first and second moments. See Section 3.2.

The Bayesian framework studied above is based on the condition of equality of the Bayes' estimator and the GMM estimator. Now, let us consider the case of the optimal Bayes' estimator and the optimal GMM estimator.

By the same reason as in Hansen (1982) the optimal Bayes' estimator, having the shortest posterior probability interval of a given probability content, is obtained by setting[7]

$$W_T = \{E^{\pi}[T g_T(\theta) g_T(\theta)']\}^{-1}: \tag{3:15}$$

[7] The result (3.16) in Lemma 2 or the result (3.23) in Lemma 4 implies that under (3.15) the asymptotic posterior second moment of θ, $\Sigma_T(\hat{\theta})$, in (3.19) is equal to $\{T A_T\}^{-1}$ or $\{T B_T(\hat{\theta})\}^{-1}$. We can show that the shortest posterior probability interval of a given probability content is obtained in this case by the same reason as in Hansen (1982).

Notice that from (3.7) and (3.8), we have

$$A_T = E^{\pi}[T\Xi_T g_T(\theta) g_T(\theta)' \Xi_T'];$$

and from (3.13) and from the definitions of Ξ_T and η_T in Lemma 1 and in (3.7), respectively, we have

$$B_T(\hat{\theta}) = [\Xi_T W_T^{-1} \Xi_T']:$$

Therefore, at the optimal Bayes' estimate with $W_T = \{E^{\pi}[T g_T(\theta) g_T(\theta)']\}^{-1}$, we have the following result:

Lemma 2 Let $W_T = \{E^{\pi}[T g_T(\theta) g_T(\theta)']\}^{-1}$. Then, it is true that

$$(3:16) \qquad A_T = B_T(\hat{\theta}):$$

Now, consider the case when π_T is such that the optimal weighting matrix W_T given in (3.15) is equal to the optimal GMM weighting matrix[8]:

$$(3:17) \qquad W_T = \{E^{\pi}[T g_T(\theta) g_T(\theta)']\}^{-1} = \hat{S}^{-1}:$$

where $\hat{S}$ is a consistent estimator of S in (2.4). The condition (3.17) implies the following result:

Lemma 3 Let the true posterior π be such that $W_T = \{E^{\pi}[T g_T(\theta) g_T(\theta)']\}^{-1} = \hat{S}^{-1}$: Also, let A_T^0 and B_T^0 be A_T and B_T at such π and W_T. Denote by $V_T(\hat{\theta}_G)$ the asymptotic variance-covariance matrix of the optimal GMM estimator. Then, it is true that

$$\begin{aligned}(3:18) \qquad A_T^0 &= B_T^0(\hat{\theta}) \\ &= D_T' \hat{S}^{-1} D_T \equiv [T V_T(\hat{\theta}_G)]^{-1}\end{aligned}$$

where

$$D_T = \left\{ \frac{\partial g_T(\theta)}{\partial \theta} |_{\theta=\hat{\theta}} \right\}:$$

[8] As is clear from (3.1) a different posterior π_T yields a different Bayes' estimator. A different weighting matrix W_T gives a different Bayes' estimator as well.

3.2 Asymptotic Approximations and Properties

We first introduce a neighborhood system in Θ in which the posterior is defined. Let $N(\hat\theta_T;\delta_T)$, $T = 1,\dots,\infty$, be such that

$$N(\hat\theta_T;\delta_T) = \{\theta : |\theta_1 - \hat\theta_{T1}|^2/\delta_{T1}^2 + \cdots + |\theta_k - \hat\theta_{Tk}|^2/\delta_{Tk}^2 < 1\}$$

where $\hat\theta_{Ti}$ is the ith element of $\hat\theta_T$; $\delta_T = (\delta_{T1},\dots,\delta_{Tk})'$ is a q-vector of real numbers; $|\cdot|$ denotes the usual Euclidean norm. We consider a sequence $\{\delta_T\}$ such that δ_T becomes smaller and smaller as $T \nearrow \infty$, so that $N(\hat\theta_T;\delta_T)$ shrinks as T gets larger. Also, δ_T may depend on $\omega \in \Omega$.

Denote by $\|\cdot\|$ the matrix norm: For an $m \times m$ matrix A, $\|A\| = \sup |Ax|/|x|$, where $|Ax|$ is the usual Euclidean norm on $\mathbb{R}^m$. For notational convenience, define

$$\Sigma_T(\theta) = \left\{T B_T(\theta)' A_T^{-1} B_T(\theta)\right\}^{-1}. \tag{3.19}$$

Notice that under the optimality condition (3.17) we have

$$\Sigma_T^0(\hat\theta) = \{T B_T^0(\hat\theta)\}^{-1} = \{T A_T^0\}^{-1}. \tag{3.20}$$

where $\Sigma_T^0(\hat\theta)$ denotes $\Sigma_T(\hat\theta)$ under the optimality condition (3.17).

Now, consider the following conditions (C1) and (C2).

(C1)(a) Let $\rho_T(\hat\theta_T(\omega);\delta_T) = \sup_{\theta\in N(\hat\theta_T;\delta_T)} \|[\Sigma_T(\hat\theta_T)]^{-1}[\Sigma_T(\theta) - \Sigma_T(\hat\theta_T)]\|$. There exists a positive sequence $\{\delta_T\}_{T=1}^{\infty}$ such that $\lim_{T\nearrow\infty} P[\rho_T(\hat\theta_T(\omega);\delta_T) < \epsilon] = 1$ for each $\epsilon > 0$. (b) For δ_T satisfying (C1)(a) the absolute value of each element of the vector $\Sigma_T(\hat\theta_T)^{-1/2}\delta_T$ tends to infinity as $T \nearrow \infty$ in P-probability.

Condition (C1)(a) is a smoothness or equicontinuity condition of $\Sigma_T(\theta)$ in $N(\hat\theta_T;\delta_T)$. This condition rules out the case with a suddern 'jump' in $\Sigma_T(\cdot)$ in $N(\hat\theta_T;\delta_T)$. Condition (C1)(b) guarantees that the the neighborhood $N(\hat\theta_T;\delta_T)$ is wide enough to cover the domain of the posterior. These two conditions (C1)(a) and (b) are not really binding in many cases of practice.[9] In fact, conditions (C1) and (C2) (below) cover

[9] We can find in Kwan (1999) examples of models that violate the smoothness condition (C1)(a). Three examples are presented in Kwan (1999). Interpreting them in the model of our interest,

a very wide variety of models including the case with possible nonstationarity.[10]

By (3.18) and (3.20) we know that the condition (C1)(b) can be stated in terms of $V_T(\hat{\theta}_G)$ under (3.17) or in terms of $B^0_T(\hat{\theta})$ under (3.15). The condition (C1)(b) stated in terms of $B^0_T(\hat{\theta})$ or $V_T(\hat{\theta}_G)$ is much easier to check for a given $g_T(x_T;\theta)$ than that stated in terms of $\Sigma_T(\hat{\theta})$. Also, the following condition (C1)(a)$'$, which is much easier to check for a given $g_T(x_T;\theta)$, is sufficient for (C1)(a) under (3.17):

(C1)(a)$'$ Let $B^0_T(\theta) = D^0_T \hat{S}^{-1} D_T(\theta)$ where $D_T(\theta) = \partial g_T(\theta)/\partial\theta$. Let $m_T(\hat{\theta}_T(\omega);\delta_T) = \sup_{\theta\in N(\hat{\theta}_T;\delta_T)} \|[B^0_T(\hat{\theta})]^{-1}[B^0_T(\theta) - B^0_T(\hat{\theta})]\|$. There exists a positive sequence $\{\delta_T\}^{\infty}_{T=1}$ such that $\lim_{T\uparrow\infty} P[m_T(\hat{\theta}_T(\omega);\delta_T) < \epsilon] = 1$ for each $\epsilon > 0$.

The following condition is about asymptotic concentration of $\theta \in N(\hat{\theta}_T;\delta_T)$ in the sense of Berk (1970):

(C2) Let $\pi^*_T(\theta|x_T)$ be the posterior as given in (3.11). For δ_T satisfying (C1),

$$\int_{\Theta\setminus N(\hat{\theta}_T;\delta_T)} \pi^*_T(\theta|x_T)d\theta \;-\!\!\to\; 0$$

as $T\uparrow\infty$, i.e., θ concentrates in $N(\hat{\theta}_T;\delta_T)$ as $T\uparrow\infty$.

We can show that under the conditions (C1)-(C2) the posterior $\pi^*_T(\theta|x_T)$ is asymptotically normal. Thus, let $\phi(\cdot)$ denote the standard normal p.d.f. defined on $\mathbb{R}^q$. Also, for $a,b\in\mathbb{R}^q$, $a = (a_1,\ldots,a_q)$, etc., let (a,b) be a q-dimensional interval, that is, $(a,b) = \{y = (y_1,\ldots,y_q) : a_i < y_i < b_i,\ i = 1,\ldots,q\}$.

Theorem 2 Assume that (C1) and (C2) are satisfied. Then, for each (a,b),

$$\int_{J_{Tab}} \pi^*_T(\theta|x_T)d\theta \;-\!\!\to \int_a^b \phi(z)dz$$

Example 2 and Example 4 of Kwan (1999) are 'designed' such that the (long-run) variance of $w_t = h(x_t;\theta_0)$ or the variance of $\hat{\theta}$ has a sharp discontinuity point at some θ in the neighborhood $N(\hat{\theta}_T;\delta_T)$ and can never be smooth in the sense of (C1)(a). Example 1 in Kwan (1999), on the other hand, presents $\hat{\theta}$, an estimate of the mean, that depends on the sample mean. Anyone of these examples, however, are of little interest in practice.

[10] See Kim (1998) for examples of models satisfying (C1) and (C2). Although the examples in Kim (1998) are true likelihood functions satisfying conditions similar to (C1) and (C2), the same method can be applied to find examples of $h(x_t;\theta)$ satisfying (C1) and (C2).

in P-probability where $J_{Tab} = \{\theta : [\Sigma_T(\hat{\theta}_T)]^{-1/2}(\theta - \hat{\theta}_T) \in (a, b)\}$.

Theorem 2 states that under some conditions the posterior distribution of the parameter θ is asymptotically normal with the first moment of the distribution equal to $\hat{\theta}_T$ and the second moment $\Sigma_T(\hat{\theta}_T)$:

$$\theta|x_T \overset{a}{\sim} N(\hat{\theta}_T, \Sigma_T(\hat{\theta}_T)). \tag{3.21}$$

The result of Theorem 2 further implies that the statistics $(\hat{\theta}_T, \Sigma_T(\hat{\theta}_T))$ are jointly sufficient for θ in the posterior $\pi_T^*(\theta|x_T)$ in asymptotics, where asymptotic sufficieny is defined in the following:

Definition 2 A statistic $s(x_T)$ is asymptotically sufficient for θ in the posterior $\pi_T(\theta|x_T)$ if for each (a, b),

$$\left| \int_{J_{Tab}} \pi_T(\theta|x_T)d\theta - \int_{J_{Tab}} \pi_T(\theta|s(x_T))d\theta \right| \to 0.$$

Corollary 2 Assume that (C1) and (C2) are satisfied. Let $s(x_T) = (\hat{\theta}_T, \Sigma_T(\hat{\theta}_T))$ Then, $s(x_T)$ is asymptotically sufficient for θ in the posterior $\pi_T^*(\theta|x_T)$.

The result of Corollary 2 implies that for large sample analysis we can construct a posterior based on the statistics $(\hat{\theta}_T, \Sigma_T(\hat{\theta}_T))$ rather than based on the whole sample x_T and the full likelihood. The approximated posterior is normal with the first moment of the distribution equal to $\hat{\theta}_T$ and the second moment $\Sigma_T(\hat{\theta}_T)$:

$$\theta|s(x_T) \overset{a}{\sim} N(\hat{\theta}_T, \Sigma_T(\hat{\theta}_T)). \tag{3.22}$$

Now, consider the case of the optimal Bayes' estimator either under the condition (3.15) or under (3.17). In this case, we have the following results:

Lemma 4 Under condition (C2), if $W_T = \{E^{\pi}[g_T(\theta)g_T(\theta)']\}^{-1}$ it is true that

$$\limsup_{\theta \in N(\hat{\theta}_T, \delta_T)} \|A_T^{-1}B_T(\theta) - I\| = 0. \tag{3.23}$$

In addition, if $W_T = \{E^{\pi}[g_T(\theta)g_T(\theta)']\}^{-1} = \hat{S}^{-1}$ it is true that

$$\limsup_{\theta \in N(\hat{\theta}_T, \delta_T)} \|\Sigma_T^0(\theta) - V_T(\hat{\theta}_G)\| = 0. \tag{3.24}$$

By (3.24) the results of Theorem 2 and Corollary 2 imply the following important fact in econometrics:

Proposition 1 Let $s^0(x_T) = (\hat{\mu}_T; V_T(\hat{\mu}_G))$. Assume that the conditions in Theorem 2 hold. Then, for the optimal Bayes' estimator with $W_T = \hat{S}_T^{i1}$ it is true that

$$(3{:}25) \qquad \mu j s^0(x_T) \overset{a}{»} N(\hat{\mu}_T; V_T(\hat{\mu}_G)):$$

The result of Proposition 1 implies that the optimal I-projection posterior, which is based on the optimal Bayes' estimator with $W_T = \hat{S}_T^{i1}$, can be constructed given the GMM estimator $\hat{\mu}_T$ and its asymptotic variance $V_T(\hat{\mu}_G)$. The optimal I-projection posterior of μ is asymptotically normal with the first moment equal to $\hat{\mu}_T$ and the second moment $V_T(\hat{\mu}_G)$.

4 Alternative Approach

In the previous section we directly derived a limited information posterior based on a set of moment conditions in the Bayesian framework. In this section we derive a limited information posterior based on moment conditions of GMM in the sampling theory framework. We first derive a limited information likelihood and then get a limited information posterior through the Bayes' rule. Later on in Section 4.2 we show that the two approaches of Sections 3 and 4 yield asymptotically the same result whenever both approaches are feasible.

4.1 GMM and a Limited Information Likelihood

From the moment condition (2.1) in Section 2, we have

$$(4{:}1) \qquad E_P[g_T(x_T; \mu_0)] = 0:$$

Also, we have the second moment condition on g_T from (2.4) and (2.6):

$$(4{:}2) \qquad \lim_{T\%1} E_P[T g_T(x_T; \mu_0) g_T(x_T; \mu_0)^0] = S$$

where S is the long-run variance of $w_t = h(x_t;\theta_0)$ explained in (2.4).[11]

Given the true probability measure P with the properties in the moment conditions (4.1) and (4.2), we are interested in a probability measure Q that implies the same moment conditions: Thus, let $\mathcal{Q}$ be a family of probability measures that is absolutely continuous with respect to P such that for $\theta \in \Theta$

$$(4.3)\quad \mathcal{Q}(\theta) = \{Q : E_Q[g_T(x_T;\theta)] = 0\} \cap \left\{ Q : \lim_{T\to\infty} E_Q[T g_T(x_T;\theta) g_T(x_T;\theta)'] = S \right\}.$$

For $Q \in \mathcal{Q}$ we are interested in the one that is the closest to the true probability measure P in the entropy distance or the Kullback-Leibler information distance:

$$(4.4)\qquad Q^* = \operatorname{argmin}_{Q\in\mathcal{Q}} I(Q\|P) \equiv \int \ln(dQ/dP)dQ$$

where dQ/dP is the Radon-Nikodym derivative (or density) of Q with respect to P. We denote by $q_P^* = dQ^*/dP$ the Radon-Nikodym derivative of Q^* with respect to P. We call $q_P^*(\theta)$ a limited information likelihood or the I-projection likelihood following the notion of Csiszar (1975).

The idea of the limited information likelihood $q_P^*(\theta)$ is similar to that of the empirical likelihood studied in Owen (1988,1991), Chen (1993,1994), Kolaczyk (1994), Chen and Hall (1993), Quin (1993), Quin and Lawless (1994), DiCiccio, Hall and Romano (1989,1991), DiCiccio and Romano (1989,1990), Hall (1990), and Kitamura (1997). However, while the empirical likelihood method of these authors is based on the first order moments such as (2.1), our approach utilizes the second order moment (4.2) as well.[12]

Following Csiszar (1975), we can show that q_P^* is as in the following:

[11] Since the conditions (4.1) and (4.2) are described with respect to the probability measure P defined on $\mathcal{F}$, the existence of nonstationarity may matter for these conditions, different from (3.6) and (3.8).

[12] The likelihood obtained from the two approaches yield different MLEs: The MLE for the likelihood obtained in our approach is the same as the GMM estimator for each T while the MLE for the likelihood obtained from the empirical likelihood method is not. As is shown in Kim (2000), the MLE for the likelihood obtained from our approach matches the mean of a posterior from a flat prior and the likelihood while that obtained from the empirical likelihood method matches the median of the posterior.

Theorem 3 Under the conditions on Q,

$$(4.5)\qquad q_P^*(\theta) = K\exp\left\{\lim_{T\nearrow\infty}\kappa\cdot T g_T(x_T,\theta)' S^{-1} g_T(x_T,\theta)\right\}$$

where κ is a constant, and K is a normalizing constant.

A natural finite-sample analogue of q_P denoted by $q_{P,T}$ for a sample x_T is

$$(4.6)\qquad q_{P,T}^*(\theta) = K_T\exp\left\{\kappa\cdot T g_T(x_T,\theta)'\hat{S}_T^{-1} g_T(x_T,\theta)\right\}$$

where K_T is a normalizing constant.

It is easy to show that the maximum likelihood estimator (MLE) of the limited information likelihood $q_{P,T}^*(\theta)$ in (4.6), denoted by $\hat{\theta}_{LM}$, is the same as the optimal GMM estimator:

Lemma 5 Let $\kappa < 0$. For a given T, it is true that

$$(4.7)\qquad \text{argmax}_{\theta\in\Theta}\log q_{P,T}^*(\theta) = \text{argmin}_{\theta\in\Theta} g_T(x_T,\theta)'\hat{S}^{-1} g_T(x_T,\theta).$$

Furthermore, the asymptotic variance of $\hat{\theta}_{LM}$ is the same as the asymptotic variance of the GMM estimator, that is,

$$(4.8)\qquad \lim_{T\nearrow\infty}\left[\frac{\partial^2\ln q_{P,T}^*(\bar{\theta}^1)}{\partial\theta\partial\theta}\right]^{-1}\left[\left(\frac{\partial\ln q_{P,T}^*(\theta_0)}{\partial\theta}\right)\left(\frac{\partial\ln q_{P,T}^*(\theta_0)}{\partial\theta}\right)'\right]\left[\frac{\partial^2\ln q_{P,T}^*(\bar{\theta}^1)}{\partial\theta\partial\theta'}\right]^{-1}$$

$$= \lim_{T\nearrow\infty}\left[\left(\frac{\partial g_T(x_T,\bar{\theta}^2)}{\partial\theta}\right)\hat{S}_T^{-1}\left(\frac{\partial g_T(x_T,\bar{\theta}^2)}{\partial\theta}\right)\right]$$

where $\bar{\theta}^1\in(\hat{\theta}_{LM},\theta_0)$ and $\bar{\theta}^2\in(\hat{\theta}_G,\theta_0)$.

4.2 A Limited Information Posterior and Its Properties

For notational convenience, write $q_T^*(x_T|\theta) = q_{P,T}^*(x_T,\theta)$ the limited information likelihood (4.6) based on the sample x_T. Let $\varphi(\theta)$ be a prior density of θ. Then, a posterior can be derived from the Bayes' rule

$$(4.9)\qquad \varphi_T(\theta|x_T) = q_T^*(x_T)^{-1}\{\varphi(\theta)\cdot q_T^*(x_T|\theta)\}$$

where $q_T^*(x_T) = \int_\Theta\varphi(\theta)q_T^*(x_T|\theta)d\theta$, a normalizing factor.

As $q_T^*(x_T|\theta)$ is used as a finite sample analogue of $q_P^*(\theta)$ in (4.5), it is meaningful to study the behavior of $\varphi_T(\theta|x_T)$ in the large-sample context. Also, as the GMM is well justified for asymptotic inference, we are interested in the asymptotic properties of $\varphi_T(\theta|x_T)$. As can be shown in the following, the posterior $\varphi_T(\theta|x_T)$ in (4.9) is asymptotically equivalent to the limited information posterior $\pi_T^*(\theta|x_T)$ from the optimal GMM in (3.25).

Recall that we have a `grand' parameter space Θ in which all the likelihoods and posteriors considered in this paper are defined. The likelihood $q_T(x_T(\cdot)|\cdot)$ is assumed to be jointly measurable $\mathcal{F}\times\mathcal{G}$. Also, $\varphi_T(\theta|x_T(\cdot))$ is jointly measurable $\mathcal{G}\times\mathcal{F}$. Define

$$P_T^{\varphi}(G;\omega)=\int_G \varphi_T(\theta|X_T(\omega))d\theta$$

for any $G\in\mathcal{G}$ and $\omega\in\Omega$. Then $P_T^{\varphi}(\cdot;\omega)$ is a probability measure on Θ for every $\omega\in\Omega$, and $P_T^{\varphi}(G;\cdot)$ is a random variable for each $G\in\mathcal{G}$.

Let $L_T(\theta;\omega)=\log q_T(x_T(\omega)|\theta)$, the log-likelihood of θ. Let $N(\hat{\theta}_T;\delta_T)$, $T=1,\ldots,\infty$, be a shrinking neighborhood as is defined in Section 3. Assume that the log-likelihood $L_T(\theta)$ is twice differentiable with respect to θ in $\bigcup_{T=1}^{\infty}N(\hat{\theta}_T;\delta_T)$. Denote by $L_T''(\theta)$ the second derivative of the log-likelihood. Notice that

$$[-L_T''(\hat{\theta}_T)]^{-1}=\left[T\left(\frac{\partial g_T(x_T;\hat{\theta})}{\partial\theta}\right)\hat{S}_T^{-1}\left(\frac{\partial g_T(x_T;\hat{\theta})}{\partial\theta}\right)\right]^{-1}\equiv V_T(\hat{\theta}_G). \tag{4.10}$$

Now, consider the following conditions (D1) and (D2).

(D1)(a) Let $M_T(\hat{\theta}_T(\omega);\delta_T)=\sup_{\theta\in N(\hat{\theta}_T;\delta_T)}\|[L_T''(\hat{\theta}_T)]^{-1}[L_T''(\theta)-L_T''(\hat{\theta}_T)]\|$. There exists a positive sequence $\{\delta_T\}_{T=1}^{\infty}$ such that $\lim_{T\uparrow\infty}P[M_T(\hat{\theta}_T(\omega);\delta_T)<\epsilon]=1$ for each $\epsilon>0$. (b) For δ_T satisfying (a) the absolute value of each element of the vector $[-L_T''(\hat{\theta}_T)]^{1/2}\delta_T$ tends to infinity as $T\uparrow\infty$ in P-probability.

(D2) Let $\varphi_T(\theta|x_T)$ be the posterior as defined in (4.9). For δ_T satisfying (D1),

$$\int_{\Theta\setminus N(\hat{\theta}_T;\delta_T)}\varphi_T(\theta|Y_T)d\theta\longrightarrow 0$$

as $T\uparrow\infty$, i.e., θ concentrates in $N(\hat{\theta}_T;\delta_T)$ as $T\uparrow\infty$.

Notice that Conditions (D1) and (D2) are similar to Conditions (C1) and (C2) that are applied to different objects.

(D3) The prior density ¼(µ) is continuous in £ and 0 < ¼($µ_0$) < 1 .

We can show that a posterior formed from an I-projection likelihood and a prior satisfying (D3) is asymptotically normal:

Theorem 4 Assume that (D1) and (D2) are satis¯ed for $q^¤_T(¢;¢)$ and '(¢). Also, assume that (D3) is satis¯ed. Then, for each (a;b),

$$\mathrm{Z}_{J_{T\,ab}} '_T(µjx_T)dµ \;\; ¡ ! \;\; \mathrm{Z}_a^b Á(z)dz$$

in P-probability where J_{Tab} = fµ : [¡ $L^{00}_T(\hat{µ}_T)]^{1=2}$(µ ¡ $\hat{µ}_T$) 2 (a;b)g.

Theorem 4 states that under some conditions the posterior distribution of the parameter µ is asymptotically normal with the ¯rst moment of the distribution equal to $\hat{µ}_T$ and the second moment [¡ $L^{00}_T(\hat{µ}_T)]^{¡1}$:

$$µjx_T \overset{a}{»} N(\hat{µ}_T;[¡ L^{00}_T(\hat{µ}_T)]^{¡1}):$$

Since from (4.10) [¡ $L^{00}_T(\hat{µ}_T)]^{¡1}$ = $V_T(\hat{µ}_G)$; the result in Theorem 4 shows that the posterior $'_T(µjx_T)$ in (4.9) is asymptotically equivalent to the limited information posterior $¼^¤_T(µjx_T)$ under the optimality condition (3.17):

Proposition 2 Assume that assumptions of Proposition 1 and Theorem 4 hold. Then, under (3.17) it is true that

$$\left| \mathrm{Z}_{J^1_{T\,ab}} '_T(µjx_T)dµ \;\; ¡ \;\; \mathrm{Z}_{J^2_{T\,ab}} ¼^¤_T(µjx_T)dµ \right| ¡ ! \;\; 0$$

in P-probability where J^1_{Tab} = fµ : [$L^{00}_T(\hat{µ}_T)]^{1=2}$(µ ¡ $\hat{µ}_T$) 2 (a;b)g and J^2_{Tab} = fµ : [$V_T(\hat{µ}_T)]^{¡1=2}$(µ ¡ $\hat{µ}_T$) 2 (a;b)g.

Asymptotic equivalence of $'_T(µjx_T)$ in (4.9) and $¼^¤_T(µjx_T)$ in Section 3 proves the validity of each of the two by the other. Asymptotic su±ciency of the statistic $s(x_T)$ = ($\hat{µ}_T$; $L^{00}_T(\hat{µ}_T)$) for the posterior $'_T(µjx_T)$ also follows by the same way as in Corollary 2.

5 Selection of Models and Moment Conditions

We begin with a general setup of a model selection problem. Let M be a family of candidate models for x_T. Denote by $m_0 \in M$ the true model for x_T and $p_T(\theta; x_T)$ the true p.d.f. of x_T. A model $m_i \in M$ is associated with a parameter space Θ^i of dimension q_i for $i \in I$ where $I = \{1, \ldots, I\}$, and $\Theta^i \subset \Theta$ for $i \in I$. Assume that for each m_i a family $Q_T^i(\theta^i; x_T)$ of distribution functions, with a density $q_T^i(\theta^i; x_T)$, is defined on the measurable space $(\Theta, G) \times (\Omega, F)$. For our GMM framework, each m_i corresponds to a set of moment conditions as in (2.1). Also, each $q_T^i(\theta^i; x_T)$ is an I-projection likelihood as is defined in Section 5 corresponding to the set of moment conditions. Notice that, different from Section 4, we do not have the superscript `$*$' for the I-projection likelihood $q_T^i(\theta^i; x_T)$ for notational convenience. We assume some regular conditions on the density $q_T^i(\theta^i; x_T(\omega))$ defined on $\Theta \times \Omega$:

Assumption 3 (a) For each T, $\int_\Omega \log(p_T)dP_T$ exists and $|\log q_T^i(\theta^i; x_T)| \le \mu^i(x_T)$ for all $\theta^i \in \Theta^i$ for $i \in I$, where μ is integrable with respect to P_T. (b) For each T and m_i $\int_\Omega \log(p_T/q_T^i)dP_T$ has a unique maximum at $\theta^i_{T,0} \in \Theta^i$. (c) For each T, $\int_\Theta \log(p_T)dP_T$ exists and $|\log q_T^i(\theta^i; x_T)| \le \kappa^i(\theta)$ a.e. in Ω, where κ^i is integrable with respect to P_T. (d) For each T $\int_\Theta \log(p_T/q_T^i)dP_T$ has a maximum at $i = i^*(T)$ for an $i^* \in I$.

5.1 BIC in the Limited Information Framework

A natural approach to model selection in the Bayesian framework is to choose a model m_i for which the posterior probability is the largest. Thus, let $\Pr(m_i|X_T)$ be the posterior probability that m_i is true. By the Bayes' rule

$$\Pr(m_i|x_T) = \frac{q_T(x_T|m_i)\Pr(m_i)}{\sum_{j \in I} q_T(x_T|m_j)\Pr(m_j)} \tag{5.1}$$

where $\Pr(m_i)$ is the prior probability that m_i is true and $q_T(x_T|m_j) = q_T^i(x_T)$. If we assume that $\Pr(m_j)$ is the same for all j, the model selection rule is to choose m_i for which $q_T(x_T|m_j)$, or

$$q_T(x_T|m_j) = \int q_T(x_T|\theta^j; m_j)\varphi(\theta^j|m_j)d\theta^i = E^{m_j}[q_T(x_T|\theta^j)] \tag{5.2}$$

is the largest, where $'(\mu^j jm_j)$ is the prior density associated with the model m_j. Phillips (1996) provides another dimension of justification of Bayesian approach for model selection based on the notion of a Bayesian model measure.

The criterion (5.2) involves computation of an integral of $q_T \pounds\ '$ with respect to μ^i in $\mathbb{R}^{q_i}$. Certainly this computation is not easy even with a very fast computer. Also, the choice of the range of μ is another problem for the computation. Chib (1995), among others, applies the Gibbs sampling method to compute the marginal likelihood $q_T(X_T jm^j)$. The Gibbs sampling method is a powerful approach to computing a density that can be written as a product of several conditional densities. Sometimes, however, the result from the Gibbs sampler is sensitive to the setup of the simulation or `sampling'. Also, it is necessary that all integrating constants of the full conditional distributions in the Gibbs sampler be known (p.1314, Gibbs (1995)). On the other hand, the marginal likelihood $q_T(X_T jm^j)$ itself depends on the prior density, so that model selection based on a direct computation of $q_T(X_T jm^j)$ yields different results depending on the choice of the prior.

In the following we provide an approximation to the integral in (5.2) that is valid for large sample analysis. It is computationally simple to handle and yet has sound theoretical justification.

Lemma 6 Assume that the prior $'(\mu)$ is continuous in $\pounds$ and bounded at μ_0. Then, under the assumptions (D1) and (D2)

$$(5{:}3)\qquad \begin{aligned} &\log E^{m_j}[q_T^j(x_T j\mu^j)] \\ &= \log(q_T(x_T j\hat{\mu}_T))\ ¡\ (1{=}2)\log(j[\ ¡\ L_T^{00}(\hat{\mu})]j) + (q{=}2)\log(2¼) + \log('(\mu_0)) + R_0; \end{aligned}$$

where R_0 is of $o_p(1)$.

From Lemma 6, an approximation of the criterion (5.2) is

$$\text{Choose the model } j \text{ that maximizes}$$
$$(5{:}4)\qquad \log(q_T(x_T j\hat{\mu}_T))\ ¡\ (1{=}2)\log(j[\ ¡\ L_T^{00}(\hat{\mu})]j){:}$$

Now, consider the criterion (5.4) for the optimal GMM estimate. From (4.6)

$$q_T(\mu) = K_T \exp \mathbf{n}\ \cdot\ ¢\, T g_T(x_T; \mu)^0 \hat{S}_T^{¡1} g_T(x_T; \mu)\ \mathbf{o}\ :$$

Also, from (4.10)

$$[-L''_T(\hat{\theta}_T)]^{-1} = V_T(\hat{\theta}_G)$$

Therefore, from (4.6), (3.23) and (4,10) the criterion (5.4) for the optimal GMM estimate is such that

Choose the model j that maximizes

$$(5.5) \qquad -\kappa \cdot T g_T(x_T;\hat{\theta})'\hat{S}^{-1}g_T(x_T;\hat{\theta}) + \log K_T + (1/2)\log(|V_T(\hat{\theta}_G)|).$$

Notice that the criterion (5.5) is for selecting the moment for GMM. The constants κ and K_T and $V_T(\hat{\theta}_G)$ can be obtained for a given $g_T(x_T;\theta)$ and $\hat{S}$.

As an example, consider the case of linear regression

$$(5.6) \qquad x_{1t} = x'_{2t}\theta + \varepsilon_t$$

with the following moment condition

$$(5.7) \qquad E[x_{2t}\varepsilon_t] = 0.$$

In this case,

$$g_T(x_T;\theta) = \frac{1}{T}\sum_{t=1}^{T} x_{2t}(x_{1t} - x'_{2t}\theta).$$

Notice that

$$g_T(x_T;\hat{\theta}) = \frac{1}{T}\sum_{t=1}^{T} x_{2t}\hat{\varepsilon}_t = 0$$

where $\hat{\varepsilon}_t = x_{1t} - x'_{2t}\hat{\theta}$. Also, for the model (5.6)-(5.7) we can show that

$$V_T(\hat{\theta}) = \hat{\sigma}^2[X'_2X_2]^{-1};$$

$$K_T = \hat{\sigma}^{-T}$$

where $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t$ and $X_2 = (x'_{21},\ldots,x'_{2T})'$. Therefore, for the model (5.6)-(5.7) the moment selection criterion (5.5) is as follows

Choose the model j that maximizes

$$(5.8) \qquad (T - q_j)\log\hat{\sigma}^2/T + \log(|X_2^{(j)'}X_2^{(j)}|)/T.$$

Notice that the criterion (5.8) is exactly the same as the Bayesian model selection criterion for the regression (5.6) with $\varepsilon_t \sim i.i.d.N(0;\sigma^2)$ instead of the moment condition (5.7) which is derived in Kim (1998).

5.2 Consistency

We can show that the above criterion (5.2) leads to the choice of the true model with unit probability. Denote by $p_T(x_T|\theta)$ the true likelihood. Recall that $\pi_T(\theta|x_T)$ is the true posterior density of θ from $p_T(x_T|\theta)$ and a prior $\pi(\theta)$. Denote by $\varphi^i_T(\theta|x_T)$ the posterior density of θ from the likelihood q^i_T and a prior $\varphi^i(\theta^i)$.

The following theorem shows that the decision rule (5.2) chooses the true model m_0 in a set of alternatives under some condition:

Theorem 5 Assume that for $i \in I$

$$(5.9) \qquad \int \ln\bigl((p_T\pi)/(q^i_T\varphi^i)\bigr)\pi_T d\theta - \int \ln\bigl(\pi_T/\varphi^i_T\bigr)\pi_T d\theta \geq 0$$

with equality only if $\varphi^i_T = \pi_T$. Then, for any $i \in I$,

$$(5.10) \qquad E^{m_i}[q^i_T(x_T|\theta^i)] \leq E^{m_0}[p_T(x_T|\theta)];$$

with equality only if $q^i_T\varphi^i = p_T\pi$. That is, $\Pr[m_i|X_T] \leq \Pr[m_0|x_T]$ for all $i \in I$ with equality only if $m_i = m_0$.

Notice that (5.10) implies that given $x_T(\omega)$ for a $\omega \in \Omega$ the decision rule (5.2) chooses the true model under the condition (5.9).

The condition (5.9) generally holds in some probabilistic sense for a sufficiently large sample. For example, we can show that under (D1)-(D3) the condition (5.9) holds asymptotically.

5.3 Size Adjustment and Power Comparison

References

[1] Andrews, D.W.K. and B. Lu (1999) “Consistent Model and Moment Selection Criteria for GMM Estimation with Application to Dynamic Panel Data Models," Mimeo, Yale University.

[2] Andrews, D.W.K. (1991). “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," Econometrica 59:817-58.

[3] Andrews, D.W.K. and J.C. Monahan (1992). \An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," Econometrica 60:953-66.

[4] Chen, S. X. (1993) \On the Accuracy of Empirical Likelihood Confidence Regions for Linear Regression Models", Ann. Inst. Statist. Math., 45, 621-637.

[5] Chen, S. X. and P. Hall (1993) \Smoothed Empirical Likelihood Confidence Intervals for Quantiles," The Annals of Statistics, 21, 1166-1181.

[6] Cover, T. M. and J. A. Thomas (1991), Elements of Information Theory, New York: J. Wiley & Sons, Inc.

[7] Csiszar, I. (1975) \I-divergence Geometry of Probability Distributions and Minimization Problems," The Annals of Probability, 3, 1, 146-158.

[8] DiCiccio, T., Hall, P. and J. Romano (1989) \Comparison of Parametric and Empirical Likelihood Functions," Biometrika, 76, 465-476.

[9] DiCiccio, T., Hall, P. and J. Romano (1991) \Empirical Likelihood Is Bartelett-correctable," The Annals of Statistics, 19, 1053-1061.

[10] DiCiccio, T. and J. Romano (1989) \On Adjustments to the Signed Root of the Empirical Likelihood Statistics," Biometrika, 76, 447-456.

[11] Gallant, A.R. (1987). Nonlinear Statistical Models, New York: Wiley.

[12] Green, E.J. and W.E. Strawderman (1994), \A Bayesian Growth and Yield Model for Slash Pine Plantations," Department of Natural Resources and Department of Statistics, Rutgers U., New Brunswick, NJ.

[13] Hall, P. (1990) \Pseudo-likelihood Theory for Empirical Likelihood," The Annals of Statistics, 18, 121-140.

[14] Hansen, L.P. (1982). \Large Sample Properties of Generalized Method of Moments Estimators," Econometrica, 50, 1029-1054.

[15] Jaynes, E. T. (1982a) Papers on Probability, Statistics and Statistical Physics, Dordrecht, Netherlands: Reidel.

[16] Jaynes, E. T. (1982b), \On the Rationale of Maximum-Entropy Methods," Proceedings of the IEEE, 70, 939-952.

[17] Kim, J. Y. (1994) \Bayesian Asymptotic Theory in an AR(1) with a Possible Unit Root," Econometric Theory, 10, 764-773.

[18] Kim, J. Y. (1998) \Large Sample Properties of Posterior Densities in a Time Series with Nonstationary Components, Bayesian Information Criteriorn, and the Likelihood Principle," Econometrica, 66, 2, 359-380.

[19] Kim, J. Y. (2000) \Empirical Likelihood Methods and the Generalized Method of Moments," manuscript.

[20] Kitamura, Y. (1997) \Empirical Likelihood Methods with Weakly Dependence Processes," The Annals of Statistics, 25, 5, 2084-2102.

[21] Kitamura, Y., and P.C.B. Phillips (1997). \Fully Modified IV, GIVE and GMM Estimation with Possibly Non-stationary Regressors and Instruments," Journal of Econometrics, 80, 1, 85-123.

[22] Kwan, Y.K. (1999) \Asymptotic Bayesian Analysis Based on a Limited Information Estimator," Journal of Econometrics, 88, 1, 99-121.

[23] Newey, W.K., and K.D. West. (1987). \A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," Econometrica 55:703-8.

[24] Owen, A. (1988) \Empirical Likelihood Ratio Confidence Intervals for a Single Functional," Biometrika, 75, 237-249.

[25] Owen, A. (1991) \Empirical Likelihood for Linear Models," The Annals of Statistics, 19, 1725-1747.

[26] Quin, J. (1993) \Empirical Likelihood in Biased Sample Problems," The Annals of Statistics, 21, 1182-1196.

[27] Quin, J. and J. Lawless (1994) \Empirical Likelihood and General Estimating Equations," The Annals of Statistics, 23, 300-325.

[28] Shore, J. E. and R. W. Johnson (1980), \Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy," IEEE Transactions, IT-26, 1, 26-37.

[29] Sims, C. A. (1988) \Bayesian Skepticism on Unit Root Econometrics," Journal of Economic Dynamics and Control, no.12, 463-474.

[30] Sims, C. A. and H. Uhlig (1991) \Understanding Unit Rooters: A Helicopter Tours," Econometrica, 59, 1591-1599.

[31] Soofi, E. S. (1994), \Capturing the Intangible Concept of Information," Journal of the American Statistical Association, 89, 428, 1243-1254.

[32] Zellner, A. (1996) \Bayesian Method of Moments/Instrumental Variable (BMOM/IV) Analysis of Mean and Regression Model," in Modelling and Prediction Honoring Seymour Geisser. by Lee, J.C., Johnson, W.C., Zellner, A. (eds.), Springer, New York, 61-74.

[33] Zellner, A. (1997) \The Bayesian Method of Moments (BMOM): Theory and Application," in Advances in Econometrics by Fomby, T., Hill, R.C. (eds.).

[34] Zellner, A. (1998) \The Finite Sample Properties of Simultaneous Equations' Estimates and Estimators: Bayesian and Non-Bayesian Approaches," Journal of Econometrics, 83, 185-212.

[35] Zellner, A. and R. A. Highfield (1988), \Calculation of Maximum Entropy Distributions and Approximation of Marginal Posterior Distributions," Journal of Econometrics, 37, 195-210.