Bayesian Bootstrap Multivariate Regression

Thomas Heckelei[1] and Ron C. Mittelhammer[2]
University of Bonn and Washington State University

TWP-2002-109

Abstract

A Bayesian Bootstrap Multivariate Regression (BBMR) procedure is presented that allows robust Bayesian posterior analysis of traditional multivariate regression models. BBMR does not require the specification of a parametric family for the likelihood function and instead uses a bootstrapped likelihood based on the sampling distribution of location and scale estimators. A mixing algorithm for implementing the BBMR procedure automatically incorporates the scale invariant ignorance prior on the covariance matrix. BBMR also allows a flexible choice of prior distributions and can be implemented as a generic algorithm in standard statistical software independently of the actual choice of prior distribution. Monte Carlo evidence is provided showing accuracy and robustness of the approach in representing posterior distributions even under the application of highly non-linear mappings.

***JEL Classification*: C11, C15, C30**

***Keywords:* Bayesian Inference, bootstrapping, robust likelihood, multivariate regression, simultaneous equation mappings**

---

[1] University of Bonn, Germany; tel: +49-228-732323; e-mail: heckelei@agp.uni-bonn.de

[2] Washington State University, USA; tel: (509) 335-1706; e-mail: mittelha@wsu.edu

We are grateful for valuable discussions during the course of this research with Arnold Zellner, Luc Bauwens, Hermann van Dijk, and John Monahan. We also thank three anonymous referees and the associate editor for their constructive and very helpful comments. All remaining errors are the sole responsibility of the authors.

1. Introduction

This paper presents a Bayesian Bootstrap Multivariate Regression (BBMR) procedure that allows robust Bayesian analysis of a system of regression equations. The BBMR approach is easily automated, widely applicable, robust with respect to the likelihood function associated with the underlying data generating process, and flexible with respect to the choice of prior distribution. These characteristics remove impediments and contribute to a wider use of Bayesian techniques in applied econometric research.

BBMR uses a bootstrapped form of Monte Carlo Integration (MCI) to analyze posterior distributions of model parameters. In this regard it is related to the work of Kloek and van Dijk (1978), van Dijk and Kloek (1980), Zellner, Bauwens and van Dijk (1988), and Geweke (1989 and 1991). With continuously increasing computing power, MCI is a convincing solution to the problem of analytical tractability in multidimensional integration problems often encountered in Bayesian analysis of econometric models. It allows a very flexible choice of prior distribution and can be implemented as a generic algorithm in standard statistical software independently of the actual choice of prior distribution (Geweke, 1991).

In addition, BBMR does not require the specification of a parametric family for the likelihood function, and in this and some other respects shares conceptual parallels (discussed further ahead) with the Bayesian Method of Moments introduced by Zellner (1996). In lieu of a likelihood specification, BBMR uses a bootstrapped likelihood based on the joint sampling distribution of location and scale estimators. The outcomes of this bootstrap procedure serve as outcomes of an importance function for the MCI evaluation of posterior expectations based on an importance sampling scheme (see Geweke, 1989 and 1991 for importance sampling in the case of normality). The BBMR method is distinct from the Bayesian Bootstrap originated by Rubin(1981), and surveyed by Rubin and Shenkar (1998) and Shao and Tu(1995), in that the latter uses order statistics from a Uniform(0,1) distribution to assign posterior probabilities to bootstrap samples, is intricately related to the Dirichlet distribution in terms of both conjugate priors and resulting posterior distributions of parameters, and is difficult to generalize and implement in multi-equation, highly parameterized models. The BBMR, on the other hand, defines posteriors via bootstrapping the empirical distribution function of the data, is very general with regard to prior distribution choice, and can be applied straightforwardly to highly parameterized multi-equation models.

The paper is organized as follows: First, a short introduction to MCI analysis of the traditional multivariate regression model is given to establish notation and to set the inferential context for the BBMR approach. Then, the BBMR-algorithm is introduced as a robust method for performing posterior inference. Extensions of BBMR to Bayesian analyses of simultaneous equation models follow next, based on a generalized version of the unrestricted reduced form mappings presented by Zellner, Bauwens and van Dijk (1988). Simulation results are presented in the penultimate section to assess the performance of the BBMR in the simultaneous equations context using "Klein Model I", as summarized in Theil (1971). The final section presents conclusions and areas for future research.

2. Monte Carlo Integration Analysis of the Traditional Multivariate Regression Model

We begin with an overview of a general computer-driven approach for performing Bayesian analysis of the traditional multivariate regression model. We do not require that the joint sampling distribution be in the normal parametric family, and we deviate from standard textbook expositions (e.g. Zellner; 1971) by arranging terms differently in the representation of the posterior distribution to provide a better motivation for the BBMR procedure introduced in the next section. Since later sections will utilize multivariate regression formulations in the context of analyzing reduced forms of simultaneous equations systems, we use the conventional notation for the reduced form in representing our system of regression equations. In particular, let

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\Pi} + \mathbf{V}, \tag{1}$$

where $\mathbf{Y}$ is a $(n\times m)$ matrix of observations on m endogenous variables, $\mathbf{X}$ is a $(n\times k)$ matrix of observations on k exogenous variables, $\boldsymbol{\Pi}$ is a $(k\times m)$ matrix of regression coefficients, and $\mathbf{V}$ is a $(n\times m)$ matrix representing n iid outcomes of a $1\times m$ disturbance vector having a joint density function $g(\mathbf{V}|\mathbf{0},\boldsymbol{\Sigma})$ with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. Then $\mathbf{Y}$ has some corresponding probability density function $f(\mathbf{Y}|\boldsymbol{\Pi},\boldsymbol{\Sigma})$. Assuming that a prior probability density on the model parameters, $p(\mathbf{\Pi},\mathbf{\Sigma})$, is independent regarding $\mathbf{\Pi}$ and $\mathbf{\Sigma}$ so that $p(\mathbf{\Pi},\mathbf{\Sigma}) = p(\mathbf{\Pi})\, p(\mathbf{\Sigma})$ and letting $L(\mathbf{\Pi},\mathbf{\Sigma}|\mathbf{Y}) \equiv f(\mathbf{Y}|\mathbf{\Pi},\mathbf{\Sigma})$ denote the likelihood function, the marginal posterior probability density of the regression coefficients can be represented as

$$\begin{aligned} h(\mathbf{\Pi}|\mathbf{Y}) &\propto p(\mathbf{\Pi}) \int p(\mathbf{\Sigma})\, L(\mathbf{\Pi},\mathbf{\Sigma}|\mathbf{Y})\, d\mathbf{\Sigma} \\ &\propto p(\mathbf{\Pi}) \frac{\int p(\mathbf{\Sigma})\, L(\mathbf{\Pi},\mathbf{\Sigma}|\mathbf{Y})\, d\mathbf{\Sigma}}{\int \left[\int p(\mathbf{\Sigma})\, L(\mathbf{\Pi},\mathbf{\Sigma}|\mathbf{Y})\, d\mathbf{\Sigma} \right] d\mathbf{\Pi}} \\ &= p(\mathbf{\Pi})\; L_{\mathbf{\Sigma}}(\mathbf{\Pi}|\mathbf{Y}), \end{aligned} \tag{2}$$

where $L_{\mathbf{\Sigma}}(\mathbf{\Pi}|\mathbf{Y})$ is to be interpreted as a marginal likelihood, derived via taking the expectation of $L(\mathbf{\Pi},\mathbf{\Sigma}|\mathbf{Y})$ with respect to the prior information $p(\mathbf{\Sigma})$ and then normalizing the result to have unit total mass. If an ignorance prior on $\mathbf{\Sigma}$ is employed, which will be the case for the remainder of this paper, then the representation of the posterior in (2) separates the available information on $\mathbf{\Pi}$ into two parts: 1) the prior information $p(\mathbf{\Pi})$, and 2) the information contained in the data as represented via the model, $L_{\mathbf{\Sigma}}(\mathbf{\Pi}|\mathbf{Y})$.

Expectations of various functions of $\mathbf{\Pi}$ taken with respect to the posterior (2) are typical measures used to summarize the information contained in the posterior about the unknown model parameters. The posterior mean represents the Bayesian point estimate for the unknown true value of $\mathbf{\Pi}$ that minimizes a quadratic loss function (Mittelhammer, Judge, and Miller, 2000, p. 662). Posterior variances and tail probabilities measure the precision of the posterior knowledge relating to $\mathbf{\Pi}$. Posterior probabilities associated with specified subsets of the parameter space provide measures of confidence regarding the location of model parameters. In general, posterior expectations of functions of $\mathbf{\Pi}$, say $E[g(\mathbf{\Pi})]$, are represented by

$$E[g(\mathbf{\Pi})] = \int g(\mathbf{\Pi})\, h(\mathbf{\Pi}|\mathbf{Y})\, d\mathbf{\Pi}\,. \tag{3}$$

Monte Carlo Integration (MCI) is often the preferred method of evaluating the integrals that define posterior expectations, since flexibility in choosing prior densities, the use of complicated functions of $\mathbf{\Pi}$ and/or high parameter dimensionality prohibit the use of analytical integration tools (Geweke, 1989). Conceptually, Monte Carlo evaluation of expectations is a straightforward application of Kolmogorov's Strong Law of Large Numbers. In particular, if N iid outcomes from h($\mathbf{\Pi}$|$\mathbf{Y}$) in equation (3) are available, say $\mathbf{\Pi}_i$'s $\sim$ *iid* h($\mathbf{\Pi}$|$\mathbf{Y}$) , i = 1,...,N, and if |E[g($\mathbf{\Pi}$)]| = c < ∞, then

$$\frac{1}{N}\sum_{i=1}^{N} g(\mathbf{\Pi}_i) \xrightarrow{as} E[g(\mathbf{\Pi})]. \tag{4}$$

Thus for large enough N, a simple average of the outcomes of g($\mathbf{\Pi}_i$) provides an arbitrarily close approximation to E[g($\mathbf{\Pi}$)].

Unfortunately, it is often the case that a flexible choice of prior distribution and likelihood function will result in a situation where random sampling from the posterior distribution of $\mathbf{\Pi}$ is difficult or impossible. In this situation, an importance sampling approach can be useful with L_Σ($\mathbf{\Pi}$|$\mathbf{Y}$) serving as the importance function. Referring to (2), the expectations with respect to the posterior h($\mathbf{\Pi}$|$\mathbf{Y}$) can be expressed as

$$E[g(\mathbf{\Pi})] = \frac{\int g(\mathbf{\Pi})\, p(\mathbf{\Pi})\, L_\Sigma(\mathbf{\Pi}|\mathbf{Y})\, d\mathbf{\Pi}}{\int p(\mathbf{\Pi})\, L_\Sigma(\mathbf{\Pi}|\mathbf{Y})\, d\mathbf{\Pi}} . \tag{5}$$

Letting $\mathbf{\Pi}_i$'s $\sim$ *iid* $L_\Sigma(\mathbf{\Pi}|\mathbf{Y})$ for i=1,...N, and applying Kolmogorov's Strong Law of Large Numbers to both the numerator and denominator of (5), E[g($\mathbf{\Pi}$)] can be approximated by a prior-weighted average of the form

$$E[g(\mathbf{\Pi})] = \frac{\frac{1}{N}\sum_{i=1}^{N} g(\mathbf{\Pi}_i)\, p(\mathbf{\Pi}_i)}{\frac{1}{N}\sum_{i=1}^{N} p(\mathbf{\Pi}_i)} = \frac{\sum_{i=1}^{N} g(\mathbf{\Pi}_i)\, p(\mathbf{\Pi}_i)}{\sum_{i=1}^{N} p(\mathbf{\Pi}_i)} . \tag{6}$$

In case of an ignorance prior on $\mathbf{\Pi}$ (p($\mathbf{\Pi}$) = constant) it is clear that estimated expectations generated via (6) reduce to a simple average, $\frac{1}{N}\sum_{i=1}^{N} g(\mathbf{\Pi}_i)$, of functions of the outcomes $\mathbf{\Pi}_i$ (see also Zellner, Bauwens, and van Dijk, 1988, p. 47).

If a normal likelihood function is assumed then L_Σ($\mathbf{\Pi}$|$\mathbf{Y}$), defined in equation (2), is a matrix student-t distribution denoted by T(v,$\mathbf{\Pi}$,$\mathbf{S}$), and defined by

$$L_\Sigma(\mathbf{\Pi} \mid \mathbf{Y}) \propto |\mathbf{S} + (\mathbf{\Pi} - \hat{\mathbf{\Pi}})\mathbf{X}'\mathbf{X}(\mathbf{\Pi} - \hat{\mathbf{\Pi}})|^{\,(v+k)/2}, \tag{7}$$

where v = n-k and $\mathbf{S} = (\mathbf{Y} - \mathbf{X}\,\boldsymbol{\Pi})'(\mathbf{Y} - \mathbf{X}\,\boldsymbol{\Pi}) = \mathbf{V}'\mathbf{V}$ is the residual sum of squares and cross products matrix. Outcomes of T(v,$\boldsymbol{\Pi}$,$\mathbf{S}$) can be generated easily (an algorithm is given, for example, in the appendix of Zellner, Bauwens and van Dijk, 1988), so that it can serve as the importance function for the evaluation of posterior expectations. Geweke (1986) used this approach to impose inequality restrictions on regression coefficients.

The preceding discussion indicates that for the normal multivariate regression model a general and easily implementable approach for analyzing posterior distributions of regression coefficients is available. However, leaving the realm of normality requires the development of some other (importance) sampling approach that may or may not be as tractable as the preceding one. The need to develop methodological solutions on a case-by-case basis is unappealing and certainly impedes the use of Bayesian techniques in applied econometric analyses. Furthermore, the parametric family of the error distribution is almost always unknown, and requiring that one be chosen anyway can lead to a high propensity for model specification error. We suggest an alternative approach, BBMR, that is 1) robust with respect to the underlying error distribution, 2) preserves the advantages of the normal distribution-MCI approach regarding flexibility in choosing the prior and 3) can be implemented generically, once and for all, in standard statistical software.

3. Bayesian Bootstrap Multivariate Regression

In order to obtain an approach to Bayesian analysis of the multivariate regression model that is robust with respect to the joint probability distribution of the underlying data generating process, we first substitute for the marginal likelihood function, $L_{\Sigma}(\boldsymbol{\Pi}|\mathbf{Y})$, a "Multivariate Regression-Structure Likelihood," $L_{\Sigma}(\boldsymbol{\Pi}|\hat{\boldsymbol{\Pi}},\mathbf{S})$, conditioned on the joint outcome of location and scale estimators representing information contained in the data. We then approximate this likelihood by an empirical likelihood defined via a bootstrap procedure that simultaneously provides a random sample from the empirical likelihood to use in place of outcomes of $L_{\Sigma}(\boldsymbol{\Pi}|\mathbf{Y})$ for approximating posterior expectations via (6).

3.1 Multivariate Regression-Structure Likelihood

Analogous to $L_{\Sigma}(\mathbf{\Pi}|\mathbf{Y})$ in equation (2) we define the normalized (to unit mass) Multivariate Regression-Structure-Likelihood of $\mathbf{\Pi}$ as

$$L_{\Sigma}(\mathbf{\Pi}|\hat{\mathbf{\Pi}},\mathbf{S}) = \frac{\int p(\mathbf{\Sigma})\, L(\mathbf{\Pi},\mathbf{\Sigma}|\hat{\mathbf{\Pi}},\mathbf{S})\, d\mathbf{\Sigma}}{\int \left[\int p(\mathbf{\Sigma})\, L(\mathbf{\Pi},\mathbf{\Sigma}|\hat{\mathbf{\Pi}},\mathbf{S})\, d\mathbf{\Sigma}\right] d\mathbf{\Pi}} \tag{8}$$

The likelihood function $L(\mathbf{\Pi},\mathbf{\Sigma}\,|\,\hat{\mathbf{\Pi}},\mathbf{S})$ in (8) is effectively conditioned on least squares estimators of the parameters $(\mathbf{\Pi},\mathbf{\Sigma})$, namely $\hat{\mathbf{\Pi}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ and $n^{-1}\mathbf{S} = n^{-1}\hat{\mathbf{V}}'\hat{\mathbf{V}}$, that are derived from the structure of the multivariate regression problem (1), hence the name "Multivariate Regression-Structure Likelihood." The idea of constructing a likelihood conditional on estimators of unknown parameters has been used previously in related contexts by Boos and Monahan (1986), Doksum and Lo (1990), Davison, Hinkley, and Worton (1992), and Pettit (1982, 1983), among others, and in the same context for the univariate linear model by Heckelei and Mittelhammer (1996). Monahan and Boos (1992) have presented criteria that can be used to determine when such likelihoods are defensible via the probability calculus for Bayesian posterior inference in the case where proper prior densities are used. Essentially, such likelihoods for a parameter vector $\mathbf{\Theta}$ are defensible when they are based on the densities of statistics that are parameterized entirely by $\mathbf{\Theta}$.

It is apparent that in the current context information contained in the data is now exclusively represented via the information contained in the estimators $\hat{\mathbf{\Pi}}$ and $\mathbf{S}$. This is akin to the Bayesian Method Of Moments (BMOM) approach in which the specification of moment information inherent in the least squares estimates $\hat{\mathbf{b}}$ and s^2 form the basis for defining posterior moments of the $\boldsymbol{\beta}$ vector in a regression model setting in the absence of an explicit likelihood function specification (Zellner, 1996). A fundamental difference between BMOM and the BBMR method introduced in this paper, which will be developed ahead, is that posterior distributions in the BMOM context are derived analytically from moment constraints that are assumed to hold and that satisfy a maximum entropy criterion (Zellner, 1996; Zellner and Highfield, 1988; Ryu, 1993; Mittelhammer, Judge, and Miller, 2000), whereas the BBMR approach uses bootstrap data resampling, and a nonparametric empirical distribution function representation of the likelihood function for simulating posterior probabilities and expectations.

In the case where $\hat{\mathbf{\Pi}}$ and $\mathbf{S}$ are sufficient statistics for $(\mathbf{\Pi},\mathbf{\Sigma})$, $L_{\mathbf{\Sigma}}(\mathbf{\Pi}|\mathbf{Y})$ and $L_{\mathbf{\Sigma}}(\mathbf{\Pi}|\hat{\mathbf{\Pi}},\mathbf{S})$ are informationally identical, and in fact are both equal to the aforementioned matrix student-t distribution under normality. Whenever $\hat{\mathbf{\Pi}}$ and $\mathbf{S}$ are not sufficient statistics the use of $L_{\mathbf{\Sigma}}(\mathbf{\Pi}|\hat{\mathbf{\Pi}},\mathbf{S})$ leads to some loss of information on the parameter vector $\mathbf{\Pi}$. However, this type of information loss is to some extent inevitable in empirical analyses, being completely avoidable only in cases of perfect knowledge regarding the parametric family of distributions underlying the data generating process. Moreover, the common assumption of multivariate normality also represents all data information via the statistics $\hat{\mathbf{\Pi}}$ and $\mathbf{S}$, and thus precipitates information loss, and in addition constitutes a specification error, when normality does not hold.

3.2 Mixing Algorithm for Likelihood Sampling

Analogous to the multivariate regression model under normality summarized in section 2, one can approximate posterior expectations using (6) if a random sample from the Multivariate Regression-Structure Likelihood, $L_{\mathbf{\Sigma}}(\mathbf{\Pi}|\hat{\mathbf{\Pi}},\mathbf{S})$, can be obtained. In order to define an algorithm for sampling from this likelihood, we express $L_{\mathbf{\Sigma}}(\mathbf{\Pi}|\hat{\mathbf{\Pi}},\mathbf{S})$ in terms of a mixed distribution involving the marginal posterior distribution of $\mathbf{\Sigma}$ (which does the mixing) and the marginal likelihood function of $\mathbf{\Pi}$, conditional on $\mathbf{\Sigma}$ (which is mixed over the support of $\mathbf{\Sigma}$):

$$L_{\mathbf{\Sigma}}(\mathbf{\Pi}|\hat{\mathbf{\Pi}},\mathbf{S}) \propto \int L(\mathbf{\Pi}|\hat{\mathbf{\Pi}},\mathbf{S},\mathbf{\Sigma})\, h(\mathbf{\Sigma}|\hat{\mathbf{\Pi}},\mathbf{S})\, d\mathbf{\Sigma}, \tag{9}$$

where $h(\mathbf{\Sigma}|\hat{\mathbf{\Pi}},\mathbf{S}) \propto p(\mathbf{\Sigma})\, L(\mathbf{\Sigma}|\hat{\mathbf{\Pi}},\mathbf{S})$ is the marginal posterior of $\mathbf{\Sigma}$. It follows from this representation that a random outcome from $L_{\mathbf{\Sigma}}(\mathbf{\Pi}|\hat{\mathbf{\Pi}},\mathbf{S})$ can be obtained by first drawing a random outcome of $\mathbf{\Sigma}$ from $h(\mathbf{\Sigma}|\hat{\mathbf{\Pi}},\mathbf{S})$, say $\mathbf{\Sigma}_*$, and then drawing a random outcome of $\mathbf{\Pi}$ from $L(\mathbf{\Pi}|\hat{\mathbf{\Pi}},\mathbf{S},\mathbf{\Sigma}_*)$ (normalized to unit mass).

In order to operationalize this mixing algorithm, we first show how it relates to the sampling distributions of $\hat{\mathbf{\Pi}}$ and $\mathbf{S}$. We consider a slightly modified version of model (1) given by

$$\mathbf{Y} = \mathbf{X}\,\mathbf{\Pi} + \mathbf{U}\mathbf{T}, \tag{1'}$$

where the rows of the $(n\times m)$ matrix of errors, $\mathbf{U}$, are iid outcomes from $g(\mathbf{U}|\ \mathbf{0},\mathbf{I})$ having a mean vector of $\mathbf{0}$ and a covariance matrix of $\mathbf{I}$, the density of $\mathbf{V}_i = \mathbf{V[i,.]} = \mathbf{U[i,.]T}$ is $g(\mathbf{V}_i \mid \mathbf{0},\mathbf{T'T})$ for any conformable $\mathbf{T}$ with full column rank, and the $(m\times m)$ matrix $\mathbf{T}$ is a matrix for which $\boldsymbol{\Sigma} = \mathbf{T}'\mathbf{T}$, so that $\mathbf{U[i,.]T} = \mathbf{V}_i = \mathbf{V[i,.]} \sim g(\mathbf{V}_i \mid \mathbf{0},\boldsymbol{\Sigma})\ \forall i$. Everything else is as defined in (1).

Note that the foregoing distributional assumptions characterize the class of error density families for which the ensuing mixing algorithm will be robust. In particular, when m=1, the robust class includes all symmetric density families with mean zero, as well as any family of skewed densities that can be defined via scaling of a random variable having a parameterless density with mean zero. For m≥2, all density families in the elliptically contoured class having mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$ are in the robust class, including families such as Pearson II, Pearson VII, multivariate T, LaPlace, Bessel, Uniform (elliptical), and multivariate normal (Johnson, 1987, chapter 6; Johnson and Kotz, 1972, p. 297). This follows straightforwardly from the fact that the characteristic function of an elliptically contoured random vector $\mathbf{V}_i$ with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$ is given by $\phi_{\mathbf{V}_i}(\mathbf{t}) = \psi(c\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})$ for some function $\psi(\cdot)$, where c is a known numerical constant that is specific to a density family (Cambanis, et. al, 1981, p. 368 and Theorem 4).

Given the preceding assumptions, the outcomes of the estimators $\hat{\boldsymbol{\Pi}}$ and $\mathbf{S}$ are distributionally equivalent, in the sense of their marginal distributions, to the outcomes of

$$\begin{Bmatrix} \hat{\boldsymbol{\Pi}} = \boldsymbol{\Pi} + (\mathbf{X'X})^{-1}\mathbf{X'UT} \\ \mathbf{S} = \mathbf{T}'(\mathbf{U'MU})\mathbf{T} \end{Bmatrix}. \tag{10}$$

where $\mathbf{M} = \mathbf{I}-\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X}'$. Fixing $\hat{\boldsymbol{\Pi}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ and $\mathbf{S} = \hat{\mathbf{V}}'\hat{\mathbf{V}}$ at their observed values and interpreting $\boldsymbol{\Pi}$ and $\boldsymbol{\Sigma}$ as random variables, outcomes of $\boldsymbol{\Pi}$ and $\boldsymbol{\Sigma}$ can be solved for as[3]

$$\begin{Bmatrix} \boldsymbol{\Pi} = \hat{\boldsymbol{\Pi}} - (\mathbf{X'X})^{-1}\mathbf{X'UT} \\ \boldsymbol{\Sigma} = \mathbf{S}^{1/2}(\mathbf{U'MU})^{-1}\mathbf{S}^{1/2} \end{Bmatrix}. \tag{11}$$

[3] Here, and henceforth, we use the matrix square root notation $\mathbf{Q}^{1/2}$ to denote any matrix square root of $\mathbf{Q}$, such that $\mathbf{Q} = \mathbf{Q}^{1/2}\mathbf{Q}^{1/2}$ for symmetric matrix square roots, and $\mathbf{Q} = \mathbf{Q}^{1/2\,\prime}\mathbf{Q}^{1/2}$ for nonsymmetric matrix square roots.

To show that the second equation in (11) holds, pre and post multiply the second equation in (10) by $\mathbf{S}^{-1/2}$ to obtain $\mathbf{I} = \mathbf{S}^{-1/2}\mathbf{T}'(\mathbf{U}'\mathbf{M}\mathbf{U})\mathbf{T}\mathbf{S}^{-1/2}$. It follows that $\mathbf{S}^{-1/2}\mathbf{T}' = (\mathbf{U}'\mathbf{M}\mathbf{U})^{-1/2}$ satisfies the preceding equation, so that $\mathbf{T}' = \mathbf{S}^{1/2}(\mathbf{U}'\mathbf{M}\mathbf{U})^{-1/2}$, in which case $\boldsymbol{\Sigma} = \mathbf{T}'\mathbf{T} = \mathbf{S}^{1/2}(\mathbf{U}'\mathbf{M}\mathbf{U})^{-1}\mathbf{S}^{1/2}$.

If the parametric family of the underlying error distribution were known and new samples could be drawn from it, then (11) identifies an empirical method of implementing the mixing concept presented in (9). Specifically, one could draw a random sample of size n from g($\mathbf{U}$| $\mathbf{0}$,$\mathbf{I}$), resulting in the (n×m) matrix $\mathbf{U}_*$, and then calculate using (11) and the given value of $\mathbf{S}$, an outcome of $\boldsymbol{\Sigma}$, say $\boldsymbol{\Sigma}_*$, which is interpretable as an outcome from h($\boldsymbol{\Sigma}$| $\hat{\boldsymbol{\Pi}}$,$\mathbf{S}$) (further motivated below). Then $\mathbf{U}_*$ can be post-multiplied by a matrix $\mathbf{T}_*$, for which $\boldsymbol{\Sigma}_* = \mathbf{T}_*'\mathbf{T}_*$, to generate an outcome of $\boldsymbol{\Pi}$ that is conditioned on the covariance matrix $\boldsymbol{\Sigma}_*$. as $\boldsymbol{\Pi} = \hat{\boldsymbol{\Pi}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}_*\mathbf{T}_*$. This outcome of $\boldsymbol{\Pi}$ is effectively an outcome from (the normalized to unit total mass) L($\boldsymbol{\Pi}$| $\hat{\boldsymbol{\Pi}}$,$\mathbf{S}$,$\boldsymbol{\Sigma}_*$). Repeating the procedure N times produces a random sample of size N from L_Σ($\boldsymbol{\Pi}$| $\hat{\boldsymbol{\Pi}}$,$\mathbf{S}$), which is equivalent to mixing L($\boldsymbol{\Pi}$| $\hat{\boldsymbol{\Pi}}$,$\mathbf{S}$,$\boldsymbol{\Sigma}$) over h($\boldsymbol{\Sigma}$| $\hat{\boldsymbol{\Pi}}$, $\mathbf{S}$) and random sampling from the mixture.

3.3 Ignorance Prior on $\boldsymbol{\Sigma}$ Implied by the Mixing Algorithm

We now examine the type of marginal posterior distribution for $\boldsymbol{\Sigma}$ that is implied by the transformation in (11) and find that it automatically incorporates the standard ignorance prior on $\boldsymbol{\Sigma}$ that is typically used in Bayesian analyses of the multivariate regression model. Under the conditions given for the distributional model in (1') and denoting the sampling distribution of $\mathbf{S}$ as f($\mathbf{S}$|$\boldsymbol{\Sigma}$), we have

$$\mathbf{S} = \mathbf{T}'(\mathbf{U}'\mathbf{M}\mathbf{U})\mathbf{T} \sim f(\mathbf{S}|\boldsymbol{\Sigma}) = f(\mathbf{S}|\mathbf{T}) \tag{12}$$

where we can alternatively parameterize on the matrix $\mathbf{T}$ because the distribution of $\mathbf{UT}$ depends only on the value of $\mathbf{T}'\mathbf{T} = \boldsymbol{\Sigma}$. Letting $\mathbf{W} = \mathbf{U}'\mathbf{M}\mathbf{U} = (\mathbf{T}')^{-1}\mathbf{S}\mathbf{T}^{-1}$ and noting that $|\partial\mathbf{W} / \partial\mathbf{S}| = |\mathbf{T}|^{-(m+1)}$, because of the symmetry of $\mathbf{S}$ (Deemer and Olkin, 1951) the distribution of $\mathbf{W}$ is given via change of variables as

$$\eta(\mathbf{W}) = f(\mathbf{T}'\mathbf{W}\mathbf{T}|\mathbf{T})|\mathbf{T}|^{m+1}. \tag{13}$$

For a given $\mathbf{S}$ it follows from (11) that $\mathbf{W} = \mathbf{S}^{1/2}\boldsymbol{\Sigma}^{-1}\mathbf{S}^{1/2}$, so that differentiating with respect to $\boldsymbol{\Sigma}$ yields $|\partial\mathbf{W} / \partial\boldsymbol{\Sigma}| = |\mathbf{S}|^{(m+1)/2}\,|\boldsymbol{\Sigma}|^{-(m+1)}$ (Deemer and Olkin, 1951). The change of variable transformation from the distribution of $\mathbf{W}$ to the distribution of $\boldsymbol{\Sigma}$ then yields

$$\begin{aligned} h(\boldsymbol{\Sigma}|\mathbf{S}) &\propto f(\mathbf{T}'\,\mathbf{S}^{1/2}\,\boldsymbol{\Sigma}^{-1}\mathbf{S}^{1/2}\,\mathbf{T}|\mathbf{T})|\mathbf{T}|^{m+1}|\boldsymbol{\Sigma}|^{-(m+1)} \\ &\propto f(\mathbf{S}|\boldsymbol{\Sigma})|\boldsymbol{\Sigma}|^{(m+1)/2}|\boldsymbol{\Sigma}|^{-(m+1))} = f(\mathbf{S}|\boldsymbol{\Sigma})|\boldsymbol{\Sigma}|^{-(m+1)/2} \end{aligned} \tag{14}$$

since $|\mathbf{T}| = |\mathbf{T}'| = |\mathbf{T}'\mathbf{T}|^{1/2} = |\boldsymbol{\Sigma}|^{1/2}$, which implies that the regression structure likelihood for $\boldsymbol{\Sigma}$, represented by $f(\mathbf{S}|\boldsymbol{\Sigma})$, is post-multiplied by the standard ignorance prior $p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(m+1)/2}$ to obtain the marginal posterior for $\boldsymbol{\Sigma}$, $h(\boldsymbol{\Sigma}|\hat{\boldsymbol{\Pi}},\mathbf{S}) \equiv h(\boldsymbol{\Sigma}|\mathbf{S})$. It can be straightforwardly verified that in the multivariate normal case, (14) yields the appropriate inverted Wishart distribution.

3.4 Bootstrapping the Mixing Algorithm: The BBMR Algorithm

If the error distribution family were known and random sampling from this distribution were reasonably straightforward, then the mixing algorithm described in the preceding section, together with (6), would represent a feasible method of conducting Bayesian analysis of the multivariate regression model. Of course, other computationally intensive and possibly more efficient approaches might be available for performing Bayesian inference when the error distribution is known and sampling from it is tractable, such as MCI and importance sampling. Nevertheless, the mixing approach will approximate analytical posterior inference arbitrarily closely for large enough samples if $\hat{\boldsymbol{\Pi}}$ and $\mathbf{S}$ are sufficient statistics for the parameters of the regression model. Furthermore, even in cases where there is some information loss associated with the mixing approach, computational simplicity may favor its use. As a practical matter, problem situations in which the error distribution is known are rare, and thus the theoretical ideal of lossless data information transfer to the posterior is, in actuality, generally unachievable.

In the more typical case where the error distribution family is unknown, a robust variation on the mixing approach can be defined using the empirical distribution function of the regression residuals, $\mathrm{EDF}(\hat{\mathbf{V}})$, to approximate the true error distribution, $g(\mathbf{V}|\,\mathbf{0},\boldsymbol{\Sigma})$, and ultimately to approximate $L_{\boldsymbol{\Sigma}}(\boldsymbol{\Pi}|\hat{\boldsymbol{\Pi}},\mathbf{S})$ by a nonparametric estimate $L_{\boldsymbol{\Sigma}^*}(\boldsymbol{\Pi}|\hat{\boldsymbol{\Pi}},\mathbf{S})$. In this case bootstrapping is a relevant sampling technique, where new $n{\times}m$ residual matrices are constructed by random sampling residual vectors, with replacement, from the observed row vectors of residuals contained in the matrix $\hat{\mathbf{V}} = \mathbf{MUT}$ where $\mathbf{\Sigma} = \mathbf{T'T}$.

Denote a bootstrap sample of the OLS-residual vectors as $\mathbf{V}_* = \mathbf{U}_*(\mathbf{S}/n)^{1/2}$, where $\mathbf{U}_* = \mathbf{V}_*(\mathbf{S}/n)^{-1/2}$ is then used to approximate a random sample of size n from g($\mathbf{U}$| $\mathbf{0}$, $\mathbf{I}$). According to (11) a bootstrapped outcome of $\mathbf{\Pi}$, $\mathbf{\Pi}_*$, can be generated by first computing a bootstrapped outcome of $\mathbf{\Sigma}$, say $\mathbf{\Sigma}_*$, using $\mathbf{U}_*$ in place of $\mathbf{U}$ in the second equation of (11), and then substituting the square root $\mathbf{\Sigma}_*^{1/2}$ into the first equation in (11) in place of $\mathbf{T}$. Considering this procedure in detail allows substitutions to be made that eliminate the need for sequential calculations of $\mathbf{\Sigma}_*$ and $\mathbf{\Pi}_*$ as follows:

$$\left\{\begin{array}{l} \mathbf{\Pi}_* = \hat{\mathbf{\Pi}} - (\mathbf{X'X})^{-1}\mathbf{X'V}_*\mathbf{S}^{-1/2}(\mathbf{SS}_*^{-1}\mathbf{S})^{1/2} \\ \mathbf{\Sigma}_* = \mathbf{S}^{1/2}\left[\left(\mathbf{V}_*(\mathbf{S}/n)^{-1/2}\right)'\mathbf{M}\left(\mathbf{V}_*(\mathbf{S}/n)^{-1/2}\right)\right]^{-1}\mathbf{S}^{1/2} = n^{-1}\mathbf{SS}_*^{-1}\mathbf{S} \end{array}\right\} \tag{15}$$

with $\mathbf{S}_* = \mathbf{V}_*'\mathbf{MV}_*$.

In combination with a specification of the prior distribution, p($\mathbf{\Pi}$), the N outcomes of $\mathbf{\Pi}_*$ from (15) can be used to approximate posterior expectations utilizing equation (6). As the data sample size increases, and EDF($\hat{\mathbf{V}}$) $\rightarrow$ g($\mathbf{V}$| $\mathbf{0}$,$\mathbf{\Sigma}$), the bootstrap sampling becomes equivalent to random sampling from the true error distribution. As the bootstrap sample size increases, posterior expectations become equivalent to analytical expectations based on EDF($\hat{\mathbf{V}}$). Thus, for large enough data and bootstrap sample sizes, the BBMR algorithm will produce posterior expectations based on the true regression structure likelihood for the parameters of the regression model.

4. Extending BBMR to Robust Bayesian Analysis of Simultaneous Equations

A considerable portion of applied econometric analyses involves structural equations that have endogenous explanatory variables, i.e., simultaneous equation systems. Zellner, Bauwens and van Dijk (1988) apply several mappings of unrestricted reduced form coefficients that allow for Bayesian estimation and specification analysis of structural equations based on random samples from the posterior distribution of reduced form coefficients. In this context, a "2SLS-

Mapping" and a "3SLS-Mapping" (see also van Dijk, 1985) are described ahead and later used in Monte Carlo simulations to evaluate the performance of the BBMR approach.

4.1 2SLS and 3SLS Mappings of the Reduced Form

Consider a system of m structural equations represented by

$$\mathbf{Y\Gamma} + \mathbf{XB} + \mathbf{U} = [\mathbf{0}] \qquad (16)$$

where $\mathbf{Y}$ is a n×m matrix of endogenous variables, $\mathbf{X}$ is a (n×k) matrix of predetermined variables, $\mathbf{\Gamma}$ and $\mathbf{B}$ are (m×m) and (k×m) matrices of coefficients of endogenous and predetermined variables, respectively, and $[\mathbf{0}]$ is a n×m matrix of zeros. Assume the system is normalized so that $\mathbf{\Gamma}$ has negative unit values on its diagonal. The reduced form of (16) can then be written as

$$\mathbf{Y} = \mathbf{X\Pi} + \mathbf{V} \qquad (17)$$

with

$$\mathbf{V} = -\mathbf{U\Gamma}^{-1} \text{ and } \mathbf{\Pi} = -\mathbf{B\Gamma}^{-1}. \qquad (18)$$

Now let $\mathbf{\Gamma}_0 = \mathbf{\Gamma} + \mathbf{I}$, with $\mathbf{I}$ being a m×m identity matrix, and rearrange the second equation in (18) to define

$$\mathbf{\Pi}(\mathbf{\Gamma}_0-\mathbf{I}) = -\mathbf{B} \Leftrightarrow \mathbf{\Pi} = \mathbf{\Pi\Gamma}_0 + \mathbf{B} \qquad (19)$$

Premultiplying by $\mathbf{X}$ yields a generalized version of equation 2.24 in Zellner, Bauwens and van Dijk (1988) as

$$\mathbf{X\Pi} = \mathbf{X\Pi\Gamma}_0 + \mathbf{XB} = \overline{\mathbf{Z}}\delta, \qquad (20)$$

where $\overline{\mathbf{Z}} = [\mathbf{X\Pi} \;\; \mathbf{X}]$ and $\delta = \begin{bmatrix} \mathbf{\Gamma}_0 \\ \mathbf{B} \end{bmatrix}$. To allow for possible errors in the exact restrictions implied by (20) a (n×m) discrepancy matrix Δ (instead of a vector in the case of single equation analyses) is introduced leading to an exact relationship between $\mathbf{X}$, $\overline{\mathbf{Z}}$, and $\mathbf{\Lambda}$ given by

$$\mathbf{X\Pi} = \overline{\mathbf{Z}}\,\delta + \Delta. \qquad (21)$$

Let Δ_i be the ith column of Δ, which represents discrepancies in the exact restrictions of (20) corresponding to the ith structural equation. Minimizing each of the discrepancy functions $\Delta_i'\Delta_i$, for i = 1,...,m, separately with respect to δ_i (representing the ith column of δ) defines what Zellner, Bauwens and van Dijk call the 2SLS-mapping, because it resembles the way 2SLS-estimation maps Π into δ, as

$$\delta_i^{2SLS} \; = \; (\overline{\mathbf{Z}}_i'\overline{\mathbf{Z}}_i)^{-1}\overline{\mathbf{Z}}_i'\mathbf{X}\mathbf{\Pi}_i, \; i = 1,\ldots,m. \tag{22}$$

Here $\overline{\mathbf{Z}}_i$ denotes a matrix of the columns of $\mathbf{X\Pi}$ and $\mathbf{X}$ that appear in the ith equation, and $\mathbf{\Pi}_i$ denoting the ith column of $\mathbf{\Pi}$.

Expanding the single equation mapping concept to systems of equations, a 3SLS Mapping can be defined from (21) as

$$\delta^{3SLS} = \left(\tilde{\mathbf{Z}}'(\mathbf{\Omega}\otimes\mathbf{I})^{-1}\tilde{\mathbf{Z}}\right)^{-1}\tilde{\mathbf{Z}}'(\mathbf{\Omega}\otimes\mathbf{I})^{-1}\operatorname{vec}(\mathbf{X\Pi}) \tag{23}$$

with

$$\tilde{\mathbf{Z}} = \begin{Bmatrix} \overline{\mathbf{Z}}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{Z}}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \overline{\mathbf{Z}}_m \end{Bmatrix}, \qquad \hat{\mathbf{\Omega}} = \begin{Bmatrix} \hat{\omega}_{11} & \hat{\omega}_{12} & \cdots & \hat{\omega}_{1m} \\ \hat{\omega}_{21} & \hat{\omega}_{22} & \cdots & \hat{\omega}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\omega}_{m1} & \hat{\omega}_{m2} & \cdots & \hat{\omega}_{mm} \end{Bmatrix},$$

and $\hat{\omega}_{ij} \; = \; (X\mathbf{\Pi}_i - \overline{\mathbf{Z}}_i\delta_i^{2SLS})'(X\mathbf{\Pi}_j - \overline{\mathbf{Z}}_j\delta_j^{2SLS}) / n$, so that this mapping takes covariances between the different errors in (21) into account. The 3SLS-mapping, δ^{3SLS}, is the solution for δ in (21) that minimizes the quadratic discrepancy function defined by $\operatorname{vec}(\Delta)'(\hat{\mathbf{\Omega}}\otimes\mathbf{I})^{-1}\operatorname{vec}(\Delta)$.

4.2 BBMR and Robust Structural Equation Analysis

The 2SLS- and 3SLS-mappings can be combined with the BBMR algorithm to perform robust Bayesian estimation and specification analysis of the structural coefficients of a system of simultaneous equations. In the first step, BBMR is used to provide random samples from a robust ignorance prior-based posterior of the reduced form coefficients, replacing the random samples based on the matrix student t-distribution used by Zellner, Bauwens and van Dijk (1988). In the second step, these reduced form parameter outcomes are substituted for $\mathbf{\Pi}_i$ and $\mathbf{\Pi}$ in (22) and (23) to calculate outcomes of the 2SLS- and 3SLS-mappings. The mappings represent information about structural coefficients that is contained in the data and bootstrapped outcomes can be interpreted as being drawn from an approximation to the marginal likelihood function of the structural coefficients. Finally, posterior expectations of structural coefficients, or functions thereof ($g(\boldsymbol{\delta})$), are evaluated using a prior ($p(\boldsymbol{\delta})$)-weighted average of the mapping outcomes analogous to equation (6), e.g., for the 3SLS-mapping as

$$\frac{\sum_{i=1}^{N} g(\delta_i^{3SLS})\, p(\delta_i^{3SLS})}{\sum_{i=1}^{N} p(\delta_i^{3SLS})} \quad . \tag{24}$$

where N denotes the bootstrap sample size.

5. Motivation, Design, and Results of Monte Carlo Simulations

The theoretical validity of BBMR is based on the bootstrap's ability to consistently estimate the sampling distribution of $\hat{\mathbf{\Pi}}$ and $\mathbf{S}$. As the data sample size n approaches infinity, the empirical distribution function of the observed residuals, EDF($\hat{\mathbf{V}}$), converges to the true distribution
$g(\mathbf{V}|\ \mathbf{0},\mathbf{\Sigma})$ and accordingly $L_{\Sigma^*}(\mathbf{\Pi}|\hat{\mathbf{\Pi}},\mathbf{S})$ converges to $L_{\Sigma}(\mathbf{\Pi}|\hat{\mathbf{\Pi}},\mathbf{S})$. Finite sample properties of the bootstrap, however, are generally unknown. Monte Carlo simulations of a simpler Bayesian bootstrap regression procedure for single equation models have been quite promising, even for rather small data sample sizes (see Heckelei and Mittelhammer, 1996) but performance of the more complex and more highly parameterized multivariate procedure requires independent investigation. Furthermore, the use of BBMR-outcomes to analyze structural equations through 2SLS- and 3SLS-mappings involves nonlinear functions that might render the approximation error of the bootstrap more significant.

In order for the Monte Carlo experiments to provide a realistic illustration of the performance of the BBMR, we evaluate the procedure within a simultaneous equations structure that has been prevalent in the applied econometrics literature. The design of the Monte Carlo simulations allow an assessment of the crucial properties of a robust estimator, namely the efficiency loss relative to the normal distribution-based approach when the normal probability model is true, and the ability of the procedure to accurately represent characteristics of the true posterior distribution under different assumptions regarding the true probability distribution underlying the data generating process.

In order to measure the approximation error of BBMR in representing posterior distributions, we contrast means, variances, and tail probabilities of the bootstrapped marginal posterior distributions $h_*(\delta_j|\hat{\mathbf{\Pi}},\mathbf{S})$ with their parametric counterparts from $h(\delta_j|\hat{\mathbf{\Pi}},\mathbf{S})$ based on both 2SLS and 3SLS-mappings. The intent of the appraisal is not to evaluate the performance of the unrestricted reduced form mappings per se, but rather to provide measures of the approximation error of the BBMR procedure in the reduced form (i.e., the multivariate regression) setting. The highly nonlinear mappings can accentuate approximation errors in the reduced form posterior, thereby providing a more stringent and/or broader scope to the validation of the methodology, as opposed to only evaluating the accuracy of BBMR in representing the posterior distributions of reduced form coefficients. We also provide results that focus on the multivariate regression setting directly, where we compare bootstrapped marginal posterior distributions $h_*(\pi_j|\hat{\mathbf{\Pi}},\mathbf{S})$ with their parametric counterpart $h(\pi_j|\hat{\mathbf{\Pi}},\mathbf{S})$[4]. Under an ignorance prior on $\mathbf{\Pi}$ these distributions are equivalent to $L_{\mathbf{\Sigma}*}(\pi_j|\hat{\mathbf{\Pi}},\mathbf{S})$ and $L_{\mathbf{\Sigma}}(\pi_j|\hat{\mathbf{\Pi}},\mathbf{S})$, respectively.

We do not evaluate the efficiency loss resulting from the use of $\hat{\mathbf{\Pi}}$ and $\mathbf{S}$ as the only source of data information relative to a case where the functional form of the likelihood function $L(\mathbf{\Pi},\mathbf{\Sigma}|\mathbf{Y})$ is known and $\hat{\mathbf{\Pi}}$ and $\mathbf{S}$ are not sufficient statistics. Such efficiency loss is in a sense unavoidable in empirical work since the analyst rarely knows the true functional form of the likelihood function and, consequently, this comparison is of less practical interest in our context.

Two important sources of approximation error remain in the current simulation context: (1) A finite collection of bootstrapped error samples $\mathbf{V}_*$ from EDF($\hat{\mathbf{V}}$) does not completely represent the full informational content of EDF($\hat{\mathbf{V}}$), so that the appropriate bootstrap sample size is an issue, and (2) EDF($\hat{\mathbf{V}}$) itself is only an approximation to the true error distribution $g(\mathbf{V}|\,\mathbf{0},\mathbf{\Sigma})$ raising questions regarding required data sample sizes and also regarding the usefulness of

---

[4] This comparison was suggested by an anonymous reviewer.

certain transformations of the observed residuals $\hat{\mathbf{V}}$ suggested in the bootstrapping literature (Efron, 1979; Shao and Tu, 1995). In all cases except one (noted ahead) we report the results for sampling with replacement of the untransformed OLS-residuals of the reduced form equations exactly as described in the BBMR-algorithm above, because it was shown to perform as well or better than any of the alternative residual scaling mechanisms (see also Heckelei and Mittelhammer, 1996).

The only exception to reporting the results of bootstrapping only untransformed residuals relates to a promising second order bootstrap correction technique. The technique transforms the collection of bootstrap samples to achieve exact second moment characteristics of the "infinite" bootstrap. In particular, letting $\mathbf{V}_{*i}$, i = 1,...,N, be the collection of N bootstrapped (n×m) residual matrices, the second order bootstrap correction is defined as

$$\mathbf{V}^{S}_{*i} = \mathbf{V}_{*i}\mathbf{P}\mathbf{L}^{-1/2}\hat{\mathbf{S}}^{1/2},\ i = 1,\ldots,N \tag{25}$$

where $\mathbf{P}$ and $\boldsymbol{\Lambda}$ are the eigenvector and diagonal eigenvalue matrices corresponding to the matrix $\overline{\boldsymbol{\Psi}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{V}_{*i}'\,\mathbf{V}_{*i}$, and $\hat{\boldsymbol{\Sigma}}^{1/2}$ is the symmetric square root matrix corresponding to $\hat{\boldsymbol{\Sigma}} = n^{-1}\mathbf{S} = (\mathbf{Y}\text{-}\mathbf{X}\hat{\boldsymbol{\Pi}})'(\mathbf{Y}\text{-}\mathbf{X}\hat{\boldsymbol{\Pi}})/n$. It follows that

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{V}^{S\prime}_{*i}\mathbf{V}^{S}_{*i} = \hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{\Lambda}^{-1/2}\mathbf{P}'\left[\frac{1}{N}\sum_{i=1}^{N} \mathbf{V}_{*i}'\mathbf{V}_{*i}\right]\mathbf{P}\boldsymbol{\Lambda}^{-1/2}\hat{\boldsymbol{\Sigma}}^{1/2} = \hat{\boldsymbol{\Sigma}}, \tag{26}$$

which is the appropriate covariance matrix associated with EDF($\hat{\mathbf{V}}$) and corresponds to the limiting covariance matrix that would be estimated from bootstrapped residuals if the bootstrap sample size were increased toward ∞. This technique is a more generally applicable alternative to "Second Order Balancing" (Graham et al., 1990), and places none of the latter approach's restrictions on bootstrap and data sample sizes that are imposed by Latin Squares designs.

5.1 Simulation Design

We follow Zellner, Bauwens, and Van Dijk (1988) and base our Monte Carlo simulations on Klein's Model I. Variable definitions and additional information about the model that is not reported here can be found in their work, and also in Theil (1971). Interested readers can compare simulated posterior expectations and standard deviations relating to simultaneous equations mappings reported in Zellner, Bauwens and Van Dijk with results presented here.

The simulation results are generated via the following sequence of steps:

1.) $n_{rep} = 1000$ data samples of sample size n = 21 are drawn from Klein's model using the 3SLS-estimates reported by Theil as the representation of the "true" values of the model parameters. Data on the predetermined variables are the actual historical values for the 1921-1941 period reported by Klein (1950). Conditional on the predetermined variables, and given the preceding values of the model parameters, the data is drawn from a simultaneous equation system that contains three behavioral equations (represented by the first three equations in the system) and three identities (the last three equations in the sytem). Specifically, the system takes the following form:

$$\mathbf{Y}_t\mathbf{\Gamma} = \mathbf{X}_t\mathbf{B} + \mathbf{U}_t \tag{27}$$

where

$$\mathbf{Y}_t = \left\{ C_t \;\; I_t \;\; W_t^I \;\; X_t \;\; P_t \;\; D_t \right\},$$

$$\mathbf{\Gamma} = \left\{ \begin{array}{cccccc} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 \\ -0.7901 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & -0.4005 & 1 & -1 & 0 \\ -0.1249 & 0.0131 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right\},$$

$$\mathbf{X}_t = \left\{ 1 \quad t-1931 \quad W_t^G \quad T_t \quad G_t \quad P_{t-1} \quad K_{t-1} \quad X_{t-1} \right\}$$

$$\mathbf{B} = \begin{Bmatrix} 16.44 & 28.18 & 1.8 & 0 & 0 & 0 \\ 0 & 0 & 0.1497 & 0 & 0 & 0 \\ 0.7901 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0.1631 & 0.7557 & 0 & 0 & 0 & 0 \\ 0 & 0.1948 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0.1813 & 0 & 0 & 0 \end{Bmatrix}.$$

The structural errors, $\mathbf{U}_t$, are sampled iid from multivariate probability densities having mean vector $\mathbf{0}$ and a covariance submatrix for the three behavioral equations equal to (all other entries of the complete covariance matrix are zero)

$$\mathbf{\Omega} = \begin{Bmatrix} 4.459 & 2.057 & -1.968 \\ 2.057 & 10.47 & 2.015 \\ -1.968 & 2.015 & 2.600 \end{Bmatrix}.$$

The elements of the $\mathbf{\Omega}$ matrix were chosen to be five times the values of the actual estimated contemporaneous covariance matrix elements calculated from 3SLS residuals and the historical data. The additional variation was introduced to insure that any observed accuracy of the BBMR was not due primarily to the relatively good historical fit of Klein's model. However, we also show simulation results based on the original smaller contemporaneous covariance matrix estimated from the 3SLS residuals for comparison purposes (Table 2). The data sample was generated sequentially (because of the lagged endogenous variables in $\mathbf{X}_t$) using $\mathbf{Y}_t = \mathbf{X}_t\mathbf{B}\mathbf{\Gamma}^{-1} + \mathbf{U}_t\mathbf{\Gamma}^{-1}$. All reported results are restricted to the first model equation, i.e. the consumption function, which is representative of the performance for the other equations.

2.) In the description of how posterior distributions are evaluated, only the notation for structural coefficients is used ahead. However the simulation design for the evaluation of the posteriors of reduced form coefficients is analogous, i.e., π_j can be substituted for δ_j in the ensuring description. The means, $\bar{\delta}_j$, variances, Var(δ_j), and values of δ_j corresponding to the 2nd, 5th, 10th, 90th, 95th, and 98th percentile, denoted as δ_j^{ith}, are calculated from the parametric marginal posterior distributions of the structural coefficients, $h(\delta_j|\hat{\mathbf{\Pi}},\mathbf{S})$. For the case of normally distributed errors this is done using the sampling procedure from the appropriate matrix T-distribution of reduced form coefficients as described in Zellner, Bauwens and van Dijk (1988) followed by an application of the 2SLS/3SLS-mapping. In the case of T-distributed errors we use a procedure equivalent to the BBMR algorithm except that the errors are drawn from the known parametric family parameterized by **S**. For simplicity, and to avoid any unintended alteration of the evaluation of posterior precision, we employ an ignorance prior on δ. Note that the nonlinear character of the mappings of the reduced form coefficients should sufficiently challenge the ability of the BBMR to approximate multivariate posterior distributions without the addition of a complex prior distribution. All sampling based "parametric solutions" are based on 25000 samples.

3.) For each data sample, the BBMR-outcomes of the posterior expectations based on bootstrap sample sizes of n_b = 1000 and 5000 are compared with their parametric counterparts from the same data sample. The reported measures for each marginal posterior $h_j(\delta_j|\hat{\boldsymbol{\Pi}},\mathbf{S})$ (and equivalently for $h_j(\pi_j|\hat{\boldsymbol{\Pi}},\mathbf{S})$), are calculated as follows (the subscript j is suppressed below and bootstrapped outcomes have a "*" subscript):

(a) Root Mean Square Error of bootstrapped posterior mean estimate (denoted in Tables as $(MSE)^{1/2}$, Mean):

$$\sqrt{\frac{1}{n_{rep}}\sum_{i=1}^{n_{rep}}(\bar{\delta}_{*i}-\bar{\delta}_i)^2} \ .$$

b) Bias of posterior mean estimate (Bias, Mean):

$$\frac{1}{n_{rep}}\sum_{i=1}^{n_{rep}}(\bar{\delta}_{*i}-\bar{\delta}_i) \ .$$

c) Root Mean Square Error of bootstrapped posterior variance estimate($(MSE)^{1/2}$, Variance):

$$\sqrt{\frac{1}{n_{rep}}\sum_{i=1}^{n_{rep}}\left[\mathrm{Var}(\delta_{*i}) - \mathrm{Var}(\delta_i)\right]^2} \ .$$

d) Bias of bootstrapped posterior variance estimate (Bias, Variance):

$$\frac{1}{n_{rep}}\sum_{i=1}^{n_{rep}}(\mathrm{Var}(\delta_{*i}) - \mathrm{Var}(\delta)) \ .$$

e) The average bootstrapped probabilities (various quantiles):

$$\frac{1}{n_{rep}} \sum_{i=1}^{n_{rep}} prob_{*i} ,$$

where $prob_{*i}$ is calculated as the proportion of the bootstrapped δ_*'s that are below δ_j^{ith}, relative to the bootstrap sample size.

f) Standard deviation of bootstrapped probabilities (STDV):

$$\sqrt{\frac{1}{n_{rep}-1} \sum_{i=1}^{n_{rep}} (prob_{*i} - \frac{1}{n_{rep}} \sum_{i=1}^{n_{rep}} prob_{*i})^2} .$$

5.2 Simulation Results

5.2.1 Reduced Form Parameters

Performance measures for reduced form coefficients relating to the consumption equation, based on the simulation procedure described in the previous section involving a five-fold inflation of historical variation, a multivariate normal error distribution, and for arbitrarily selected parameters, is given in Table 1. In this case (only) the parametric solution is evaluated analytically based on the appropriate marginal t-distributions of the parameters. The accuracy of the nonparametric mean and variance estimates as measured by the root mean square error is quite good. Since the respective bias measures are smaller than the root mean square errors by an order of magnitude, the largest component of $MSE^{1/2}$ relates to variance. Therefore, one would expect that the general approximation accuracy of BBMR will increase for higher bootstrap sample sizes, second order corrected samples, or lower error variances in the data generating process. Results presented below for the nonlinear mappings partly confirm this hypothesis.

The accuracy in representing the full parametric posterior is remarkable, as evidenced by the estimates of the posterior quantiles and the small deviations associated with those estimates. The standard deviations barely exceed one percentage point in the worst case.

5.2.2 2SLS-Mappings Under Normality – Comparison Between Inflated and Original Variance

Table 2 presents performance measures for 2SLS-mappings. These highly nonlinear mappings of the reduced form coefficients have the potential to accentuate the approximation errors of the BBMR. The results indicate an approximation accuracy with respect to mean and variance measures that is comparable with the preceding results reported for the reduced form coefficients. The representation of the tails of the posterior distribution are also very accurate, where average bootstrapped probabilities show only slightly larger errors than in the previous comparison. Standard deviations are only slightly above those for the reduced form coefficients.

Table 3 is based on the original, non-inflated covariance matrix estimated from 3SLS residuals based on Klein's original historical data. The use of a covariance matrix that is scaled by a factor of 1/5 relative to the covariance matrix employed in all other simulations leads to a uniform reduction of all root mean square errors and standard deviations of probabilities compared to Table 2, as expected. The bias measures, however, can hardly be distinguished between the higher and lower variation runs. The smaller error variation also had an impact on the accuracy of the average bootstrapped probabilities, where all coefficients and quantiles averages are closer to the true values.

5.2.3 2SLS-Mappings Under Normality - Residual Correction and Increased Bootstrap Sample Size

The performance of different estimation methods is certainly more interesting than impacts of changes in the data generation process from a practical point of view, because only the former are generally controllable by the empirical analyst. Two approaches promising an improvement in posterior approximation accuracy are the aforementioned second order sample correction technique and increasing the bootstrap sample size. Table 4 allows these two cases to be compared with the "reference" scenario in Table 2 and with each other for the coefficients δ_2 and δ_4.

Both approaches improve upon the accuracy of the mean and variance estimates presented in Table 2. Increasing the bootstrap sample size from 1000 to 5000 outperforms the second order correction technique if measured with respect to root mean square errors whereas the bias results and the average bootstrapped probabilities indicate no clear advantage for either approach. Correcting the covariance matrix of the reduced form errors alone apparently is not as effective in stabilizing the expectations of the nonlinear mappings as is the large increase in bootstrap sample size, and the latter is certainly quite feasible in applications. This is confirmed upon examining the standard deviations of the bootstrapped probabilities, where the reduction for the increased bootstrap sample size is around 50% compared to the reference scenario whereas the second order correction procedure reduces them by a little more than 10%. These results are somewhat discordant with the single equation case in Heckelei and Mittelhammer (1996) where the second order correction had a stronger impact on distance measures at comparable bootstrap sample sizes ($n_b = 900$) and the impact of increasing the bootstrap sample size was less significant. This may again be explained by the multivariate setting and the nonlinear mappings employed here that place higher demands on the bootstrap. Higher bootstrap sample sizes in the 1000 to 5000 range still seem to improve considerably upon the bootstrap's ability to represent the information contained in the multivariate empirical distribution function.

5.2.4 2SLS-Mappings For T-Distributed Errors

Another important trait of a robust estimation approach is its performance under different probability distributions for the underlying data generating process. Table 5 presents performance measures between BBMR based 2SLS-Mappings and those calculated via a parametric regression structure likelihood, using the mixing algorithm described above, when the disturbances are assumed to be multivariate T-distributed instead of normally distributed. Compared with the results for the normal distribution in Table 2, root mean square errors of mean and variance estimates and standard deviations of bootstrapped probabilities are only slightly increased for T-distributed errors. Mean and variance bias as well as average bootstrapped probabilities indicate somewhat less accuracy than for the normal error case. However, the overall approximation accuracy of the BBMR-based 2SLS-mappings for the multivariate T-distribution is still very good and does not differ from the normal case by amounts of any practical importance. It is interesting to note that analogous simulations that (incorrectly) assume normality and proceed with an analytical parametric Bayesian analysis in the case of T-distributed errors, the details of which are not reported here, are significantly less accurate than the BBMR results with regard to bias measures and average posterior probabilities.

5.2.5 3SLS mappings

All of the types of simulations based on 2SLS-mappings reported here were also done for 3SLS-mappings, but only for the bootstrap sample size of 1000. The general approximation accuracy as well as the findings regarding variations on error variance, second order correction and error distribution paralleled those for the 2SLS-mappings. Therefore, in Table 6, only distance measures for the simulation scenario relating to the five-fold inflated covariance matrix, analogous to the simulation context of Table 2, are presented. The comparison with Table 2 demonstrates the similarity of the results.

6. Conclusions

This paper provided the conceptual rationale and Monte Carlo evidence illustrating the posterior approximation accuracy associated with Bayesian Bootstrap Multivariate Regression (BBMR). The procedure allows for a generic, computer algorithmic Bayesian analysis of the traditional multivariate regression model without specification of a likelihood function and without restrictions on the form of prior densities. Combining BBMR with 2SLS- and 3SLS-mappings allows Bayesian analysis of simultaneous equation systems based on unrestricted reduced forms. Monte Carlo results for a simultaneous equation setting presented in the paper, including posterior distributions of both reduced form parameters (i.e., multivariate regression) and structural equation parameters indicate that the efficiency loss relative to a parametric approach under normality is small. Simulation results with multivariate T-distributed error terms suggest that the BBMR-algorithm may also be robust in the class of elliptically contoured distributions, given that the data information is represented exclusively via a regression structure likelihood defined in terms of the sampling distribution of location and scale estimators. Since the underlying probability model for the regression disturbances is generally unknown, BBMR can be a robust and useful alternative to assuming normality and performing parametric Bayesian analysis with a potentially incorrect likelihood function.

Several areas in need of further research can be identified. First of all, Monte Carlo studies are always of limited generality and additional simulations using other error distributions would provide additional insights into the robustness of BBMR. Second, in the context of non-normal error distributions it would be interesting to further compare the parametric normality-based approach with the BBMR both to evaluate the robustness of the normality assumption, and to assess possible relative improvements in posterior analyses that BBMR can provide over incorrectly assuming normality. Third, the development of a bootstrap algorithm for performing restricted reduced form analyses of simultaneous equation systems would be desirable for cases in which the analyst felt that over identifying restrictions could be imposed with certainty. In this context in which overidentifying information is taken explicitly into account, the important issue of weak identification on the performance of BBMR could be evaluated. Finally, one might consider BBMR based on other robust estimators of location and scale to generate samples from likelihood functions that are potentially more robust than the Regression-Structure-Likelihood.

7. References

Boos, D.D., and J.F. Monahan, 1986. Bootstrap methods using prior information. Biometrika 73, 77-83.

Cambanis, S., S. Huang, and G. Simons, 1981. On the theory of elliptically contoured distributions. J. of Multivariate Analysis 11, 368-385.

Davison, A.C., D.V. Hinkley, and B.J. Worton, 1992. Bootstrap likelihoods. Biometrika 79, 113-30.

Deemer, W.L. and I. Olkin, 1951. The jacobians of certain matrix transformations useful in multivariate analysis. Biometrika 38, 345-67.

Doksum, K.A. and A. Y. Lo, 1990. Consistent and robust Bayes procedures for location based on partial information. Ann. Stat. 18, 443-453.

Efron, B. , 1979. Bootstrap Methods: Another Look at the Jackknife., Annals of Statistics 7, 1-26.

Geweke, J., 1986. Exact inference in the inequality constrained normal linear regression model. J. of Appl. Econometrics 1, 127-41.

________, 1989. Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57, 1317-39.

________, 1991. Generic, algorithmic approaches to Monte Carlo integration in Bayesian inference. Contemporary Mathematics 115, 117-34.

Graham, R.L., D.V. Hinkley, P.W.M. John, and S. Shi, 1990. Balanced design of bootstrap simulations. J. R. Statist. Soc. B. 52, 185-202.

Heckelei, Th. and R.C. Mittelhammer, 1996. Bayesian bootstrap inference via regression structure likelihood, in: T.B. Fomby and R.C. Hill, eds., Advances in Econometrics, Vol. 11A.

Johnson, M.E., 1987. Multivariate statistical simulation. John Wiley & Sons (New York).

Johnson, N.L. and S. Kotz, 1972. Distributions in statistics: Continuous multivariate distributions. John Wiley & Sons (New York).

Judge, G.G., R.C. Hill, W.E. Griffiths, H. Lhtkepohl, and T.C. Lee, 1988. Introduction to the Theory and Practice of Econometrics, 2nd ed., John Wiley & Sons (New York).

Klein, L., 1950. Economic Fluctuations in the United States, 1921-1941. John Wiley and Sons (New York).

Kloek, T. and H.K. van Dijk, 1978. Bayesian estimates of equation system parameters: An application of integration by Monte Carlo. Econometrica 46, 1-19.

Mittelhammer, R.C., G. Judge, and G. Miller. Econometric Foundations, Cambridge University Press, 2000.

Monahan, J.F. and D.B. Boos, 1992. Proper likelihoods for Bayesian analysis. Biometrika 79, 271-8.

Pettit, A.N., 1982. Inference for the linear model using a likelihood based on ranks. J.R. Statist. Soc. B 44, 234-43.

________, 1983. Likelihood based inference using signed ranks for matched pairs. J.R. Statist. Soc. B 45, 287-96.

Rubin, D.B. 1981. The Bayesian Bootstrap. Annals of Statistics 9, 130-134.

Rubin, D.B. and Shenkar, N. 1998. Imputation. In Encyclopedia of Statistical Sciences, Update Volume 2 (S. Kotz, C.B. Read, and D.L. Banks, editors) 336-342, John Wiley and Sons (New York).

Ryu, H. , 1993. Maximum Entropy Estimation of Density and Regression Functions. Journal of Econometrics 56, 397-440.

Shao, J. and D. Tu, 1995. The Jackknife and Bootstrap. Springer-Verlag (New York).

Theil, H., 1971. Principles of Econometrics. John Wiley & Sons (New York).

van Dijk, H.K., 1985: A note on simple algebraic relations between structural and reduced form parameters. Report no. 8549, Econometric Institute, Erasmus University, Rotterdam.

van Dijk, H.K. and T. Kloek, 1980. Further experience in Bayesian analysis using Monte Carlo integration. J. of Econometrics 14, 307-28.

Zellner, A. 1996. Bayesian Method of Moments/Instrumental Variable (BMOM/IV) Analysis of Mean and Regression Models. In J.C. Lee, W.O. Johnson, and A. Zellner (Eds.), Modeling and Prediction Honoring Seymour Geisser, Springer-Verlag (New York).

Zellner, A., 1971. An introduction to Bayesian inference in econometrics. John Wiley & Sons (New York).

Zellner, A., Bauwens, L., and H.K. van Dijk, 1988. Bayesian Specification analysis and estimation of simultaneous equation models using Monte Carlo methods. J. of Econometrics 38, 39-72.

Zellner, A. and R.A. Highfield, 1988. Calculation of Maximum Entropy Distributions and Approximations of Marginal Posterior Distributions. Journal of Econometrics 37, 195-209.

Table 1: BBMR Performance under Normally Distributed Errors and Inflated Error Variance – Reduced Form Coefficients

	Reduced Form Coefficients (True Value)			
Distance Measures	π_1(65.06)	π_3(1.750)	π_5(1.189)	π_7(-0.2460)
$MSE^{1/2}$, Mean	1.162	0.0802	0.0136	0.0046
Bias, Mean	0.0296	-4.6E-4	5.4E-04	1.3E-04
$MSE^{1/2}$, Variance	82.03	0.3938	0.0114	0.0013
Bias, Variance	9.060	0.0417	0.0019	1.5E-04
Posterior Quantiles				
2%	2.00	2.03	2.04	2.03
(STDV)	(0.4652)	(0.4747)	(0.5768)	(0.4813)
5%	4.99	5.04	5.04	5.06
(STDV)	(0.7373)	(0.7437)	(0.8572)	(0.7423)
10%	10.04	10.06	10.03	10.05
(STDV)	(0.9859)	(1.010)	(1.067)	(0.9851)
90%	89.94	89.98	89.89	89.99
(STDV)	(0.9849)	(0.9886)	(1.046)	(0.9909)
95%	94.95	94.97	94.90	94.99
(STDV)	(0.7535)	(0.7600)	(0.8452)	(0.7254)
98%	97.95	97.98	97.93	98.01
(STDV)	(0.4774)	(0.4815)	(0.5912)	(0.4593)

NOTE: $n_b = 1000$, $n_{rep} = 1000$, MSE = Mean Square Error, STDV = Standard Deviation. The coefficients reported are from the reduced form consumption equation and represent the constant (π_1), and the coefficients of government wages(π_3), government non-wage expenditures (π_5), and lagged capital stock(π_7).

Table 2: BBMR Performance under Normally Distributed Errors and Inflated Error Variance – 2SLS-Mappings

	Structural Coefficients (True Value)			
Distance Measures	δ_1(16.44)	δ_2(0.1249)	δ_3(0.1631)	δ_4(0.7901)
$MSE^{1/2}$, Mean	0.0847	0.0046	0.0039	0.0025
Bias, Mean	-3.0E-04	-1.8E-05	-2.3E-05	-6.9E-06
$MSE^{1/2}$, Variance	0.5409	0.0018	0.0012	4.8E-04
Bias, Variance	0.0044	8.2E-06	1.4E-05	-1.6E-05
Posterior Quantiles				
2%	1.99	1.96	1.96	1.98
(STDV)	(0.5099)	(0.4598)	(0.4858)	(0.5026)
5%	5.01	4.98	4.95	4.97
(STDV)	(0.8001)	(0.7628)	(0.7816)	(0.7674)
10%	10.02	9.97	9.95	9.97
(STDV)	(1.104)	(1.055)	(1.078)	(1.021)
90%	89.90	89.97	89.89	89.92
(STDV)	(1.072)	(1.020)	(1.008)	(1.012)
95%	94.93	94.92	94.95	94.93
(STDV)	(0.7949)	(0.7596)	(0.7281)	(0.7395)
98%	97.93	97.94	97.96	97.95
(STDV)	(0.5157)	(0.4891)	(0.4805)	(0.4695)

NOTE: n_b = 1000, n_{rep} = 1000, MSE = Mean Square Error, STDV = Standard Deviation. The coefficients of government wages (W^G) and industry wages (W^I) are set equal in model estimation (δ_4) so that only four instead of five coefficients are reported here.

Table 3: BBMR Performance under Normally Distributed Errors and Original Error Variance – 2SLS-Mappings

	Structural Coefficients (True Value)			
Distance Measures	δ_1(16.44)	δ_2(0.1249)	δ_3(0.1631)	δ_4(0.7901)
$MSE^{1/2}$, Mean	0.0500	0.0029	0.0029	0.0012
Bias, Mean	6.4E-04	1.7E-05	2.1E-05	-3.2E-05
$MSE^{1/2}$, Variance	0.1370	7.8E-04	5.3E-04	7.0E-05
Bias, Variance	-0.0025	8.2E-06	-7.7E-06	-2.1E-06
Posterior Quantiles				
2%	1.96	1.94	1.96	1.96
(STDV)	(0.4398)	(0.4317)	(0.4340)	(0.4181)
5%	4.96	4.96	4.94	4.96
(STDV)	(0.6588)	(0.6687)	(0.6506)	(0.6268)
10%	9.96	9.97	9.96	10.01
(STDV)	(0.8865)	(0.9300)	(0.8761)	(0.8613)
90%	89.97	89.94	89.97	89.94
(STDV)	(0.8859)	(0.8318)	(0.8908)	(0.8821)
95%	94.95	94.93	94.96	94.95
(STDV)	(0.6394)	(0.6329)	(0.6534)	(0.6613)
98%	97.94	97.94	97.96	97.96
(STDV)	(0.4326)	(0.4386)	(0.4351)	(0.4260)

NOTE: See Table 2.

Table 4: BBMR Performance under Normally Distributed Errors and Inflated Error Variance – 2SLS-Mappings, Second Order Corrected Bootstrap Samples and Large Bootstrap Sample Size

	Structural Coefficients (True Value)			
Distance Measures	δ_1(16.44)	δ_2(0.1249)	δ_3(0.1631)	δ_4(0.7901)
$MSE^{1/2}$, Mean	0.0037	0.0024	0.0029	0.0014
Bias, Mean	5.9E-05	-5.7E-05	2.5E-04	-1.2E-05
$MSE^{1/2}$, Variance	0.0017	4.5E-04	0.0010	2.7E-04
Bias, Variance	2.6E-05	6.1E-06	4.6E-05	2.7E-06
Posterior Quantiles				
2%	1.94	1.99	1.96	1.98
(STDV)	(0.4294)	(0.4340)	(0.2539)	(0.3182)
5%	4.92	4.97	4.95	4.99
(STDV)	(0.6632)	(0.6564)	(0.4002)	(0.4780)
10%	9.96	9.96	9.94	9.97
(STDV)	(0.8705)	(0.8993)	(0.5661)	(0.6113)
90%	89.91	89.99	89.87	89.97
(STDV)	(0.9177)	(0.9085)	(0.6367)	(0.5373)
95%	94.92	94.97	94.90	94.96
(STDV)	(0.6955)	(0.6493)	(0.4631)	(0.4176)
98%	97.94	97.97	97.92	97.97
(STDV)	(0.4607)	(0.4408)	(0.2772)	(0.2781)

NOTE: n_b = 1000 for first two columns and n_b = 5000 for last two columns, n_{rep}= 1000, MSE = Mean Square Error, STDV = Standard Deviation. The coefficients of government wages (W^G) and industry wages (W^I) are set equal in model estimation (δ_4).

Table 5: BBMR Performance under T-Distributed Errors and Inflated Error Variance – 2SLS-Mappings

	Structural Coefficients (True Value)			
Distance Measures	δ_1(16.44)	δ_2(0.1249)	δ_3(0.1631)	δ_4(0.7901)
$MSE^{1/2}$, Mean	0.0966	0.0054	0.0042	0.0028
Bias, Mean	0.0085	-0.0013	6.3E-04	1.2E-04
$MSE^{1/2}$, Variance	0.6482	0.0018	0.0011	5.8E-04
Bias, Variance	0.0166	-1.2E-04	-1.0E-04	-3.4E-05
Posterior Quantiles				
2%	1.93	1.96	1.91	1.91
(STDV)	(0.5321)	(0.4869)	(0.4869)	(0.5249)
5%	4.93	4.99	4.88	4.88
(STDV)	(0.8430)	(0.7572)	(0.7511)	(0.7923)
10%	9.89	10.08	9.83	9.88
(STDV)	(1.175)	(1.062)	(1.050)	(1.073)
90%	89.84	90.21	89.87	89.95
(STDV)	(1.122)	(1.108)	(1.059)	(1.082)
95%	94.89	95.13	94.95	94.95
(STDV)	(0.8582)	(0.8361)	(0.7719)	(0.7860)
98%	97.93	98.05	97.97	97.96
(STDV)	(0.5666)	(0.5184)	(0.5068)	(0.4906)

NOTE: See Table 2.

Table 6: BBMR Performance under Normally Distributed Errors and Inflated Error Variance – 3SLS-Mappings

	Structural Coefficients (True Value)			
Distance Measures	δ_1(16.44)	δ_2(0.1249)	δ_3(0.1631)	δ_4(0.7901)
$MSE^{1/2}$, Mean	0.0884	0.0043	0.0037	0.0027
Bias, Mean	-0.0015	3.0E-04	-1.0E-04	-6.4E-05
$MSE^{1/2}$, Variance	0.4513	0.0016	0.0011	4.8E-04
Bias, Variance	-0.0334	5.9E-05	8.3E-05	1.5E-05
Posterior Quantiles				
2%	1.99	1.94	1.96	1.97
(STDV)	(0.5145)	(0.4702)	(0.4732)	(0.4969)
5%	5.00	4.94	4.99	4.96
(STDV)	(0.8218)	(0.7360)	(0.7319)	(0.7805)
10%	9.97	9.91	10.00	9.93
(STDV)	(1.111)	(1.037)	(1.023)	(1.040)
90%	89.92	89.85	89.97	89.93
(STDV)	(1.046)	(1.071)	(0.9912)	(1.024)
95%	94.91	94.91	94.93	95.96
(STDV)	(0.8022)	(0.7518)	(0.7439)	(0.7609)
98%	97.93	97.91	97.94	97.97
(STDV)	(0.5316)	(0.4800)	(0.4861)	(0.4965)

NOTE: See Table 2.