BAYESIAN ECONOMETRICS

VICTOR CHERNOZHUKOV

Bayesian econometrics employs Bayesian methods for inference about economic questions using economic data. In the following, we briefly review these methods and their applications.

Suppose a data vector $X = (X_1, ..., X_n)$ follows a distribution with a density function $p_n(x|\theta)$ which is fully characterized by some parameter vector $\theta = (\theta_1, ..., \theta_d)'$. Suppose that the prior belief about θ is characterized by a density $p(\theta)$ defined over a parameter space Θ, a subset of a Euclidian space $\mathbb{R}^d$. Using Bayes' rule to incorporate the information provided by the data, we can form posterior beliefs about the parameter θ, characterized by the posterior density

$$p_n(\theta|X) = p_n(X|\theta)p(\theta)\mathrm{c}, \quad c = 1/\int_\Theta p_n(X|\tilde{\theta})p(\tilde{\theta})d\tilde{\theta}. \tag{1}$$

The posterior density $p_n(\theta|X)$, or simply $p_n(\theta)$, describes how likely it is that a parameter value θ has generated the observed data X. We can use the posterior density to form optimal point estimates and optimal hypotheses tests. The notion of optimality is minimizing mean posterior loss, using various loss functions. For example, the posterior mean

$$\hat{\theta} = \int_\Theta \theta p_n(\theta)d\theta, \tag{2}$$

is the point estimate that minimizes posterior mean squared loss. The posterior mode θ^* is defined as the maximizer of the posterior density, and it is the decision that minimizes the posterior mean Dirac loss. When the prior density is flat, the posterior mode turns out to be the maximum likelihood estimator. The posterior quantiles characterize the posterior uncertainty about the parameter, and they can be used to form confidence regions for the parameters of interest (Bayesian credible regions). The posterior α-quantile $\hat{\theta}_j(\alpha)$ for θ_j (the j-th component of the parameter vector) is the number c such that $\int_\Theta 1\{\theta_j \le c\} p_n(\theta) d\theta = \alpha$.

Properties of Bayesian procedures in both large and small samples are as good as the properties of the procedures based on maximum likelihood. These properties have been developed by Laplace (1818), Bickel and Yahav (1969), and Ibragimov and Hasminskii (1981), among others. With mild regularity conditions (which hold in many econometric applications), the properties include (a) consistency and asymptotic normality of the point estimates, including asymptotic equivalence and efficiency of the posterior mean, mode, and median, (b) asymptotic normality of the posterior density, and (c) asymptotically correct coverage of Bayesian confidence intervals, (d) average risk optimality of Bayesian estimates in small and hence large samples. The regularity conditions for properties (a) and (b) require that the true parameter θ_0 is well identified and that the data's density $p_n(x|\theta)$ is sufficiently smooth in the parameters. Mathematically, property (a) means that

$$\sqrt{n}(\hat{\theta} - \theta_0) \approx \sqrt{n}(\theta^* - \theta_0) \approx \sqrt{n}(\hat{\theta}(1/2) - \theta_0) \approx_d N(0, J^{-1}), \tag{3}$$

where J equals the information matrix $\lim_n -\frac{1}{n}\frac{\partial^2 E \ln p_n(X|\theta_0)}{\partial\theta\partial\theta'}$, $\approx$ indicates agreement up to a stochastic term that approaches zero in large samples, and $\approx_d N(0, J^{-1})$ means "approximately distributed as a normal random vector with mean 0 and variance matrix J^{-1}." These estimators are asymptotically efficient in the sense of having smallest variance J^{-1} in the class of asymptotically unbiased estimators. Property (b) is that $p_n(\theta)$ is approximately equal to a normal density with mean $\hat{\theta}$ and variance J^{-1}/n. Property (c) means that in large samples

$$Prob[\hat{\theta}_j(\alpha/2) \le \theta_{0j} \le \hat{\theta}_j(1 - \alpha/2)] \approx 1 - \alpha. \tag{4}$$

In non-regular cases, such as in structural auction and search models, consistency and correct coverage properties also continue to hold (Chernozhukov and Hong 2004). Property (d) is implied by the defining property of the Bayes estimators that they minimize the posterior mean risk (Lehmann and Casella 1998). The property continues to hold in non-regular cases, which proved especially useful in non-regular econometric models (Hirano and Porter 2003, Chernozhukov and Hong 2004).

The explicit dependency of Bayesian estimates on the prior is both a virtue and a drawback. Priors allow us to incorporate information available from previous studies and various economic restrictions. When no prior information is available, diffuse priors can be used. Priors can have a large impact on inferential results in small samples and in any other cases where the identifiability of parameters crucially relies on restrictions brought by the prior. In such cases, selection of priors requires a substantial care: see Chamberlain and Imbens (2003, 2004) for an example concerning simultaneous equations and Uhlig (2005) for an example dealing with sign restrictions in structural vector autoregressions. On the other hand, priors should have little impact on the inferential results when the identifiability of parameters does not crucially rely on the prior and when sample sizes are large.

The appealing theoretical properties of Bayesian methods have been known for many years, but computational difficulties prevented their wide use. Closed form solutions for estimators such as (2) have been derived only for very special cases. The recent emergence of Markov Chain Monte Carlo (MCMC) algorithms has diminished the computational challenge and made these methods attractive in a variety of practical applications, see e.g. Robert and Casella (2004) and Liu (2001). The idea of MCMC is to simulate a possibly dependent random sequence, $(\theta^{(1)}, ..., \theta^{(B)})$, called a chain, such that stationary density of the chain is the posterior density $p_n(\theta)$. Then we approximate integrals such as (2) by the averages of the chain, that is $\hat{\theta} \approx \sum_{k=1}^{B} \theta^{(k)}/B$. For computation of posterior quantiles, we simply take empirical quantiles of the chain. The leading MCMC method is the Metropolis-Hastings (MH) algorithm, which includes, for example, the random walk algorithm with Gaussian increments generating the candidate points for the chain. Such random walk is characterized by an initial point u_0 and a one-step move that consists of drawing a point η according to a Gaussian distribution centered on the current point u with covariance matrix $\sigma^2 I$, then, moving to η with probability $\rho = \min\{p_n(\eta)/p_n(u), 1\}$ and staying at u with probability $1 - \rho$. The MH algorithm is often combined with the Gibbs sampler, where the latter updates components of θ individually or in blocks. The Gibbs sampler can also speed up computation when the posterior for some components of θ is available in a closed form. MCMC algorithms have been shown to be computationally efficient in a variety of cases.

The classical econometric applications of Bayesian methods mainly dealt with the classical linear regression model and the classical simultaneous equation model, which admitted closed form solutions (Zellner 1996, Poirier 1995). The emergence of MCMC has enabled researchers to attack a variety of complex non-linear problems. The recent examples of important problems that have been solved using Bayesian methods include: (1) discrete choice models (Albert and Chib 1993, Lancaster 2004), (2) models with limited-dependent variables (Geweke 2005), (3) non-linear panel data models with individual heterogeneity (McCulloch and Rossi 1994, Lancaster 2004), (4) structural vector autoregressions in macroeconomics, including models with sign restrictions (Uhlig 2005), (5) dynamic discrete decision processes (Geweke, Keane, and Runkle 1997, Geweke 2005), (6) dynamic stochastic equilibrium models (Smets and Wouters 2003, Del Negro and Schorfheide 2004), (7) time series models in finance (Fiorentini, Sentana, and Shephard 2004, Johannes and Polson 2003), and (8) unit root models (Sims and Uhlig 1991). Econometric applications of the methods are rapidly expanding.

There are also recent developments that break away from the traditional parametric Bayesian paradigm. Ghosh and Ramamoorthi (2003) develop and review several nonparametric Bayesian methods. Chamberlain and Imbens (2003) develop Bayesian methods based on the multinomial framework of Ferguson (1973, 1974). In models with moment restrictions and no parametric likelihood available, Chernozhukov and Hong (2003) propose using an empirical likelihood function or a generalized method-of-moment criterion function in place of the unknown likelihood $p_n(X|\theta)$ in equation

(1). This permits the application of MCMC methods to a variety of moment condition models. As a result there are a growing number of applications of the latter approach to nonlinear simultaneous equations, empirical game-theoretic models, risk forecasting, and asset-pricing models. The literature both on theoretical and practical aspects of various non-parametric Bayesian methods is rapidly expanding.

The following bibliography includes some of the classical works as well as a sample of contemporary works on the subject. The list is by no means exhaustive.

References

Albert, J. H., and S. Chib (1993): "Bayesian analysis of binary and polychotomous response data," *J. Amer. Statist. Assoc.*, 88(422), 669–679.

Bickel, P. J., and J. A. Yahav (1969): "Some Contributions to the Asymptotic Theory of Bayes Solutions," *Z. Wahrsch. Verw. Geb*, 11, 257–276.

Chamberlain, G., and G. Imbens (2004): "Random effects estimators with many instrumental variables," *Econometrica*, 72(1), 295–306.

Chamberlain, G., and G. W. Imbens (2003): "Nonparametric applications of Bayesian inference," *J. Bus. Econom. Statist.*, 21(1), 12–18.

Chernozhukov, V., and H. Hong (2003): "An MCMC approach to classical estimation," *J. Econometrics*, 115(2), 293–346.

——— (2004): "Likelihood estimation and inference in a class of nonregular econometric models," *Econometrica*, 72(5), 1445–1480.

Del Negro, M., and F. Schorfheide (2004): "Priors from General Equilibrium Models for VARs," *International Economic Review*, 45, 643–673.

Ferguson, T. S. (1973): "A Bayesian analysis of some nonparametric problems," *Ann. Statist.*, 1, 209–230.

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

Fiorentini, G., E. Sentana, and N. Shephard (2004): "Likelihood-based estimation of latent generalized ARCH structures," *Econometrica*, 72(5), 1481–1517.

Geweke, J. (2005): *Contemporary Bayesian econometrics and statistics*, Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ.

Geweke, J. F., M. P. Keane, and D. E. Runkle (1997): "Statistical inference in the multinomial multiperiod probit model," *J. Econometrics*, 80(1), 125–165.

GHOSH, J. K., AND R. V. RAMAMOORTHI (2003): *Bayesian nonparametrics*, Springer Series in Statistics. Springer-Verlag, New York.

HIRANO, K., AND J. R. PORTER (2003): "Asymptotic efficiency in parametric structural models with parameter-dependent support," *Econometrica*, 71(5), 1307–1338.

IBRAGIMOV, I. A., AND R. Z. HASMINSKII (1981): *Statistical estimation*, vol. 16 of *Applications of Mathematics*. Springer-Verlag, New York, Asymptotic theory, Translated from the Russian by Samuel Kotz.

JOHANNES, M., AND N. POLSON (2003): "MCMC Methods for Continuous-Time Financial Econometrics," in *Handbook of Financial Econometrics*. North-Holland, Amsterdam, forthcoming.

LANCASTER, T. (2004): *An introduction to modern Bayesian econometrics*. Blackwell Publishing, Malden, MA.

LAPLACE, P.-S. (1818): *Théorie analytique des probabilités*. Éditions Jacques Gabay (1995), Paris.

LEHMANN, E., AND G. CASELLA (1998): *Theory of Point Estimation*. Springer.

LIU, J. S. (2001): *Monte Carlo strategies in scientific computing*, Springer Series in Statistics. Springer-Verlag, New York.

MCCULLOCH, R., AND P. E. ROSSI (1994): "An exact likelihood analysis of the multinomial probit model," *J. Econometrics*, 64(1-2), 207–240.

POIRIER, D. J. (1995): *Intermediate statistics and econometrics*. MIT Press, Cambridge, MA, A comparative approach.

ROBERT, C. P., AND G. CASELLA (2004): *Monte Carlo statistical methods*, Springer Texts in Statistics. Springer-Verlag, New York, second edn.

SIMS, C. A., AND H. UHLIG (1991): "Understanding unit rooters: a helicopter tour," *Econometrica*, 59(6), 1591–1599.

SMETS, F., AND R. WOUTERS (2003): "An estimated dynamic stochastic general equilibrium model of the euro area," *J. European Economic Association*, 1, 527–549.

UHLIG, H. (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 52, 381–419.

ZELLNER, A. (1996): *An introduction to Bayesian inference in econometrics*, Wiley Classics Library. John Wiley & Sons Inc., New York, Reprint of the 1971 original, A Wiley-Interscience Publication.