CHAPTER 3

Model and Moment Selection in Moment Inequality Models

3.1. Introduction

Similar to the selecting the moment conditions in GMM, there is a moment and model selection problem in moment inequality models. Suppose there are p candidate moment inequalities

$$Em_j(X, \theta) \geq 0, j = 1, ..., p$$

with a k-dimensional parameter vector $\theta = (\theta_1, ..., \theta_k)^T$ that belongs to the parameter space $\Theta_1 \times \cdots \times \Theta_k$. The moment selection problem refers to selecting the best subset of the moment inequalities among all the possible candidates, while the model selection procedure addresses the problem of selecting the best model that is characterized by setting some components of the parameter to be zero. Such a candidate model can be a parameter subspace like $\{0\} \times \Theta_2 \times \cdots \times \Theta_k$. Therefore, the moment/model selection procedure produces a combination of moment inequalities and a parameter subspace. Consider the following example:

Example 3.1.1 (Interval censored regression)**.** (See, e.g., Example 1 of CHT 2007.) Let Y be a real valued random variable which lies in $[Y_1, Y_2]$ almost surely; Y_1 and Y_2 are observed random variables, but Y is not observed. (Sometimes one may assume that $Y_2 = Y_1 + 1$ as in the case when Y_1 is the recorded integer part of Y.) . Assume that

$$Y = X^T\theta + \epsilon$$

where X is a regressor vector. In addition, there exists a observable random vector Z such that $E(\epsilon|Z) = 0$. Here Z can be part of the regressors in X or a set of instrumental variables when X is endogenous. It the follows that $E(ZY) = E(ZX^T)\theta$. Due to $Y_1 \leq Y \leq Y_2$, we then have moment inequalities

$$EZ(Y_2 - X^T\theta) \geq 0, EZ(X^T\theta - Y_1) \geq 0 \tag{3.1}$$

In this example, the moment selection problem can correspond to selecting the instrumental variables (components of Z), while the model selection problem is related to selecting the useful explanatory variables (components of X) that have nonzero regression coefficients.

□

A similar selection problem in point identified case was previously considered by Andrews and Lu (2001), where they applied their approach to dynamic panel data models. We provide a similar example as follows.

Example 3.1.2 (Dynamic Panel Data)**.** Consider a dynamic panel data model

$$y_{it} = z_{it}'\theta_t + \eta_i + v_{it}$$

Here y_{it} is the dependent variable, censored between $y_{it}^L \leq y_{it} \leq y_{it}^U$. Hence instead of y_{it}, econometricians can only observe y_{it}^L, y_{it}^U. Also, v_{it} is an unobserved error, η_i is an unobserved individual effect, and θ_t are unknown parameters of interest. The distributions of η_i and v_{it} are not specified, and hence the limited information likelihood based on some moment conditions given below may be preferable. All of the random variables are assumed to be independent across individuals i.

The regressor $z_{it} = (X_{it}, w_{it})$ is an observed vector, where $X_{it} = (x_{it-1}, ..., x_{it-L})$ includes L lags of some covariates that may be exogenous, predetermined, or endogenous, where $L \geq 0$. The true lag length L_0 may be unknown. In addition, z_{it} also includes exogenous variables w_{it}, which are contained in an observed vector Z_{it}. The vector Z_{it} may also contain variables that do not enter the regression function. Such variables can be employed as instrumental variables. The following assumptions are imposed to derive the moment conditions.

$E\eta_i = Ev_{it} = 0, \forall t = 1, ..., T.$

$Ev_{it}Z_{it+1} = ... = Ev_{it}Z_{iT} = 0, \forall t = 1, ..., T.$

We may further partition Z_{it} into variables that are either uncorrelated with η_i or not, and achieve additional moment conditions. We do not do so here for simplicity. The moment conditions implied are

$$\begin{aligned}
&EZ_{it}(y_{it} - y_{it-1}) = EZ_{it}(z_{it}'\theta_t - z_{it-1}'\theta_{t-1}), \forall t = 1, ..., T \\
&Ey_{it} = Ez_{it}'\theta_t, \forall t = 1, ..., T.
\end{aligned}$$

Assume $\{Z_{it} : i = 1, ..., n,\}$ is generated from a distribution with support $supp(Z_t)$, which is compact for each t. Since one can always transform Z_{it} into $Z_{it} - \inf supp(Z_t)$, hence without loss of generality, we assume $Z_{it} \geq 0$. As y_{it} is censored in $[y_{it}^L, y_{it}^U]$, we have moment inequalities

$$\begin{aligned}
&EZ_{it}(y_{it}^L - y_{it-1}^U) \leq EZ_{it}(z_{it}'\theta_t - z_{it-1}'\theta_{t-1}) \leq EZ_{it}(y_{it}^U - y_{it-1}^L), \forall t = 1, ..., T \\
&Ey_{it}^L \leq Ez_{it}'\theta_t \leq Ey_{it}^U, \forall t = 1, ..., T.
\end{aligned}$$

Setting different lag coefficients to zero yields models with different number of lags in X_{it}. Therefore, the model selection refers to selecting the lagged variables of X_{it}. In addition, θ_t

also contains covariates of other variables w_{it}. If the effects of some particular components of w_{it} are of special interest, we should always keep the corresponding coefficients in the model. $\square$

As illustrated by the previous example, by allowing the dimension of the parameter space to change, we consider the case where the parameter vector may incorporate several models. By setting different elements of θ equal to zero, one obtains different models. Andrews and Lu (2001) also gave another example where a model may have structural breaks in the parameters. For example, when $t \leq t_0$, $\theta_t = \theta_0$, which is the pre-break value; when $t > t_0$, $\theta_t = \theta_0 + d\theta_1$, which adds a post-break deviation to the pre-break value. Here t_0 may be unknown. There may be multiple time breaks, and one can stack the break values into a single vector of parameters $\theta = (\theta_0, d\theta_1, ...)$. Different sets of post-break deviations can denote changes at different times. If the post-break deviations are set equal to zero, then one obtains the model with no structural breaks at that time.

3.2. Posterior Setup

Suppose we have p candidate moment inequalities

$$Em_j(X, \theta) \geq 0, j = 1, ..., p$$

with a k-dimensional parameter vector $\theta = (\theta_1, ..., \theta_k)^T \in \Theta_1 \times \cdots \times \Theta_k$. Here the possible moment inequalities and corresponding subsets of the parameter space are known. What is not known is which ones are the best.

Instead of selecting the moment inequalities and the parameter subspace as two separate procedures, I select them as a combination simultaneously. The selection procedure is based on the posterior probabilities. I assign prior probabilities to each candidate moment/model, and then derive the posterior probabilities based on the limited information likelihood described previously in Chapter 2, by integrating out the structural and nuisance parameters (θ, λ).

Let us define a combination $C_s = (M_{s_1}, \Theta_{s_2})$, with a vector index $s = (s_1, s_2)$, $s_1 \in \{1, 2, ..., 2^p - 1\}$, and $s_2 \in \{1, ..., 2^k\}$. Here M_{s_1} denotes a subset of moments, for instance, $M_{s_1} = \{m_1\}$, or $M_{s_1} = \{m_1, m_2\}$, etc. Then there are $2^p - 1$ number of such possible subsets. In addition, we denote by Θ_{s_2} as the parameter subspace corresponding to the selected model. By definition, Θ_{s_2} is the subset of vectors with one or more components fixed to be zero. There are 2^k possible Θ_{s_2}'s. (Notice that we can select none of the parameters, in which case the model is a reduced model, for example, in Cox proportional hazard model, if all the parameters are set to be zero, we get the baseline model.) The combination C_s combines both the candidate moment functions and the parameter subspace together. When selecting a subset of moment inequalities, we also specify a subspace of the structural parameter.

Example 3.2.1 (Example 3.1.1 continued)**.** Let $\Theta_1 \times \Theta_2$ be the parameter space for (θ_1, θ_2), chosen large enough so that $\{(\theta_1, \theta_2) : 0.2 \leq \frac{1}{3}\theta_1 + \theta_2 \leq 0.4, -0.1 \leq \theta_2 \leq 0.1\} \subset \Theta_1 \times \Theta_2$. A scope of candidate combinations can be any of the following:

$$
\begin{array}{ll}
\{E(Z_1X^T\theta - Z_1Y_1)\}, & \Theta_1 \times \Theta_2 \\
\{E(Z_1X^T\theta - Z_1Y_1), E(Z_1Y_2 - Z_1X^T\theta)\}, & \Theta_1 \times \Theta_2 \\
\{E(Z_2X^T\theta - Z_2Y_1)\}, & \{0\} \times \Theta_2 \\
\{E(Z_1X^T\theta - Z_1Y_1), E(Z_1Y_2 - Z_1X^T\theta), E(Z_2Y_2 - Z_2X^T\theta)\}, & \Theta_1 \times \{0\} \\
 & \vdots
\end{array}
$$

$\{E(Z_2Y_2 - Z_2X^T\theta)\},$ $\Theta_1 \times \Theta_2$

Definition 3.2.1. *A combination $C_s = (M_{s_1}, \Theta_{s_2})$ is compatible if and only if*

$$\inf_{\theta \in \Theta_{s_2}, \lambda \in [0,\infty)^m} ||EM_{s_1}(X,\theta) - \lambda||^2 = 0$$

where m denotes the number of candidate moment functions in M_{s_1}.

Assumption 3.2.1. *(i) $\Theta = \Theta_1 \times ... \times \Theta_k$ is compact.*
(ii) $\forall j = 1, ..., p$, $Em_j(X, .) : \Theta \to \mathbb{R}$ is continuous.

Lemma 3.2.1. *Under Assumption 3.2.1, the following statements are equivalent.*
(i) C_s is compatible.
(ii) $\Omega_s = \{\theta \in \Theta_{s_2} : EM_{s_1}(X,\theta) \geq 0\}$ is not empty.
(iii) For all positive definite V_0,

$$\inf_{\theta \in \Theta_{s_2}, \lambda \in [0,\infty)^m} (EM_{s_1}(X,\theta) - \lambda)^T V_0 (EM_{s_1}(X,\theta) - \lambda) = 0.$$

Let us partition the parameters θ and λ into "restricted" and "unrestricted" parts according to the biases of the selected and unselected moment functions. Formally, let

$$\lambda = EM(X,\theta)$$

where $M(X,\theta) = (m_1(X,\theta), ..., m_p(X,\theta))^T$, the vector of all the candidate moments, and $\theta = (\theta_1, ..., \theta_k)^T$, the vector of full parameters supported on $\Theta_1 \times ... \times \Theta_k$. Suppose a combination $C_s = (M_{s_1}, \Theta_{s_2})$ selects m moment conditions M_{s_1}, and leaves the rest of the moments (denoted by $M_{s_1}^c$ unused). It also selects a submodel parameterized by $\theta_s \in \Theta_{s_2}$, setting all the other components of θ, (which is denoted by θ_s^c) to be zero.

One can view model selection as placing a restriction on θ, while moment selection can be reviewed as placing a restriction on λ. Let λ_s be the subvector of λ corresponding to the selected moments. Let λ_s^c be the remaining components of λ corresponding to $M_{s_1}^c$. Then we have

$$EM_{s_1}(X, \theta_s) = \lambda_s, \lambda_s \geq 0$$

$$EM_{s_1}^c(X, \theta_s) = \lambda_s^c, \lambda_s^c \in \mathbb{R}^{p-m}$$

The bias λ_s for the selected moments is restricted to be nonnegative, while the bias λ_s^c for the unselected moments is left unrestricted. We thus have partitioned the moment functions into $M(X, \theta_s) = (M_{s_1}(X, \theta_s)^T, M_{s_1}^c(X, \theta_s)^T)^T$, and λ into $\lambda = (\lambda_s, \lambda_s^c)$. We put prior:

$$\begin{aligned} p(\lambda_s^c | C_s) &\sim N_{p-m}(0, \Sigma) \\ p(\lambda_s | C_s) &\sim Exp(\psi) \end{aligned} \tag{3.2}$$

where N_{p-m} denotes the $p - m$ dimensional multivariate normal distribution, assumed to be a priorily independent so that $\Sigma = diag\{\sigma_1^2, ...\sigma_{p-m}^2\}$. $Exp(\psi)$ is the exponential distribution with parameter ψ, as in Chapter 2.

We include both selected M_s and unselected M_s^c to construct the limited information likelihood, which depends only on the unrestricted θ_s since $\theta_s^c = 0$.

$$L(X^n | \theta_s, \lambda, C_s) = \frac{1}{\sqrt{\det(\frac{2\pi}{n} V)}} e^{-\frac{n}{2}(\bar{M}(\theta_s) - \lambda)^T V^{-1} (\bar{M}(\theta_s) - \lambda)} \tag{3.3}$$

where $\bar{M}(\theta_s) = \frac{1}{n}\sum_{i=1}^{n} M(X_i, \theta_s)$. The prior of C_s is imposed. Then the posterior of C_s can be obtained by integrating out θ_s and $\lambda = (\lambda_s^T, \lambda_s^{cT})^T$, which is proportional to the "integrated likelihood":

$$p(C_s|X^n) \propto \iint_{\Theta_{s_2}\times[0,\infty)^m\times\mathbb{R}^{p-m}} L(X^n|\theta_s,\lambda,C_s)p(\theta_s|C_s)p(\lambda_s|C_s)p(\lambda_s^c|C_s) \cdot p(C_s)d\theta_s d\lambda_s d\lambda_s^c \tag{3.4}$$

3.3. Posterior Consistency of Liao and Jiang (2010)

This chapter is actually written based on the materials of my published paper Liao and Jiang (2010, hereafter LJ), which considered the moment/model selection problem in interval censored regression problem. Due to the limited space, in this section I only briefly go over the main results, and details can be found in the published paper. The selection problems considered in LJ consist of two parts: selecting the compatible combinations of moment /model, and among the compatible combinations, selecting the "optimal" one. The optimal compatible combination is defined as the one with maximal $\dim(M_{s_1}) - \dim(\Theta_{s_2})$. This is because it is desirable that the optimal combination should contain as many moment inequalities as possible, since intuitively the more moment inequalities, the smaller the identified region, and hence the more information we have about the parameter. Meanwhile, it is required that the model should be as simple as possible, since simpler models are easier to interpret.

The selection procedure was known as the maximal posterior criterion (MPC), by maximizing the posterior of the combinations. In order for the MPC procedure to asymptotically select the optimal combination, the variance covariance matrix Σ in prior (3.2) should depend on the sample size:

$$\sigma_i^2 = \sigma_n^2, \text{ where } \sigma_n^2 \to \infty \text{ but not exponentially fast}$$

$$p(\theta_s|C_s) \sim N_t(0, n\sigma_n^2 I_t), \text{ where } t = \dim(\theta). \tag{3.5}$$

Under further regularity assumptions (see LJ Assumptions 4.2-4.6), it can be shown that

Theorem 3.3.1 (LJ Theorem 4.3, MPC Consistency)**.** *Let*

$$C^* = \arg\max_{C_s} p(C_s|X^n)$$

where $p(C_s) > 0$ *does not depend on* n *for each* C_s*, and priors on specified by (3.2) and (3.5) for interval censored regression model, with probability approaching one,* $C^* = (M_{s_1}, \Theta_{s_2})$ *then is compatible and has the largest* $\dim(M_{s_1}) - \dim(\Theta_{s_2})$*.*

We can impose the following assumption which is similar to Assumption IDbc in Andrews and Lu (2001):

Assumption 3.3.1. *The true model and moment combination is the unique combination of* (M, Θ)*, such that it has the maximal* $\dim(M) - \dim(\Theta)$*.*

If this assumption holds, the previous theorem implies that by maximizing the combination posterior, we can asymptotically select the actual true combination of model and moments. In particular, when the dimension of the true parameter is fixed, and we are only selecting the corresponding moment inequalities that are satisfied by the true parameter of interest, Assumption 3.3.1 becomes: There exists a unique set of maximal number of moment inequalities that are satisfied by the true parameter. In this case, MPC consistently selects all the true moment inequalities.

3.4. More Reliable Setting

Note that Assumption 3.3.1 plays the central role of identifying the “true” moment inequalities and the parameter space. When the parameter space is fixed, meaning that the true parameter is assumed to exist, the selection problem then becomes to select the “true” moment inequalities that are satisfied by the true parameter. In this case, MPC procedure in LJ asymptotically selects the true moment inequalities.

However, in practice Assumption 3.3.1 is not satisfied naturally, and when it is not, MPC may select a set of incorrect moment inequalities with probability approaching one. The problem is that, the moment inequalities that are not satisfied by the true parameters can still be compatible.

Example 3.4.1. Suppose the true parameter $\theta_0 = 1.7$, with parameter space $\Theta = [0, 5]$. Consider the following moment inequalities

$$\theta \geq EY_1 (= 1.5) \tag{3.6}$$

$$\theta \leq EY_2 (= 2) \tag{3.7}$$

$$\theta \geq EY_3 (= 3) \tag{3.8}$$

$$\theta \geq EY_4 (= 3.5) \tag{3.9}$$

Apparently, only (3.6) and (3.7) are satisfied by θ_0, which correspond to interval $[1.5, 2]$. However, the MPC procedure will select all the other three inequalities (3.6), (3.8) and (3.9), because their combination has the maximal (3) number of inequalities, and the corresponding interval in Θ is $[3.5, 5]$. In this example, the compatible interval defined by the maximal number of inequalities does not contain the true parameter.

□

The following theorem shows that, when Assumption 3.3.1 is relaxed and priors (3.2) and $p(\theta_s)$ are data-independent, the posterior probability of incompatible combinations (where the corresponding identified region is empty) is still exponentially small, as opposed to compatible combinations, whose posterior is proportional to a positive constant multiplied by the combination prior.

Assumption 3.4.1. *For any compatible $C_s(M_{s_1}, \Theta_{s_2})$, $p(\theta|C_s)$ is uniformly bounded on Θ_{s_2}.*

Theorem 3.4.1. *Under Assumption 3.2.1, 3.4.1, the parameter priors are given in (3.2), and $V > 0, \Sigma > 0$ are fixed,*

(1) *If C_s is compatible and $p(C_s) > 0$, in probability*

$$\liminf_{n\to\infty} p(C_s|X^n) > 0$$

(2) *If C_s is not compatible, then for some $\alpha > 0$,*

$$p(C_s|X^n) = o_p(e^{-\alpha n})p(C_s)$$

The MPC procedure consistently selects the maximal $\dim(M) - \dim(\Theta)$, because the data-size-dependent priors (3.5) for unrestricted parameters (θ_s, λ_s^c), corresponding to unselected moments and selected parameters, have very thick tails asymptotically, which force the posterior of combinations with many unrestricted parameters to be very small. As illustrated in the previous example, however, the true parameter (if any) may satisfy only a few inequalities. Therefore a more reliable setting is to use data-independent prior for unrestricted parameters. If the prior of (θ_s, λ) is jointly specified as $p_{\theta,\lambda}(\theta_s, \lambda|C_s)$, which does not depend on the sample size n, the selection among compatible combinations using posterior probabilities is no longer consistent in terms of selecting the maximal $\dim(M_{s_1}) - \dim(\Theta_{s_2})$, because when Assumption 3.3.1 is relaxed, we fail to identify the true set of inequalities. Suppose $C_s(M_{s_1}, \Theta_{s_2})$ is a compatible combination. Write λ as the parameter that satisfies the moment condition $\lambda = EM(X, \theta)$, and λ is ordered and partitioned as (λ_s, λ_s^c), then λ takes its value in $\Lambda = [0, \infty)^m \times \mathbb{R}^{p-m}$, where m denotes the dimension of M_{s_1}, i.e., the number of selected moments. We impose the following regularity conditions on the parameter prior. Note that Condition (i) can be achieved if $p_{\theta,\lambda}(\theta, \lambda|C_s)$ is uniformly bounded by a constant $k > 0$ on $\Theta_{s_2} \times \Lambda$, given that Θ_{s_2} is bounded.

Assumption 3.4.2. *(i) For any C_s, and $\theta \in \Theta_{s_2}$, there exists $g(\theta) > 0$ satisfying $\int_{\Theta_{s_2}} g(\theta)d\theta < \infty$, such that $p_{\theta,\lambda}(\theta, \lambda|C_s) \leq g(\theta)$ for all $\lambda \in \Lambda$.*
(ii) For any fixed θ, $p_{\theta,\lambda}(\theta, \lambda|C_s)$ is continuous with respect to λ on Λ.

The following theorem shows that, in this case, the posterior heavily depends on the prior of combinations $p(C_s)$, which may be obtained by, if any, a priori information about some specific moment inequalities/ submodels.

Let $\Omega(\Theta, \Lambda) = \{\theta \in \Theta : EM(X, \theta) \in \Lambda\}$.

Theorem 3.4.2. *Under Assumptions 3.2.1 and 3.4.2, with fixed $V > 0$, in probability,*

$$plim_{n\to\infty} p(C_s|X^n) = \frac{p(C_s)\int_{\Omega(\Theta,\Lambda)} p_{\theta,\lambda}(\theta_s, EM(X,\theta)|C_S)d\theta}{\sum_{C_s} p(C_s)\int_{\Omega(\Theta,\Lambda)} p_{\theta,\lambda}(\theta_s, EM(X,\theta)|C_S)d\theta} \tag{3.10}$$

We can see from this theorem, that asymptotically the posterior depends on $p(C_s|X^n)$, the combination's prior, and on $p_{\theta,\lambda}$, which is the prior distribution of (θ, λ) on the identified region. Therefore, the posterior is sensitive to the prior specification.

On the other hand, note that this is not a consistency result: Consider two compatible combinations C_1 and C_2 with the same parameter subspace but C_1 is nested with C_2 and contains more moment inequalities than C_2. Since C_1 has smaller identified region $\Omega(\Theta, \Lambda)$, by (3.10), it may have smaller posterior, even asymptotically. Therefore, compatible combinations with more inequalities do not necessarily have larger posteriors. This result is quite different from those in regular moment selection procedures with point identification (for example, in Andrews (1999)), which is reasonable, however, in moment inequalities problems, because of three reasons:

(1) First, as we have seen, the posterior is sensitive to the choice of priors. The penalty term against selecting fewer moment inequalities in the posterior is hidden in

$\int_{\Omega(\Theta,\Lambda)} p_{\theta,\lambda}(\theta_s, EM(X,\theta)|C_S)d\theta$, which does not involve the sample size.

(2) Second, unlike the moment selection problem with over-identification by moment equalities (Andrews 1999, Andrews and Lu 2001), in moment inequalities models, compatible combinations may not be correct, meaning that if the true parameter of interest is assumed to be fixed, some combinations may still be compatible even though they do not contain the true parameter. Therefore, the true inequalities are not necessarily the maximal set of compatible inequalities.

(3) Finally, by allowing the parameter space to change, we allow for the model uncertainty. In this case, it is reasonable for the result to heavily reply on the prior beliefs of the useful parameter components, and of the corresponding moment inequalities.

In some special cases, however, it is possible that the posterior favors the maximal number of inequalities. Consider the following example:

Example 3.4.2. Suppose the selected moment conditions satisfy $EM_s(X,\theta) = \lambda_s \in [0,M]^m$, and the unselected moment conditions satisfy $EM_s^c(X,\theta) = \lambda_s^c \in [-M,M]^{p-m}$, for some large constant $M > 0$. Hence $\Lambda = [0,M]^m \times [-M,M]^{p-m}$. Suppose $q(\lambda)$, the prior of λ, is uniformly distributed on $[-M,M]^p$, hence $q(\lambda) = (2M)^{-p} I(\lambda \in [-M,M]^p)$. Then

$$p_\lambda(\lambda|C_s) = \frac{q(\lambda)I(\lambda\in\Lambda)}{q(\lambda\in\Lambda)} = \frac{2^m I(\lambda\in\Lambda)}{(2M)^p}$$

In addition, suppose $p_{\theta,\lambda}(\theta,\lambda|C_s) = p_\theta(\theta|C_s)p_\lambda(\lambda|C_s)$. We have

$$\int_{\Omega(\Theta_s,\Lambda)} p_{\theta,\lambda}(\theta, EM(X,\theta)|C_s)d\theta = P(\theta\in\Omega(\Theta_s,\Lambda)|C_s)\frac{2^m}{(2M)^p}$$

Assume that $P(\theta \in \Omega(\Theta_s,\Lambda)|C_s) > 0$ if C_s is compatible. Hence 2^m is the reward of more moment inequalities. However, such a reward term does not depend on the sample size n. Q.E.D.

The moment equality condition case in the literature is significantly different than the problem considered here. For the sake of comparison, I briefly illustrate it here. Consider p-dimensional candidate moment equalities $EM(X,\theta) = (Em_1, ..., Em_p) = 0$. Suppose we select m moment conditions $EM_s = 0$, and partition the conditions into selected and unselected pair $M = (M_s, M_s^c)$. As before, we use the limited information likelihood to construct the posterior:

$$L(X^n|\lambda,\theta,C_s) = \det(2\pi V/n)^{-1/2}\exp\left(-\frac{n}{2}(\bar{M}_s(\theta)^T, \bar{M}_s^c(\theta)^T - \lambda^T)V^{-1}\begin{pmatrix}\bar{M}_s(\theta)\\ \bar{M}_s^c(\theta)-\lambda\end{pmatrix}\right)$$

Straightforward calculation yields that

(3.11)

$$p(C_s|X^n) \approx Const \times n^{\dim(M_s)/2}\int_\Theta p_{\theta,\lambda}(\theta, EM_s^c(\theta) + \Sigma_2^T\Sigma_3 EM_s(\theta))e^{-\frac{n}{2}EM_s(\theta)^T\Sigma_1 EM_s(\theta)}d\theta$$

where $V^{-1} = \begin{pmatrix}\Sigma_1 & \Sigma_3\\ \Sigma_3^T & \Sigma_2\end{pmatrix}$. When M_s is incompatible, meaning that $\{\theta \in \Theta : EM_s(X,\theta) = 0\}$ is empty, $e^{-\frac{n}{2}EM_s(\theta)^T\Sigma_1 EM_s(\theta)} < e^{-an}$ for some $a > 0$. But when $EM_s(X,\theta)$ over-identifies some element θ in Θ, the posterior then replies on $n^{\dim(M_s)/2}$, which is a penalty term that rewards the use of more moment conditions. Note that this penalty term depends on the sample size, hence is not sensitive to the prior specification. In addition, by applying the Laplace expansion to the integrand of the right hand side, the posterior criterion (3.11) can be shown to be equivalent to Andrews(1999)'s MSC.