CONTENTS IN BRIEF

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

APPROXIMATE BAYESIAN INFERENCE IN MODELS DEFINED THROUGH ESTIMATING EQUATIONS

This chapter reviews recent approaches to Bayesian inference when the only information linking parameters and data is in the form of estimating equations. Bayesian inference in these class of models is difficult to implement since the likelihood function is left unspecified. The approaches reviewed in the chapter are those that rely on posterior constructed using frequentist objective functions, such as the Generalized Method of Moments and the Generalized Empirical Likelihood, as approximate likelihoods functions. We explore the Bayesian and frequentist properties of inference based on these posteriors.

1.1 Introduction

Over the last three decades, econometric models characterized in terms of restrictions on the moments of underlying random vectors have gained a prominent role in economics and in the social sciences in general. The class of models we have in mind here are those that can be summarized by the following population equation

$$E[g(X, \theta)] = 0, \tag{1.1}$$

where the random vector X has dimension D and support $\mathcal{X} \subseteq \mathbb{R}^D$, θ is a parameter vector, an element of $\Theta \subset \mathbb{R}^K$, and the moment function $g(\cdot, \cdot) : \mathcal{X} \times \Theta \to \mathbb{R}^M$ is a known vector-valued function. In the next section we will provide some examples of models that satisfy (1.1), but it should already be clear that many conventional estimation methods such as maximum likelihood, ordinary least squares, panel data models, and instrumental variables can be casted into this framework. Keep in mind though that (1.1) is flexible enough to also accommodate specification of more complex statistical models.

While the Generalized Method of Moments (GMM) as formalized by Hansen (1982) and the Generalized Empirical Likelihood (GEL) of Newey and Smith (2004) are the de-facto standard for conducting inference about the parameter θ of (1.1), there is not an accepted way of doing so in a manner consistent with the Bayesian framework. This deficiency is problematic because, as Sims (2002) rightly puts it, models defined in terms of estimating equations are used in real decision making problems and, thus, classical confidence bands for parameters are going to be interpreted as posterior credible regions.

The key difficulty in carrying out proper Bayesian inference in this setting has to do with the fact that usually the conditional distribution of the data given the parameter—the likelihood function—is not available and thus the Bayes' theorem cannot be readily applied. There have been many attempts at overcoming such problem. For specific instances of the class of models that we are concerned with here, it is feasible to carry out a full fledged semiparametric Bayesian analysis. For example, Hirano (2002) and Conley et al. (2008) propose semiparametric Bayesian methods for inference in dynamic panel data models with fixed effects and in instrumental variables problems, respectively. It is however difficult to directly broaden the scope of these methods to other models defined through the moment condition equations in (1.1). Always for specific models, in particular linear ones, Zellner (1994) introduced the Bayesian Method of Moments (BMOM); see, for applications and further discussions, Zellner (1991, 1997), Zellner and Tobias (2001).

A more general approach to carry out Bayesian inference in the estimating equation framework has been proposed by Chamberlain and Imbens (2003). They extend the Bayesian bootstrap of Rubin (1981) to obtain a posterior distribution for θ that is justified by assuming an improper Dirichlet prior on the space of probability distributions. There are however two issues with this approach. The main problem is that it only applies directly to cases in which the dimension of θ is equal to the dimension of the moment function, the exactly identified case. When instead the model is over-identified, that is there are more equations than parameters, the method is only applicable indirectly. Furthermore, since the computation of the posterior relies on solving potentially non-linear equations for each iteration of the bootstrap, the method may become infeasible, especially when the dimension of θ is large.

A recent trend has seen the attempt of replacing the likelihood function in the Bayes' theorem with approximate likelihoods that are transformations of frequentist objective functions. Kim (2002) proposes using as a likelihood a transformation of the efficient GMM objective function. Yin (2009) shows that Kim's likelihood can potentially provide a good starting point for approximate Bayesian inference for certain instances of the class of models in (1.1). Yu and Stander (2007) develop a Bayesian framework for Tobit quantile regressions that is organized around a likelihood function based on the objective function of Powell (1986).

Lazar (2003) investigates the properties of Bayesian posteriors based on the empirical likelihood function. Empirical likelihood (EL) (Owen, 1988, 1990, 2001) acts as a nonparametric analog of the usual likelihood function. Lazar (2003) conducts simulations to asses the coverage validity of EL based posteriors and the frequentist coverage of credible intervals. Although she only considers very simple models, the simulation results allow her to conclude that it is reasonable to use empirical likelihood within the Bayesian paradigm. Schennach (2005) introduces another approximate likelihood which is similar in spirit to that based on empirical likelihood. This likelihood, aptly named Bayesian exponentially tilted empirical likelihood (BETEL), is obtained from an exponential tilted multinomial distribution (DiCiccio and Romano, 1990). Ragusa (2011) considers various other likelihoods that can be thought as extension of the BETEL.

The main objective of the chapter is to explore Bayesian and frequentist properties of inference based on posterior distributions resulting from using these likelihood functions. To be able to do so, we need to be more precise about the formulation of the statistical model under consideration.

1.2 Examples

As said, most models considered in econometrics can be expressed in terms of estimating equations. We provide same examples for which Bayesian inference is not yet developed.

Generalized linear model

The generalized linear model provides a unifying framework for the analysis of discrete and continuous outcomes. For the $i = 1, \dots, n$ subject the outcome y_i is observed together with the corresponding covariate vector X_i. The density function of y_i given X_i takes the form

$$\varphi(Y_i|X_i) = \exp\left\{\frac{Y_i\zeta_i - b(\zeta_i)}{a_i(\phi)} + c(Y_i, \phi)\right\},$$

where ζ_i and ϕ are location and scale parameters, respectively. The linear predictor $\theta^\top X_i$ is linked to the location parameter by a monotone differential function $h(\cdot)$, such that $\zeta_i = h(X_i^\top \theta)$. The conditional mean of Y_i is then given by $\mu_i = E[Y_i|X_i] = b'(\zeta_i)$, where $\mu_i = h^{-1}(X_i^\top \theta)$, and the conditional variance is $\upsilon_i = \mathrm{Var}(Y_i|X_i) = b''(\zeta_i)a_i(\phi)$. The population score equation satisfies (1.1) with

$$g(Y_i, X_i, \theta) = \frac{\partial \mu_i}{\partial \theta} \upsilon_i^{-1}(Y_i - \mu_i).$$

Linear instrumental variables

Consider the linear model

$$Y = X^\top \theta + u,$$

where the conditional mean of u given X is not constant. The parameter θ can be estimated consistently if there is a vector of instruments Z that is orthogonal to u and correlated with X. In this case, the model boils down to the moment function

$$g(Y, X, Z, \theta) = Z^\top (Y - X^\top \theta).$$

The linear instrumental variable model is overidentified if the dimension of Z is larger than the dimension of θ.

Quantile instrumental variables

Let D be a binary treatment indicator. There is a potential outcome, Y_D, associated with the potential treatment status $d \in \{0, 1\}$, $Y_{d=1}$ and $Y_{d=0}$. The outcome is latent because for each individual unit we only observe either $Y_{d=1}$ or $Y_{d=0}$. The quantile treatment response (QTR) function is

$$q(d, x, \tau) = d\,\theta_1(\tau) + x^\top \theta_2(\tau)$$

which represents the τth quantile of potential outcomes under various treatments d, conditional on observed characteristics $X = x$. The quantity of interest is the quantile treatment effect (QTE)

$$q(1, x, \tau) - q(0, x, \tau) = \theta_1(\tau), \quad \tau \in (0, 1)$$

which is the differential impact of treatment on different quantiles of Y. If the treatment is endogenous, Chernozhukov and Hansen (2005) show that the QTR can be identified by a conditional moment restriction

$$E\left[\mathbb{1}_{[Y_d \leqslant D\theta_1(\tau) + X^\top \theta_2(\tau)]} - \tau | X, Z\right] = 0,$$

where Z is an instrument that affects D but is independent of potential outcomes. Thus, the model can be casted in the form of (1.1) with the following moment function

$$g(Y_D, X, Z, \theta_1(\tau), \theta_2(\tau)) = \begin{pmatrix} X \\ Z \end{pmatrix} \mathbb{1}_{[Y_d \leqslant D\theta_1(\tau) + X^\top \theta_2(\tau)]}.$$

Side information

Estimating equations are key when marginal population information about some variables involved in a regression is available. As shown by Imbens and Lancaster (1994), this information can be easily incorporated by imposing moment restrictions. Consider the following linear model:

$$Y = X^\top \theta + u$$

where $E[u|X] = 0$. Suppose now that information is available about certain moments of the distribution of X. For instance, the researcher knows that

$$h_1^* = E[X|X > c], \quad h_2^* = E[X|X \leqslant c].$$

Knowledge of h_1^* and h_2^* restricts the marginal distribution of X in a non trivial way. The equations above lead to the following set of restrictions:

$$E[\mathbb{1}_{X>c}(h_1^* - X^\top \theta)] = 0, \quad E[\mathbb{1}_{X^\top \leqslant c}(h_2^* - X^\top \theta)] = 0.$$

The moment function corresponding to this model is

$$g(Y, X, \theta) = \begin{bmatrix} X_i(Y_i - X_i^\top \beta) \\ \mathbb{1}_{X_i > c}(h_1^* - X_i^\top \beta) \\ \mathbb{1}_{X_i \leqslant c}(h_2^* - X_i^\top \beta) \end{bmatrix}.$$

1.3 Frequentist estimation

For a frequentist point of view, the parameter θ is unknown but "fixed". To reflect this view, the model in (1.1) is often complemented by two additional assumptions. First, there exists a "true" parameter $\theta^* \in \Theta$ such that

$$E[g(X, \theta^*)] = 0. \tag{1.2}$$

Second, the model is identified in the sense that the moment conditions only hold at θ^*, or more formally,

$$\|E[g(X, \theta)]\| \neq 0, \quad \text{for all } \theta^* \neq \theta.$$

Given an i.i.d. random sample $\mathbb{X} = \{x_1, x_2, \ldots, x_n\}$ let $g_i(\theta) = g(x_i, \beta)$, $g_n(\theta) = \sum_{i=1}^n g_i(\theta)/n$ and $\widehat{\Omega}(\theta) = \sum_{i=1}^n g_i(\theta)g_i(\theta)'/n$. The objective is to carry out inference about the finite dimensional parameter β_0.

The most common frequentist procedure to estimate θ^* is the Generalized Method of Moments (GMM) of Hansen (1982). The procedure consists in minimizing a weighted quadratic expression in the empirical estimating equations, where the weighting matrix is a consistent estimator of the covariance of $\sum_{i=1}^n g(x_i, \theta)/n$. Formally, the GMM estimator is given by

$$\hat{\theta}^{GMM} = \arg\min_{\theta \in \Theta} J_n(\theta), \quad J_n(\theta) := n \cdot g_n(\theta)' \left[\frac{1}{n} \sum_{i=1}^n g_i(\bar{\theta}) g_i(\bar{\theta})^\top \right]^{-1} g_n(\theta),$$

where $\bar{\theta}$ is a preliminary consistent estimate of θ^*. In a wide array of settings (Hansen, 1982; Gallant and White, 1988; Newey and McFadden, 1994) the GMM estimator is consistent and asymptotically normal.

A number of alternative estimators to the GMM have been proposed in recent years. The motivation for these estimators is often the bias of the two-step GMM estimator and the lack of accuracy of the confidence intervals based on the normal approximation to the distribution of $\hat{\theta}^{GMM}$.

Qin and Lawless (1994) and Imbens (1997) have proposed estimators that embed estimating equations into the empirical likelihood framework (Owen, 1988, 1990). The intuition behind the empirical likelihood is important for understanding the logic of some of the approximate posteriors we will be discussing in the next sections. Suppose that the moment function does not depend on any unknown parameter and takes the form $g(x, \theta) = x$. The condition $E[g(X, \theta)] = 0$ implies that the expected value of X is 0. The empirical distribution function does not satisfy the moment condition, since with probability 1, $\sum_{i=1}^n x_i/n \neq 0$. The idea behind empirical likelihood is to modify the empirical distribution in such a way that it satisfies the moment conditions. As there exist many modifications of the empirical distribution satisfying the moment condition, the idea is to pick the one that is the closest to the empirical one. In other words, among all systems of weights $\{p_i\}_{i=1}^n$ such that $\sum_{i=1}^n p_i x_i = 0$, we pick the one that maximizes $\prod_{i=1}^n p_i$. We can think of this optimal weighting as an estimator of the distribution function of X.

Consider now the general case where the moment condition depends on the unknown parameter θ. In this case, for a given θ, the empirical likelihood estimator of the probability distribution is given by

$$\max_p \prod_{i=1}^n p_i, \text{ subject to } p_i > 0, \quad \sum_{i=1}^n p_i = 1, \quad \sum_{i=1}^n p_i g(x_i, \theta) = 0.$$

The empirical distribution for this given value of θ has a simple solution, namely

$$p_i^{EL} = \frac{1}{n}\frac{1}{1+\lambda_\theta^\top g(x_i,\theta)}, \tag{1.3}$$

where λ_θ solves the following set of nonlinear equations

$$\sum_{i=1}^{n} \frac{1}{1+\lambda_\theta^\top g(x_i,\theta)} g(x_i,\theta) = 0. \tag{1.4}$$

Substituting (1.3) into the likelihood, we obtain the "empirical likelihood" at θ:

$$f_{EL}(\mathbb{X}|\theta) = \prod_{i=1}^{n} \frac{1}{n[1+\lambda_\theta^\top g(x_i,\theta)]}. \tag{1.5}$$

The empirical likelihood estimator is obtained by maximizing $f_{EL}(\mathbb{X}|\theta)$ over $\theta \in \Theta$.

The empirical likelihood idea can be generalized by noting that maximizing $\prod_{i=1}^n p_i$ is equivalent to minimizing $-\sum_{i=1}^n \log p_i$. From this point of view, the empirical likelihood operationalizes the concept of closeness between empirical distribution and the system of weights compatible with the moment conditions. As such, other measures of distance can be employed to estimate the distribution of the underlying random variables. Kitamura and Stutzer (1997) propose to use the Kullback-Leibler distance measure. In this case, the estimator of the probability distribution is

$$\min_p \sum_{i=1}^{n} p_i \log p_i, \text{ subject to } p_i \geq 0, \quad \sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i g(x_i,\theta) = 0.$$

The solution is then

$$p_i^{ET} = \frac{e^{\lambda_\theta^\top g(x_i,\theta)}}{\sum_{i=1}^n e^{\lambda_\theta^\top g(x_i,\theta)}}, \tag{1.6}$$

and, as for the empirical likelihood case, λ_θ is chosen in such a way to guarantee that

$$\sum_{i=1}^{n} e^{\lambda_\theta^\top g(x_i,\theta)} g(x_i,\theta) = 0. \tag{1.7}$$

Kitamura and Stutzer's exponential tilting estimator of θ^* is then obtained by substituting this expression into the objective function and minimizing over $\theta \in \Theta$.

A generalization of these estimators is obtained by considering, as measure of distance between the empirical distribution and the system of weights, the

Cressie-Read divergence, which is given by

$$\frac{1}{n[\alpha(1+\alpha)]}\sum_{i=1}^{n}\left[\left(\frac{1}{np_i}\right)^{\alpha}-1\right],$$

where α is a parameter. In this case, the estimator of the probability distribution is

$$p_i^{CR}=\frac{(1+\alpha\lambda_\theta^\top g(x_i,\theta))^{1/\alpha}}{\sum_{i=1}^{n}(1+\alpha\lambda_\theta^\top g(x_i,\theta))^{1/\alpha}}$$

and, once again, the estimator of θ^* is obtained by substituting this expression in the objective function and maximizing over $\theta\in\Theta$. The empirical likelihood and the exponential tilting estimate of the probability distribution are special cases which are obtained for $\alpha\to 0$ and $\alpha\to -1$, respectively. As shown by Newey and Smith (2004), when $\alpha=-2$, the Cressie-Read estimator coincides with the Continuous Updating (CUE) estimator of Hansen et al. (1996), which is obtained by minimizing over $\theta\in\Theta$ the following objective function

$$Q_n(\theta)=n\cdot g_n(\theta)^\top S_n^{-1}(\theta)g_n(\theta), \tag{1.8}$$

where

$$S_n(\theta)=\frac{1}{n}\sum_{i=1}^{n}g_i(\theta)g_i(\theta)^\top-\frac{1}{n^2}\sum_{i=1}^{n}g(x_i,\theta)\sum_{i=1}^{n}g(x_i,\theta)^\top.$$

The main difference between the traditional GMM and the CUE is that the weighting matrix is not held fixed, but it is seen as a function of the parameter vector.

The Cressie-Read estimators are all asymptotically equivalent to the two-step GMM estimator. However, as shown by Newey and Smith (2004) and Ragusa (2011), they possess better second order properties, such as smaller bias with respect to the GMM. Furthermore, the empirical likelihood estimator is third order efficient.

1.4 Bayesian estimation

While the frequentist estimation of models specified by estimating equations is well developed, Bayesian inference is more problematic. Bayesian inference is run by the posterior distribution, which by Bayes' theorem is given by

$$\pi(\theta|\mathbb{X})=\frac{f(\mathbb{X}|\theta)\pi(\theta)}{\int f(\mathbb{X}|\theta)\pi(\theta)d\theta},$$

where $f(\mathbb{X}|\theta)$ is the likelihood function for the observed datum $\mathbb{X}=\{x_{1,\ldots,}x_n\}$ and $\pi(\theta)$ is the prior density on the parameter θ. In parametric models, the likelihood is the conditional density of the data given the parameter θ

and it arises from distributional assumptions. When the model is specified through estimating equations, it is semi-parametric and thus the likelihood is not readily available without further assumptions.

Before digging further into the issue of conducting inference, we should carefully highlight the difference of interpretation of the estimating equations in the Bayesian framework. For a frequentist, the parameter θ is unknown but fixed and it takes a true value, θ^*. Thus, the moment conditions (1.1) hold only at θ^*. From a Bayesian perspective, the parameter θ is random and the expectation should be considered with respect to the conditional distribution of X given θ, P_θ, that is

$$E[g(X,\theta)|\theta] \equiv \int g(X,\theta)dP_\theta = 0. \tag{1.9}$$

Writing the moment condition this way clarifies that in this class of models the likelihood function is left unspecified and Bayesian inference is problematic. The recent literature has proposed different approaches to obviate the lack of a likelihood function. In what follows, we review these approaches.

1.4.1 Bayesian bootstrap

Chamberlain and Imbens (2003) extend the Bayesian bootstrap (BB) of Rubin (1981) to semiparametric models by embedding the model in a multinomial framework.[1] To describe the setup, let $a = (a_1, \ldots, a_J)$ be the vector of all possible distinct values of X, and let $\xi = (\xi_1, \ldots, \xi_J)$ be the associated vector of probabilities, $P(X = a_j) = \xi_j$, $\sum_j \xi_j = 1$. Let n_j denote the number of observations equal to a_j. The likelihood of the data, $\mathbb{X} = \{x_1, \ldots, x_n\}$, is proportional to

$$\prod_{j=1}^{J} \xi_j^{n_j}.$$

The prior on ξ is assumed to belong to the Dirichlet family, that is, the prior has density proportional to $\prod_{j=1} \xi_j^{l_j - 1}$, $l_j > 0$. The posterior density of ξ is proportional to

$$\pi(\xi|\mathbb{X}) \propto \prod_{j=1}^{J} \xi_j^{n_j + l_j - 1}, \quad l_j > 0,$$

which is as well a Dirichlet with parameters $n_j + l_j$, $j = 1, \ldots, J$. So far we have not considered the estimating equation in (1.9). The idea of Chamberlain

[1]Bayesian bootstrap in nonparametric settings has also been considered by Ferguson (1973, 1974) and Gasparini (1995). Hahn (1997) studies the frequentist properties of Bayesian Bootstrap for the quantile regression case and Lancaster (1994) applies Bayesian bootstrap to the analysis of choice-based sampling. Poirier (2011) utilizes the Bayesian bootstrap to provide a Bayesian interpretation of the heteroskedastic consistent covariance estimator of White (1980).

and Imbens (2003) is to turn the posterior for ξ into a posterior distribution for the parameter θ by drawing θ's which are consistent with $\pi(\xi|\mathbb{X})$. To get the intuition, suppose ξ^s is a draw from $\pi(\xi|\mathbb{X})$; associated with this draw of ξ, there is a value of the parameter θ, say θ^s, such that

$$\sum_{j=1}^{J} g(a_j, \theta^s)\xi_j^s = 0.$$

All the θ^s consistent with all the possible draws from $\pi(\xi|\mathbb{X})$ are draws from the posterior of θ. Chamberlain and Imbens (2003) consider drawing from this posterior when $l_j \to 0$ for all j, in which case support points not observed in the sample receive posterior probability 0. In this case, to simulate from the posterior distribution of θ we can repeatedly draw exponential iid random variables, V_i^s, $i = 1, \dots, n$, and solve

$$\sum_{i=1}^{n} V_i^s g(x_i, \theta^s) = 0.$$

Repeating this for $s = 1, \dots, S$, gives us S independent draws from the posterior distribution of θ.

As pointed out by Lancaster (2004), there are few issues with this approach. First, it is not clear how to incorporate prior knowledge about θ, as the prior on θ is implicitly elicited. Second, drawings are obtained by solving a potentially high-dimensional nonlinear set of equations which can result in computational problems. Third, when the model is overidentified, that is when the dimension of the parameter space is smaller than the dimension of the moment function, the method can be applied only after projecting out the redundant equations. The projection takes place outside the model and hence the resulting inference might be arbitrary.

1.4.2 GMM-based likelihoods

Kim (2002) proposes conducting Bayesian inference by using a transformation of the GMM objective function. The procedure is, at least conceptually, extremely simple as it involves using the Bayes' formula replacing the likelihood with a transformation of the objective function $J_n(\theta)$. More specifically, for

$$f_{GMM}(\mathbb{X}|\theta) = e^{-\frac{1}{2}J_n(\theta)},$$

inference is based on

$$\pi_{GMM}(\theta|\mathbb{X}) \propto f_{GMM}(\mathbb{X}|\theta)\pi(\theta).$$

Kim (2002) motivates this choice of a likelihood by embedding the moment condition in an information theoretical framework. However, the justifications for using $\pi_{GMM}(\theta|\mathbb{X})$ are only asymptotic and no "small sample" justification is provided.

An approach similar in spirit to the one just presented is the one by Yin (2009). He considers using as posterior

$$\pi_{CU}(\theta|\mathbb{X}) \propto f_{CU}(\mathbb{X}|\theta)\pi(\theta),$$

where the likelihood function, $f_{CU}(\mathbb{X}|\theta)$, is the exponential transformation of the CUE objective function in (1.8):

$$f_{CU}(\mathbb{X}|\theta) = e^{-\frac{1}{2}Q_n(\theta)}.$$

The same approach was suggested by Chernozhukov and Hong (2003) who considered quasi posterior inference from a frequentist point of view, mainly as a computational device to construct consistent and asymptotically normal estimators of parameters defined through moment conditions.

These likelihoods can be motivated asymptotically by noticing that under fairly general conditions, as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i,\theta) \xrightarrow{d} N\left(0, S(\theta)/n\right), \quad S(\theta) = E\left[g(X_i,\theta)g(X_i,\theta)^\top\right].$$

The CU likelihood can be thought as being based on the limit distribution of the empirical estimating equation when the normalizing constant $(2\pi)^{-k/2}|S^{-1}(\theta)|^{1/2}$ is not taken into account. This suggests a modified version of the CU likelihood which also uses the empirical normalizing constant,

$$f_{NCU}(\mathbb{X}|\theta) = (2\pi)^{-m/2}|S_n^{-1}(\theta)|^{1/2}e^{-\frac{1}{2}Q_n(\theta)},$$

where $S(\theta)$ is replaced by its sample counterpart $S_n(\theta)$.

While we will carefully discuss the issue of simulating from either $\pi_{GMM}(\theta|\mathbb{X})$ or $\pi_{CU}(\theta|\mathbb{X})$ in a later section, it is instructive to compare these two posterior distributions with the "true" posterior for one simple example. To do so, we consider data generated from the following model $Y|X = x \sim N(x\theta, 1)$, $X \sim N(0,1)$, and $\theta = 0$. The parameter θ follows a normal distribution centered at 0 and variance $\tau^2 = \sqrt{.2}$. In this case, the exact posterior distribution of θ is normal with mean $\left(\sum_{i=1}^n x_i^2 + \tau^{-2}\right)^{-1}\sum_{i=1}^n x_i y_i$ and variance $\left(\sum_{i=1}^n x_i^2 + \tau^{-2}\right)^{-1}$. The posterior distributions $\pi_{GMM}(\theta|\mathbb{X})$, $\pi_{CU}(\theta|\mathbb{X})$, and $\pi_{NCU}(\theta|\mathbb{X})$ are calculated using the following estimating equations:

$$g(Y, X, \theta) = X^\top(Y - X^\top\theta).$$

Figure 1.1 plots the posterior distributions for different draws from the model and for different sample sizes. As it can be seen, when the sample size is large all the three posterior distributions based on the estimating equations are

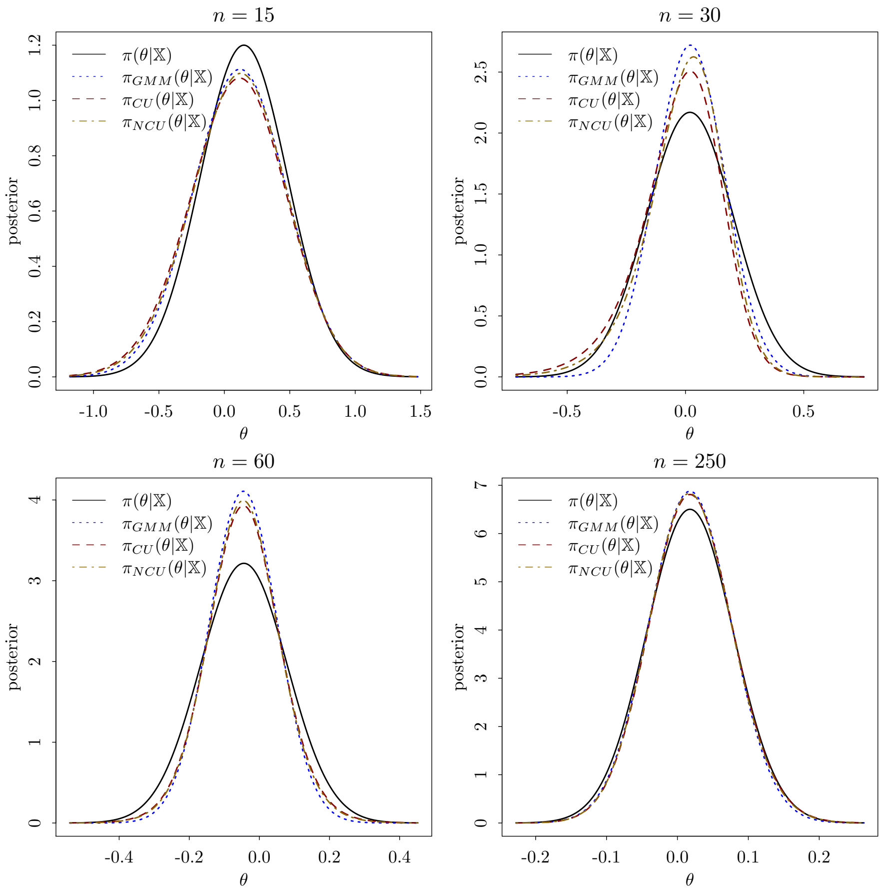

Figure 1.1: GMM, CU, and normalized CU based posteriors. The posterior distributions are based on data generated from the following model: $Y|X = x \sim N(x\theta, 1)$, $X \sim N(0,1)$, $\theta = 0$. The prior is $\theta \sim N(0, \tau^2)$, $\tau = \sqrt{.2}$. The posterior distributions $\pi_{GMM}(\theta|\mathbb{X})$, $\pi_{CU}(\theta|\mathbb{X})$, and $\pi_{NCU}(\theta|\mathbb{X})$ are calculated from the estimating equation $g(Y, X, \theta) = X^\top(Y - X^\top\theta)$. The normalizing constants are calculated by numerical integration.

close to the exact posterior, especially near the tails. At smaller sample sizes, the three posteriors put more mass toward the center, with a marked tendency of the posterior based on the GMM to have the thinnest tails. Whether this issue is specific to the linear model or it is a standard feature of the posteriors will be briefly discussed in a later section.

1.4.3 Empirical likelihood-type posteriors

As discussed in Section 1.3, the empirical likelihood behaves for some purposes as a parametric likelihood. For instance, the maximum empirical likelihood is consistent, asymptotically normal, and like the parametric MLE is third order efficient in the sense of Pfanzagl and Wefelmeyer (1978) (see, Newey and Smith, 2004; Ragusa, 2011). The empirical likelihood is also Bartlett correctable (DiCiccio et al., 1991; Chen and Cui, 2007). It is natural then to try to go a step further and see whether the empirical likelihood can be used in the Bayesian framework.

Lazar (2003) follows up on this simple idea and proposes using the empirical likelihood instead of the conditional distribution of the data in the Bayes' theorem. This amounts to consider the following posterior distribution

$$\pi_{EL}(\theta|\mathbb{X}) \propto f_{EL}(\mathbb{X}|\theta)\pi(\theta),$$

where the expression for $f_{EL}(\mathbb{X}|\theta)$ is given in equation (1.5). Instead of using the empirical likelihood, Schennach (2005) suggests a posterior distribution based on the exponentially tilted empirical distribution function. More formally, Schennach's posterior is given by

$$\pi_{ET}(\theta|\mathbb{X}) \propto f_{ET}(\mathbb{X}|\theta)\pi(\theta), \quad f_{ET}(\mathbb{X}|\theta) = \prod_{i=1}^{n} p_i^{ET}, \tag{1.10}$$

where p_i^{ET}, $i = 1, \ldots, n$, are defined in (1.6). Following the same idea, other posteriors could potentially be constructed by using other Cressie-Read empirical distributions.

Calculating the empirical likelihood and the exponentially tilted posteriors for a given value of θ amounts to calculating the weights and solving (1.4) and (1.7) for λ, respectively. At certain values of the parameter, the nonlinear equations might not have a solution. In this case, the value of the likelihood is 0, since there exists no system of weights that sets the empirical moment condition to zero.

Figure 1.2 shows $\pi_{EL}(\theta|\mathbb{X})$ and $\pi_{ET}(\theta|\mathbb{X})$ for the same model and the same data for which we calculated the posteriors presented in Section 1.4.2. A comparison with the distributions depicted in Figure 1.1 reveals that, at least for the regression model considered here, empirical likelihood and exponential tilting posteriors behave similarly to those based on GMM, CU, and the normalized CU.

1.5 Simulating from the posteriors

While the posterior distributions described in the previous section can be calculated quite easily point-wise over the space Θ, to calculate their moments one needs to resort to Monte Carlo simulation techniques. In principle any

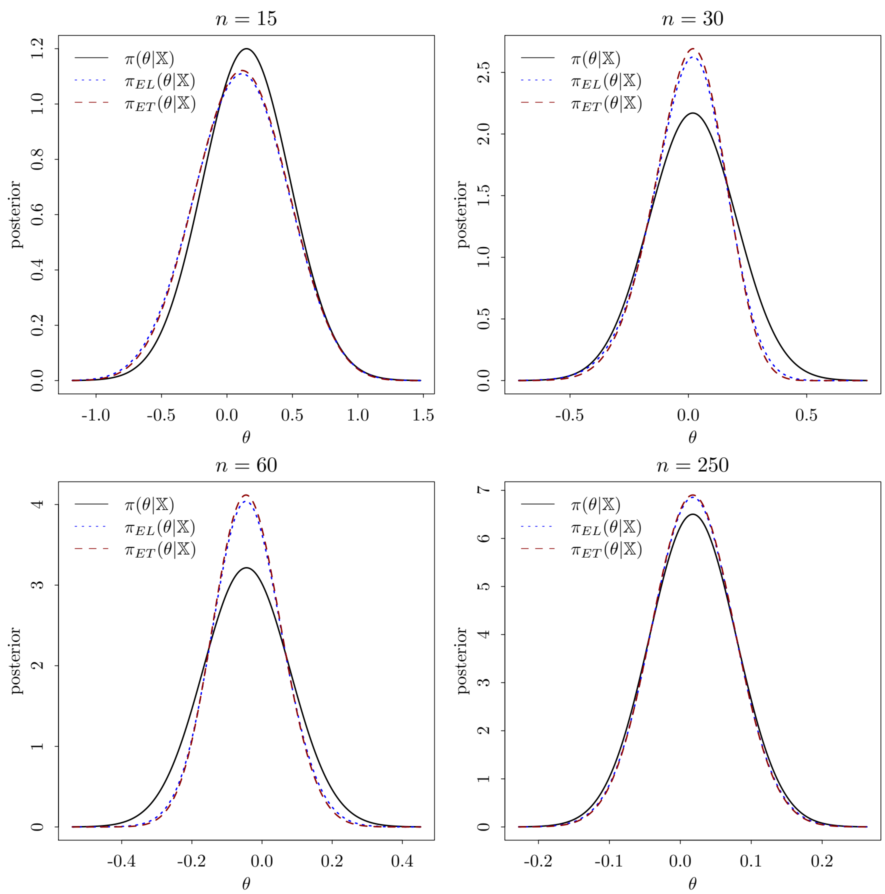

Figure 1.2: Empirical likelihood based posteriors. The posterior distributions are based on data generated from the following model: $Y|X = x \sim N(x\theta, 1)$, $X \sim N(0, 1)$, $\theta = 0$. The prior is $\theta \sim N(0, \tau^2)$, $\tau = \sqrt{.2}$. The posterior distributions $\pi_{EL}(\theta|\mathbb{X})$, $\pi_{ET}(\theta|\mathbb{X})$ are calculated from the estimating equation $g(Y, X, \theta) = X^\top(Y - X^\top\theta)$.

algorithm that allows simulating from a distribution that can be computed up to a proportionality constant can be employed. In this section, we restrict the discussion to the Metropolis-Hastings (MH) algorithm. Keep in mind that a detailed discussion of the MH algorithm and its properties is beyond the scope of the present chapter; for specific issues and implementation details see, among others, Robert and Casella (2004).

The MH algorithm is a method to produce Markov chains with a given ergodic distribution. The output is a dependent sample $(\theta^{(1)}, \dots, \theta^{(S)})$ whose empirical distribution approaches, as $S \to \infty$, $\pi_j(\theta|\mathbb{X})$, and the index j ranges over the different posteriors $j = \{GMM, CU, NCU, EL, ET\}$.[2] The ergodicity of the chain implies that

$$\frac{1}{S}\sum_{s=1}^{S} h(\theta^s) \xrightarrow[s\to\infty]{p} \int h(\theta)\pi_j(\theta|\mathbb{X})d\theta.$$

Given the posterior $\pi_j(\theta|\mathbb{X})$ and a pre-specified density $q(\theta|\theta')$, a Markov chain with limiting distribution $\pi_j(\theta|\mathbb{X})$ can be constructed iteratively as follows:

- Choose a starting value $\theta^{(0)}$
- Generate $\tilde{\theta} \sim q(\theta|\theta^{(s)})$ and calculate
$$\alpha = \inf\left\{\frac{\pi_j(\tilde{\theta}|\mathbb{X})}{\pi_j(\theta^{(s)}|\mathbb{X})}\frac{q(\theta^{(s)}|\tilde{\theta})}{q(\tilde{\theta}|\theta^{(s)})}, 1\right\}$$
- Set
$$\theta^{(s+1)} = \begin{cases} \tilde{\theta} & \text{with probability } \alpha \\ \theta^{(s)} & \text{with probability } 1-\alpha \end{cases}.$$

A common choice for the proposal density is

$$q(\theta|\theta') = \xi(|\theta - \theta'|),$$

where $\xi(\cdot)$ is a density symmetric about the origin, such as the Gaussian, the student-t, or the Cauchy density. In this case, the Markov chain is a random walk and the MH algorithm is called the random walk Metropolis-Hastings (RWMH). The advantage of RWMH over the standard HM algorithm is that it allows a local exploration of the neighborhood of the current value of the Markov chain, since the candidate distribution depends on the current state of the chain.

[2]Since most of the discussion in this section is valid for all the posterior distributions presented in Section 1.4 with the exception of the Bayesian bootstrap, we will refer to them generically with $\pi_j(\theta|\mathbb{X})$.

1.6 Asymptotic theory

Under well-known regularity conditions, standard Bayesian inference has a well defined asymptotic behavior. Ibragimov and Has'minskii (1981) have shown that in regular and correctly specified models the posterior distribution approaches a normal density. From these results it follows that posterior mean, posterior median and other symmetric moments of the posterior are consistent for the "true" parameter. Chernozhukov and Hong (2003) have extended these results to the setting considered in this chapter. In particular, they have developed an asymptotic theory for quasi-posterior distributions constructed from models whose parameters are specified through estimating equations.

Their framework can be easily adapted to study the asymptotic behavior of the following generalized bayesian estimator

$$\hat{\theta}^{\rho} = \operatorname{arginf}_{\xi} \sqrt{n} \int \rho(\theta - \xi)\pi_j(\theta|\mathbb{X})\, d\theta, \tag{1.11}$$

where $\rho(u)$ is a non-negative symmetric convex criterion function normalized in such a way that $\rho(0) = 0$ only when $u = 0$, $\rho(u) \leq 1 + |u|^p$, for some $p \geq 1$, and $\int \rho(u - \xi) \exp(-u^\top a u) du$ is uniquely minimized at some point $\xi \in \mathbb{R}^k$. For instance, $\hat{\theta}^{\rho}$ is posterior mean for $\rho(u) = u^2$; $\rho(u) = \sum_{j=1}^{k} |u_j|$ gives instead the posterior median.

Chernozhukov and Hong (2003) show that, as $n \to \infty$, $\pi_{CU}(\theta|\mathbb{X})$ concentrates on H_n, a $1/\sqrt{n}$ neighborhood of θ^*; on this neighborhood the CU posteriors is approximately normal. Formally, the result can be stated as

$$\int_{H_n} |\pi_{CU}(u|\mathbb{X}) - \pi_\infty(u)|\, du \xrightarrow{p} 0,$$

where

$$\pi_\infty(u|\mathbb{X}) = \sqrt{\frac{|V|}{(2\pi)^k}} \exp(-u^\top J u/2).$$

A consequence of the normality of the posterior is that the estimators defined in (1.11) are asymptotically normal, that is,

$$V^{-1}\sqrt{n}(\hat{\theta}^{\rho} - \theta^*) \xrightarrow{d} N(0, I_k);$$

where

$$V = \left(\frac{\partial E g(X, \theta^*)}{\partial \theta^\top}\right) E\left[g(X, \theta^*) g(X, \theta^*)^\top\right] \left(\frac{\partial E g(X, \theta^*)}{\partial \theta}\right).$$

These results are proved under very general conditions that do not require differentiability of $g(\cdot, \theta)$ by using empirical process theory and Law of Large

Numbers and Central Limit Theorem for Donsker classes. When g is differentiable, V coincides with the semiparametric efficiency bound of Chamberlain (1987).

The results of Chernozhukov and Hong (2003) can be extended to the other posteriors discussed in Section (1.4) by showing that on the same shrinking neighborhood of θ^* all the distributions are close to $\pi_{CU}(\theta|\mathbb{X})$. Indeed, it can be shown that

$$\lim_{n\to\infty} \int_{H_n} |\pi_j(\theta|\mathbb{X}) - \pi_{CU}(\theta|\mathbb{X})| = o_p(1), \quad j = \{GMM, NCU, EL, ET\},$$

which implies that the asymptotic analysis of Chernozhukov and Hong (2003) extend to the other posterior distributions.

1.7 Bayesian validity

If the approximate posteriors are to be used from a Bayesian point of view, it is important to asses whether they provide valid posterior inference. Strictly speaking, a posterior is valid if the likelihood function is based on the conditional density of the data given the parameters. If one takes such a narrow view, the posterior distributions of Section (1.4) are clearly all invalid and so is the resulting inference.

Monahan and Boos (1992) introduce a milder concept of validity for posteriors based on likelihood functions that are different from the conditional distribution of the data given the parameter. Under this definition, posterior inference is valid if the likelihood function satisfies a simple coverage condition.

Definition 1 (Validity by Coverage) *A likelihood function $f(\mathbb{X}|\theta)$ is valid by coverage if, for every continuous proper prior distribution $\pi(\theta)$, the statistic*

$$a(\mathbb{X}, \theta) = \int_{-\infty}^{\theta} \frac{f(\mathbb{X}|u)\pi(u)}{f(\mathbb{X})} du,$$

is unconditionally distributed as Unif$(0, 1)$.

The rationale behind this definition of validity is that if the posterior is based on the conditional density of the data, then $a(\mathbb{X}, \theta)$ is unconditionally Unif$(0, 1)$ for every continuous prior distribution. To see this, let $f^*(\mathbb{X}|\theta)$ denote the "true" conditional density. Then, the joint density of $(\mathbb{X}, \beta)$ is

$f^*(\mathbb{X},\beta)=f^*(\mathbb{X}|\theta)\pi(\theta)$, and thus

$$\begin{aligned}\Pr\big(a(x,\theta)\leqslant z\big) &= \int\int 1(a(\mathrm{x},u)\leq z)f^*(x,u)du d\mathrm{x}\\ &= \int\int 1(a(\mathrm{x},u)\leq z)\pi(u|\mathrm{x})du f^*(x)dx\\ &= \int z f^*(x)dx = z\end{aligned}$$

Examining the deviations of $a(\mathbb{X},\theta)$ from $\mathrm{Unif}(0,1)$ can be thought as a *goodness of fit* test and the distance between the two distributions of $a(\mathrm{x},\beta)$ provides a metric to asses the quality of resulting the Bayesian inference. The validity by coverage concept leads to a simulation method which consists in generating θ^s, $s=1,\ldots,S$, independently from the pre-specified prior $\pi(\theta)$ and then generate data $x^{(s)}$ according to $f(\cdot|\theta^s)$ and compute the statistic $a(\mathrm{x}^{(\mathrm{s})},\theta^{(s)})$. The quality of the nonparametric likelihoods can be gauged either graphically—by means of quantile-quantile plots—or formally—by calculating Kolmogorov-Smirnov distances from uniformity of the distribution of a_k.

We study the performance of the posteriors derived in Section 1.4 for a two moment, one parameter problem. The moment vector is

$$g(X,\theta)=\begin{pmatrix} X-\theta\\ X^2-\theta^2-2\theta\end{pmatrix}. \tag{1.12}$$

The distribution of X is $N(\theta,1)$. Given n realizations of X, parametric Bayesian inference about θ is based on a posterior proportional to

$$e^{-\sum_{i=1}^n(\theta-x_i)^2/2}\pi(\theta).$$

Instead of the true normal likelihood, we consider using the nonparametric likelihoods based on EL, ET and CUE implied probabilities.

As a summary, Table 1.1 gives the Kolmogorov-Smirnov distances from uniformity of the distribution of $a(\mathbb{X},\theta)$ for the case of a uniform prior with different variances. For all of cases considered, the distributions of $a(\mathbb{X},\beta)$ for $f_{EL}(\mathbb{X},\theta)$, $f_{ET}(\mathbb{X},\theta)$ and $f_{CU}(\mathbb{X},\theta)$ are statistically indistinguishable from $\mathrm{Unif}(0,1)$ when n is larger than 50. For $f_{NCU}(\mathbb{X},\theta)$, the Kolmogorov-Smirnov null hypothesis of uniformity is rejected at $n=\{20,50\}$ for the more disperse priors.

The results of this simple simulation exercise suggest that it can feasible to consider a Bayesian inferential procedure based on the posteriors discussed in this chapter. Of course, some caveats should be kept in mind. First, the set of simulations is limited: only uniform priors with different variances have been considered. The concept of validity by coverage requires that the statistics $a(\mathbb{X},\theta)$ be uniform for all possible priors. We used a uniform prior because it simplifies the numerical integrations required to calculate the quantities

			$\sup_t \lvert F_n(t) - (t) \rvert$		
			n = 20	n = 50	n = 100
Uniform Prior	$Var = .33$	$f_{EL}(\mathbb{X}\lvert\theta)$	0.040	0.024	0.023
			(0.018)	*(0.333)*	*(0.423)*
		$f_{ET}(\mathbb{X}\lvert\theta)$	0.060	0.023	0.018
			(0.000)	*(0.3854)*	*(0.710)*
		$f_{CU}(\mathbb{X}\lvert\theta)$	0.047	0.036	0.026
			(0.0000)	*(0.042)*	*(0.256)*
		$f_{NCU}(\mathbb{X}\lvert\theta)$	0.078	0.030	0.025
			(0.000)	*(0.148)*	*(0.311)*
	$Var = 1.33$	$f_{EL}(\mathbb{X}\lvert\theta)$	0.050	0.025	0.015
			(0.001)	*(0.312)*	*(0.895)*
		$f_{ET}(\mathbb{X}\lvert\theta)$	0.050	0.03	0.026
			(0.001)	*(0.147)*	*(0.264)*
		$f_{CU}(\mathbb{X}\lvert\theta)$	0.072	0.035	0.024
			(0.000)	*(0.044)*	*(0.333)*
		$f_{NCU}(\mathbb{X}\lvert\theta)$	0.070	0.047	0.029
			(0.0002)	*(0.002)*	*(0.180)*
	$Var = 3$	$f_{EL}(\mathbb{X}\lvert\theta)$	0.055	0.0290	0.0288
			(0.000)	*(0.159)*	*(0.165)*
		$f_{CU}(\mathbb{X}\lvert\theta)$	0.0661	0.0230	0.029
			(0.000)	*(0.405)*	*(0.139)*
		$f_{CU}(\mathbb{X}\lvert\theta)$	0.070	0.021	0.019
			(0.000)	*(0.470)*	*(0.640)*
		$f_{NCU}(\mathbb{X}\lvert\theta)$	0.057	0.050	0.029
			(0.000)	*(0.001)*	*(0.150)*

Table 1.1: (Overidentified location model) Kolmogorov-Smirnov distances of the empirical distributions of $a(\text{x}, \beta)$ for $\mathscr{L}^{el}$, $\mathscr{L}^{et}$, and $\mathscr{L}^{cue}$ with uniform prior centered at 0 and variances .33, 1.33, and 3, sample sizes of $n = \{20, 50, 100\}$, and simulation size of $S = 1,500$. *p-values* are reported in parentheses.

involved in the calculation of $a(\mathbb{X}, \theta)$. Second, we are considering a very simple model and it is possible that for more involved models the posteriors could be easily invalidated.

1.8 Application

In this section we conduct an empirical investigation using the approximate posteriors discussed so far. We will focus on a reanalysis of data from the Pennsylvania Reemployment Bonus Demonstration described in detail in Corson et al. (1992). The dataset collects information about randomized controlled experiment conducted in Pennsylvania by the U.S. Department of Labor between July 1988 and October 1989. The objective of the experiment was to test the effectiveness of cash bonuses paid for early reemployment in shortening the length of insured unemployment spells. During the enrollment period, claimants who became unemployed and registered for unemployment benefits in one of the 12 selected local offices throughout the state were randomly assigned either to a control group or an experimental treatment groups. In the control group the existing rules of the unemployment insurance system applied. Individuals in the treatment groups were offered a cash bonus if they became reemployed in a full-time, working more than 32 hours per week. A detailed description of the characteristics of claimants under study is presented in Koenker and Bilias (2001) which has information on age, race, gender, number of dependents, location in the state, existence of recall expectations, and type of occupation.

We follow Koenker and Xiao (2002) and we model the log-duration of the individual UI benefits as a linear conditional quantile function

$$Q_{\log(T)}(\tau|D, X) = D\beta(\tau) + X^\top\theta(\tau),$$

where the vector X contains demographic characteristics of the subject (sex, race, age indicators and number of dependents) and other variables that are relevant to the experiment and D is the treatment indicator. The parameter of interest is $\beta(\tau)$, that can be interpreted as the treatment effect at quantile τ. For brevity, we will concentrate on the median case, that is $\tau = 0.5$.

Frequentist inference is easily obtained by applying the methodology of Koenker and Bassett Jr (1978). In particular, the point estimate of $\beta(.5)$ is -0.1279, with a standard error of 0.043, numbers that indicates a statistically significant effect of the cash bonus on the UI benefits duration of about 14%. Approximate Bayesian inference can be obtained by using the posteriors based on the following estimating equations

$$E\left[\begin{pmatrix} D \\ X \end{pmatrix} \mathbb{1}_{\log(T)-D\beta(\tau)-X^\top\theta(\tau)} - \tau\right] = 0,$$

and we put a uniform independent prior $U(-10, 10)$ on all the parameters of the model. We simulate from $\pi_{EL}(\beta|\mathbb{X})$, $\pi_{ET}(\beta|\mathbb{X})$, $\pi_{CU}(\beta|\mathbb{X})$, and $\pi_{NCU}(\beta|\mathbb{X})$ using the Metropolis-Hastings algorithm described in Section 1.5.

In Figure 1.3 we present a concise visual representation of the results of the model. Each of the panels of the Figure illustrate the histogram of ten thousands draws from the posterior draws of $\beta(.5)$ for each posterior considered. Posterior medians, means, and the highest probability credible intervals are summarized in Table 1.2. All posteriors give basically very similar inference, the main difference being the credible regions that are wider for π_{CU} and π_{NCU}. Notice also, that these credible intervals are also shorter than the confidence interval based on the frequentist point estimator and the normal calibration of the asymptotic distribution of the quantile regression estimator.

1.9 Conclusions

This chapter has surveyed different approximate posteriors to conduct Bayesian inference. These posteriors are useful when the only information about the parameters of interest is expressed in terms of estimating equations. These posteriors are based on likelihood functions that use ideas behind frequentist estimations. These likelihoods have, at least in a simple model, good Bayesian and frequentist properties. We have illustrated their usefulness in a simple application using data from

	Mean	Median	lower	upper
$\pi_{EL}(\beta \mid \mathbb{X})$	-0.143	-0.145	-0.227	-0.049
$\pi_{ET}(\beta \mid \mathbb{X})$	-0.149	-0.138	-0.225	-0.049
$\pi_{CU}(\beta \mid \mathbb{X})$	-0.144	-0.146	-0.223	-0.057
$\pi_{NCU}(\beta \mid \mathbb{X})$	-0.144	-0.146	-0.223	-0.056

Table 1.2: Pennsylvania Reemployment Bonus Demonstration. Posterior inference for the median treatment effect of cash bonus on unemployment duration.

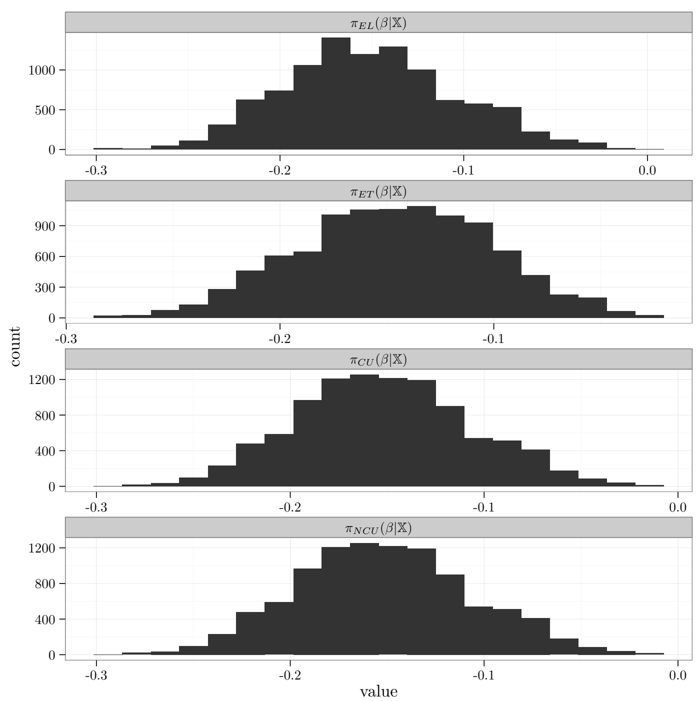

Figure 1.3: Pennsylvania Reemployment Bonus Demonstration. Posterior inference for the median treatment effect of cash bonus on unemployment duration.

Bibliography

Chamberlain, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 34, 305–334.

Chamberlain, G. and G. W. Imbens (2003): "Nonparametric Applications of Bayesian Inference," *Journal of Business and Economic Statistics*, 21, 12–18.

Chen, S. X. and H. Cui (2007): "On the second-order properties of empirical likelihood with moment restrictions," *Journal of Econometrics*, 141, 492–516.

Chernozhukov, V. and C. Hansen (2005): "An IV model of quantile treatment effects," *Econometrica*, 73, 245–261.

Chernozhukov, V. and H. Hong (2003): "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115, 293–346.

Conley, T. G., C. B. Hansen, R. E. McCulloch, and P. E. Rossi (2008): "A semi-parametric Bayesian approach to the instrumental variable problem," *Journal of Econometrics*, 144, 276–305.

Corson, W., P. Decker, S. Dunstan, and S. Keransky (1992): "Pennsylvania 'reemployment bonus' demonstration: Final Report," Tech. rep., Unemployment Insurance Occasional Paper, 92-1, Dept of Labor, Employment and Training Administration, Washington, DC.

DiCiccio, T., P. Hall, and J. Romano (1991): "Empirical likelihood is Bartlett-correctable," *The Annals of Statistics*, 19, 1053–1061.

DiCiccio, T. and J. Romano (1990): "Nonparametric confidence limits by resampling methods and least favorable families," *International statistical review*, 58, 59–76.

Ferguson, T. (1973): "A Bayesian Analysis of some nonparametric problems," *The Annals of Statistics*, 1, 209–230.

——— (1974): "Prior of distributions on spaces of probability measures," *The Annals of Statistics*, 2, 615–629.

Gallant, R. A. and H. White (1988): *A unified theory of estimation and inference for nonlinear dynamic models*, New York: Basil Blackwell.

Gasparini, M. (1995): "Exact Multivariate Bayesian Bootstrap Distributions of Moments," *The Annals of Statistics*, 23, 762–768.

Hahn, J. (1997): "Bayesian Bootstrap of the Quantile Regression Estimator: A Large Sample Study," *International Economic Review*, 38, 206–213.

Hansen, L. P. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica*, 50, 1029–54.

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

Hirano, K. (2002): "Semiparametric Bayesian inference in autoregressive panel data models," *Econometrica*, 70, 781–799.

Ibragimov, I. A. and R. Z. Has'minskii (1981): *Statistical estimation: Asymptotic theory*, New York: Springer-Verlag.

Imbens, G. W. (1997): "One-Step Estimators for Over-Identified Generalized Method of Moments Models," *Review of Economic Studies*, 64, 359–83.

Imbens, G. W. and T. Lancaster (1994): "Combining micro and macro data in microeconometric models," *The Review of Economic Studies*, 61, 655–680.

Kim, J. Y. (2002): "Limited Information Likelihood and Bayesian Analysis," *Journal of Econometrics*, 107, 175–93.

Kitamura, Y. and M. Stutzer (1997): "An information-theoretic alternative to generalized method of moments estimation," *Econometrica*, 65, 861–74.

Koenker, R. and G. Bassett Jr (1978): "Regression quantiles," *Econometrica: journal of the Econometric Society*, 33–50.

Koenker, R. and Y. Bilias (2001): "Quantile regression for duration data: A reappraisal of the Pennsylvania Reemployment Bonus Experiments," *Empirical Economics*, 26, 199–220.

Koenker, R. and Z. Xiao (2002): "Inference on the quantile regression process," *Econometrica*, 70, 1583–1612.

Lancaster, T. (1994): "Bayes WESML: Posterior Inference from Choice-Based Samples," Unpublished manuscript, Brown University, Providence, RI.

——— (2004): *An introduction to modern Bayesian econometrics*, Wiley-Blackwell.

Lazar, N. A. (2003): "Bayesian empirical likelihood," *Biometrika*, 90, 319–326.

Monahan, J. F. and D. Boos (1992): "Proper likelihoods for Bayesian analysis," *Biometrika*, 79, 271–278.

Newey, W. K. and D. McFadden (1994): "Estimation and inference in large samples," in *Handbook of Econometrics*, ed. by R. Engle and D. McFadden, Amsterdam: North-Holland, 2113–2245.

Newey, W. K. and R. J. Smith (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219–55.

Owen, A. (1990): "Empirical likelihood ratio confidence regions," *The Annals of Statistics*, 18, 90–120.

——— (2001): *Empirical likelihood*, vol. 92, Chapman & Hall/CRC.

Owen, A. B. (1988): "Empirical likelihood ratio confidence intervals for a single functional," *Biometrika*, 75, 237–49.

Pfanzagl, J. and W. Wefelmeyer (1978): "A third-order optimum property of the maximum likelihood estimator," *Journal of Multivariate Analysis*, 8, 1–29.

Poirier, D. J. (2011): "Bayesian interpretations of heteroskedastic consistent covariance estimators using the informed Bayesian bootstrap," *Econometric Reviews*, 30, 457–468.

Powell, J. (1986): "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143–155.

QIN, J. AND J. LAWLESS (1994): "Empirical likelihood and general estimating equations," *Annals of Statistics*, 22, 300–325.

RAGUSA, G. (2011): "Minimum divergence, generalized empirical likelihoods, and higher order expansions," *Econometric Reviews*, 30, 406–456.

ROBERT, C. P. AND G. CASELLA (2004): *Monte Carlo Statistical Methods*, Springer, second ed.

RUBIN, D. (1981): "Bayesian Bootstrap," *Annals of Statistics*, 9, 130–34.

SCHENNACH, S. C. (2005): "Bayesian Exponentially Tilted Empirical Likelihood," *Biometrika*, 92, 31–46.

SIMS, C. A. (2002): "Interview with Christopher A. Sims," *Journal of Business and Economic Statistics*, 20, 448–49.

WHITE, H. (1980): "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, 48, 817–838.

YIN, G. (2009): "Bayesian generalized method of moments," *Bayesian Analysis*, 4, 191–207.

YU, K. AND J. STANDER (2007): "Bayesian analysis of a Tobit quantile regression model," *Journal of Econometrics*, 137, 260–276.

ZELLNER, A. (1991): "Bayesian methods and entropy in economics and econometrics," *Maximum Entropy and Bayesian Methods*, 17–31.

——— (1994): "Bayesian method of moments/instrumental variables analysis of mean and regression models," *Manuscript, Graduate School of Business, University of Chicago.*

——— (1997): "The Bayesian Method of Moments (BMOM): Theory and Applications," *Advances in Econometrics*, 12, 85–105.

ZELLNER, A. AND J. TOBIAS (2001): "Further Results on Bayesian Method of Moments Analysis of the Multiple Regression Model," *International Economic Review*, 42, 121–139.