Simultaneous Equations Bayesian Bootstrap[1]

Thomas Heckelei and Ron C. Mittelhammer[2]

Washington State University

Selected Paper
American Agricultural Economics Association
Long Beach, CA 2002

Abstract
This paper introduces a semi-parametric bootstrapping approach to Bayesian analysis of structural parameters in simultaneous equation systems that extends the single and multivariate regression approaches of Heckelei and Mittelhammer (1996, 2002) to models with endogenous regressors. Monte Carlo evidence demonstrated the considerable accuracy of the procedure in approximating posterior distributions, even for small sample sizes.

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

Keywords: Bayesian Inference, bootstrapping, robust likelihood, simultaneous equations

---

[2] Thomas Heckelei is Assistant Professor, IMPACT Center and Department of Agricultural and Resource Economics and, and Ron Mittelhammer is a Professor in the Department of Agricultural and Resource Economics and in the Program in Statistics, Washington State University, USA. No senior authorship is assigned. tel: (509) 335-6653 or (509) 335-1706; e-mail: heckelei@wsu.edu, mittelha@wsu.edu .

1 Introduction

This paper presents a semi-parametric Bayesian approach for generating parameter estimates and conducting statistical inference within a system of simultaneous equations. The approach extends the recent Bayesian Bootstrap Multivariate Regression (BBMR, Heckelei and Mittelhammer 2002) methodology to account for endogenous regressors and over-identifying restrictions on structural parameters. The method is a completely computer-driven, simulation-based method for conducting Bayesian estimation and inference that fully avoids the oftentimes very difficult and even intractable derivations attendant to more complex Bayesian problems involving flexible combinations of prior distributions and likelihood functions. Moreover, the approach obviates the need for any specific functional specification of the likelihood function, thus eliminating the possibility of misspecification of the model in this regard and imparting a degree of model specification robustness to the analysis.

The "Simultaneous Equations Bayesian Bootstrap" (SEBB) replaces the usual explicit specification of a functional form for the likelihood function with a bootstrapped representation of the likelihood of the parameters. The representation is based on functional mappings from the error distribution to the model parameters. These mappings automatically incorporate the standard scale invariant ignorance prior on the covariance matrix of the errors. Extending the results of Zellner, Bauwens, and van Dijk (1988) and Heckelei and Mittelhammer (2002), the simulated posterior distributions incorporate any available exact and stochastic prior information identifying the structural parameters and can be used in the usual way to conduct posterior statistical analyses of the simultaneous equations model with Monte Carlo integration methods (e.g. Kloek and van Dijk 1978, Heckelei 1995 or Mittelhammer, Judge and Miller 2000 for a current textbook treatment). The approach allows for a very flexible choice of prior distributions and can be implemented as a generic computer-driven algorithm in standardized statistical software independently of the actual choice of prior distribution. It is distinguished from other approaches such as the "Bayesian Methods of Moments" (BMOM, Zellner 1996), in the way that a full representation of the posterior distributions is given together with the fact that no analytical derivations of posterior moments are necessary.

The paper is structured in the following way: First, the concept of a *Bayesian Data Information Mapping* (BDIM) is presented, which identifies semiparametric analogs to the mapping of error distributions to parameters that occurs in standard parametric Bayesian contexts. Then a brief review of the relation between reduced form and structural parameter distributions within the Bayesian paradigm is given. Third, the theory underlying the algorithm for obtaining posterior distributions of structural parameters using outcomes from an ignorance based posterior distribution of reduced form parameters is described. Fourth, the full computational algorithm is presented which allows for generating outcomes from the posterior distributions of structural parameters based on sample data. Finally, the functionality of the approach under a normal error distributions is illustrated with Monte Carlo simulation exercises based on Klein's Model I (Theil,1973).

2 Bayesian Data Information Mappings (BDIMs) For Linear Regression Models

In this section we explore semi-parametric analogues to standard parametric Bayesian mappings of data information to parameters in linear models, both of which lead to posterior density weightings on the parameters. We begin by considering single and multivariate regression settings in which the regressor matrices are orthogonal to model noise. The BDIM arguments presented here provide alternative motivation for the Bayesian Bootstrap computational algorithms presented in Heckelei and Mittelhammer (1996 and 2002). Moreover, the BDIM concept provides more fundamental motivation for posterior distribution simulations that does not begin with classical estimators of the parameters of the model as in Heckelei and Mittelhammer (1996 and 2002) but instead maps directly the probability distribution characteristics of the data sampling process into information on model parameters. Extensions of this BDIM subsequently form the basis for the semi-parametric Bayesian analysis of simultaneous equations.

Single Equation Model – Parametric Case

In order to identify what is meant by a *Bayesian Data Information Mapping* (BDIM) from a data sampling process to parameters in linear model contexts, begin with the parametric context and a single equation linear model,

(1) $$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \text{ where } \boldsymbol{\varepsilon} \sim \mathrm{N}\left(0, \sigma^2\mathbf{I}\right)$$

where we use, without loss of generality, the multivariate normal data sampling process as our benchmark parametric case. The semi-parametric case will be based on moment assumptions only, and will be developed ahead. Begin with the probability distribution of the error vector and consider moving to the likelihood function for the parameters, as is standard in Bayesian analyses of the linear model. Given the linear model structure (1) underlying the data sampling process, the probability distribution of the random vector $\boldsymbol{\varepsilon}$ can be thought as being transferred to the random vector $\mathbf{y} - \mathbf{X}\beta$, and the Jacobian of this type of transformation is always the identity matrix. Thus, in this transformation process, the argument $\boldsymbol{\varepsilon}$ in the distribution of the noise term is simply replaced by the new argument $\mathbf{y} - \mathbf{X}\beta$. Thus, functionally, we move from the PDF of the error vector,

(2) $$\mathrm{f}\left(\boldsymbol{\varepsilon} \mid \sigma\right) \propto \sigma^{-n} \exp\left(-\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{2\sigma^2}\right)$$

to the PDF of the $\mathbf{y}$ vector, as

(3) $$\mathrm{f}\left(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma\right) \propto \sigma^{-n} \exp\left(-\frac{\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)}{2\sigma^2}\right)$$

effectively by direct substitution of $\mathbf{y}-\mathbf{X}\boldsymbol{\beta}$ for $\boldsymbol{\varepsilon}$. This step in the process of defining the likelihood function is a dimension preserving transformation from R^{n} to R^{n}.

In making the final transition to the likelihood function, one engages in a subsequent dimension-reducing transformation whereby the function $f\left(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma\right)$ of the n arguments contained in $\mathbf{y}$ is changed to a function of the $k+1$ arguments $\boldsymbol{\beta}$ and σ, leading to the likelihood function

$$(4) \qquad \mathrm{L}\left(\boldsymbol{\beta}, \sigma \mid \mathbf{X}, \mathbf{y}\right) \propto \sigma^{-n} \exp\left(-\frac{\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)}{2\sigma^2}\right).$$

Thus, beginning with the probability distribution of the noise term, a sequence of functional transformations is implemented in which the dimensionality of the domain elements evolves as $\mathrm{R}^{\mathrm{n}} \rightarrow \mathrm{R}^{\mathrm{n}} \rightarrow \mathrm{R}^{\mathrm{k}+1}$, leading to the likelihood function of the parameters of the model.

Regarding the domain of the latter function and how it relates to the PDF of $\boldsymbol{\varepsilon}$ that characterized the noise term at the outset of the likelihood derivation process, first note that $\sigma \in \mathrm{R}_{+}$, and thus in the absence of prior information to the contrary, σ resides on the positive part of the real line but is otherwise unconstrained. The parameter σ can be interpreted as a scaling factor applied to the unit-variance linear model relationship

$$(5) \qquad \tilde{\mathbf{y}} = \mathbf{X}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\varepsilon}}, \quad \tilde{\boldsymbol{\varepsilon}} \sim \mathrm{N}\left(0, \mathbf{I}\right)$$

that forms the basis for an alternative characterization of the original linear model relationship as

$$(6) \qquad \mathbf{y} = \sigma\tilde{\mathbf{y}} = \mathbf{X}\left(\sigma\tilde{\boldsymbol{\beta}}\right) + \sigma\tilde{\boldsymbol{\varepsilon}} = \mathbf{X}\boldsymbol{\beta} + \sigma\tilde{\boldsymbol{\varepsilon}},$$

where $\tilde{\boldsymbol{\beta}} \equiv \sigma^{-1}\boldsymbol{\beta}$.

Regarding the $\boldsymbol{\beta}$ parameter vector, note that in the absence of prior information to the contrary, this parameter vector is unconstrained so that $\boldsymbol{\beta} \in \mathrm{R}^k$. Regarding its relationship to the PDF of the noise term, the value of $\boldsymbol{\beta}$ is clearly coincident with the value of $\boldsymbol{\varepsilon}$ that satisfies the relationship $\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, given $\mathbf{y}$ and $\mathbf{X}$. Then for any value of σ, one can think of the likelihood weighting on $(\boldsymbol{\beta},\sigma)$ to be coincident with the PDF weighting on the value of $\boldsymbol{\varepsilon}$, say $\boldsymbol{\varepsilon}(\boldsymbol{\beta})$, that corresponds to $\boldsymbol{\beta}$. That is,

(7) $$\mathrm{L}(\boldsymbol{\beta},\sigma \mid \mathbf{X},\mathbf{y}) \equiv \mathrm{f}\left(\boldsymbol{\varepsilon}(\boldsymbol{\beta}) \mid \sigma\right),$$

where we maintain the conditional density notation in this Bayesian context, but we emphasize that both the right and left sides of the identity in (7) can be interpreted as functions of both $\boldsymbol{\beta}$ and σ. Moreover, the joint posterior density of $(\boldsymbol{\beta},\sigma)$ can then be represented in the form

(8) $$\mathrm{p}(\boldsymbol{\beta},\sigma) = \mathrm{L}(\boldsymbol{\beta},\sigma \mid \mathbf{X},\mathbf{y})\,\mathrm{p}(\sigma) \equiv \mathrm{f}\left(\varepsilon(\boldsymbol{\beta}) \mid \sigma\right)\mathrm{p}(\sigma)$$

which relates the joint posterior density function to the PDF of the noise term, where we are considering the case where an improper prior is used to convey ignorance regarding the values of the unknown parameters of the model, as

(9) $$\mathrm{p}(\boldsymbol{\beta},\sigma) \propto \sigma^{-1}, \text{ where } \mathrm{p}(\sigma) \propto \sigma^{-1}.$$

Now note that the space of $\boldsymbol{\varepsilon}(\boldsymbol{\beta})$ values that are referenced through values of $\boldsymbol{\beta}$ and evaluated via $\mathrm{f}\left(\varepsilon(\boldsymbol{\beta}) \mid \sigma\right)$, lies in the subspace of R^n spanned by the *k* column vectors of $-\mathbf{X}$, translated by $\mathbf{y}$. In effect, the *relevant domain* of $\mathrm{f}(\boldsymbol{\varepsilon} \mid \sigma)$, in so far as the representation of the posterior density $\mathrm{p}(\boldsymbol{\beta},\sigma)$ is concerned, is restricted to linear functions of the columns of $\mathbf{X}$. A given value of $\boldsymbol{\varepsilon}$ is a member of this relevant domain *iff* it satisfies $\mathbf{X}\boldsymbol{\beta} = \mathbf{y} - \boldsymbol{\varepsilon}$, which in turn holds *iff*

(10) $\left(\mathbf{I}-\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)\left(\mathbf{y}-\boldsymbol{\varepsilon}\right)=\mathbf{0}$ or $\left(\mathbf{I}-\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)\boldsymbol{\varepsilon}=\hat{\boldsymbol{\varepsilon}}$,

where $\hat{\boldsymbol{\varepsilon}}$ denotes the least squares residual vector (Graybill, 1983, p.113). The solution space is a nonempty set of $(\mathrm{n}\times 1)$ vectors residing in a *k* dimensional subset of R^{n} that can be characterized by

(11) $\boldsymbol{\varepsilon}\in\Upsilon=\left\{\mathbf{e}:\ \mathbf{e}=\hat{\boldsymbol{\varepsilon}}+\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{h},\ \mathbf{h}\in\mathrm{R}^{\mathrm{n}}\right\}$

where **h** is arbitrary, and we have used the fact that $\mathbf{X}'\hat{\boldsymbol{\varepsilon}}=\mathbf{0}$ (Graybill, 1983, p. 114).

The posterior density weightings assigned to the pair $(\boldsymbol{\beta},\sigma)$ by (8) as $\boldsymbol{\beta}$ varies in R^{k}, for any given value of σ, is completely determined by the density of the noise term on the domain Υ, i.e., by a properly *normalized* version of $\mathrm{f}\left(\boldsymbol{\varepsilon}\mid\sigma\right)$ for $\boldsymbol{\varepsilon}\in\Upsilon$,

(12) $\boldsymbol{\varepsilon}_*\sim\mathrm{f}_*\left(\boldsymbol{\varepsilon}_*\mid\sigma\right)\propto\mathrm{f}\left(\boldsymbol{\varepsilon}_*\mid\sigma\right),\ \ \boldsymbol{\varepsilon}_*\in\Upsilon$.

Pursuing the relationship between the posterior (8) and the domain-restricted noise density (12) further, if $\boldsymbol{\varepsilon}_*\in\Upsilon$, then and only then (or else the equation system is inconsistent) can $\mathbf{X}\boldsymbol{\beta}=\mathbf{y}-\boldsymbol{\varepsilon}_*$ be solved uniquely for $\boldsymbol{\beta}$ via a straightforward application of the generalized inverse of $\mathbf{X}$, $\mathbf{X}^-=\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$, to yield

(13) $\mathbf{X}\boldsymbol{\beta}=\mathbf{y}-\boldsymbol{\varepsilon}_*\ \Leftrightarrow\ \boldsymbol{\beta}=\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{y}-\boldsymbol{\varepsilon}_*\right)=\hat{\boldsymbol{\beta}}-\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}_*$,

which depicts a functional mapping relationship between $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}_*$, together with the stochastic characteristics that it implies, where $\hat{\boldsymbol{\beta}}$ is the LS estimate of $\boldsymbol{\beta}$. In effect, the restriction of the noise density support, R^{n}, to the support space Υ, and the associated normalized density weighting on this support provided by (12), characterizes the contribution of the data $(\mathbf{y},\mathbf{X})$ and linear model structure $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$ to the mapping of information from the noise density to the likelihood function for the parameters. This mapping from the noise of the data sampling process, through the linear model structure and the observed data, to the likelihood function defines the BDIM in the current model context.

Single Equation Model – Semi-parametric Case

It will be useful, for purposes of constructing the semi-parametric analogue to the BDIM, to first characterize how random sampling could proceed from the restricted domain noise distribution in (12). Revisit the feasible space constraint $\left(\mathbf{I}-\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)\boldsymbol{\varepsilon}_* = \hat{\boldsymbol{\varepsilon}}$ displayed in (10) and note that we can rewrite this constraint in terms of eigenvalues and eigenvectors as

$$(14) \qquad \mathbf{P\Lambda P}'\boldsymbol{\varepsilon}_* = \hat{\boldsymbol{\varepsilon}},$$

where $\mathbf{P}$ and $\boldsymbol{\Lambda}$ are, respectively, the eigenvectors (column-wise) and the associated diagonal eigenvalue matrix for the symmetric idempotent matrix $\left(\mathbf{I}-\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)$ of rank $\mathrm{n}-\mathrm{k}$, so that $\mathbf{P\Lambda P}' = \left(\mathbf{I}-\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)$. Note the eigenvalue matrix has $\mathrm{n}-\mathrm{k}$ 1's and k 0's on its diagonal. Assuming that the columns of $\mathbf{P}$ and the associated diagonal elements of Λ are ordered so that the 1's are displayed first we obtain

$$(15) \qquad \boldsymbol{\Lambda} = \begin{bmatrix} \mathbf{I}_{\text{n-k}} & 0 \\ 0 & 0 \end{bmatrix}$$

It also follows from the orthogonality of the columns of $\mathbf{P}$ that

$$(16) \qquad \boldsymbol{\Lambda}\mathbf{P}'\boldsymbol{\varepsilon}_* = \mathbf{P}'\hat{\boldsymbol{\varepsilon}}$$

and because of the special structure of Λ exhibited in (15), only the first $\mathrm{n}-\mathrm{k}$ of these *n* constraints are linearly independent and binding constraints on the $\boldsymbol{\varepsilon}_*$ vector. The remaining k constraints are immaterial. Let $\mathbf{P}'_{(\text{n-k})}$ denote the first $\text{n}-\text{k}$ rows of $\mathbf{P}'$. Then the effective set of constraints on the feasible space of noise elements is represented by

(17) $$\mathbf{P}'_{(\text{n-k})}\boldsymbol{\varepsilon}_* = \mathbf{P}'_{(\text{n-k})}\hat{\boldsymbol{\varepsilon}},$$

which is a set of $\text{n}-\text{k}$ linearly independent constraints on $\boldsymbol{\varepsilon}_*$, given the value of $\hat{\boldsymbol{\varepsilon}}$ determined by the data $(\mathbf{y},\mathbf{X})$.

We can complete a basis for the span of noise vectors that have the original noise distribution, and that in addition respect the constraint (17), by first appending an appropriate *k*-dimensional random vector to (17) in the space orthogonal to the constraint space. In particular, note that $\mathbf{P}'_{(\text{n-k})}\mathbf{X}=\mathbf{0}$, which follows immediately from the fact that $\mathbf{P}_{(\text{n-k})}\mathbf{P}'_{(\text{n-k})} = \left(\mathbf{I}-\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)$ given (15). Then we can define the following full-rank transformation of $\boldsymbol{\varepsilon}_*$ that partitions the noise vector into the degenerate and non-degenerate subspaces, as

(18) $$\begin{bmatrix}\mathbf{P}'_{(\text{n-k})}\\ \mathbf{X}'\end{bmatrix}\boldsymbol{\varepsilon}_* = \begin{bmatrix}\mathbf{P}'_{(\text{n-k})}\hat{\boldsymbol{\varepsilon}}\\ \boldsymbol{\eta}\end{bmatrix},$$

where $\boldsymbol{\eta}=\mathbf{X}'\boldsymbol{\varepsilon}_* \sim \mathbf{X}'\boldsymbol{\varepsilon} \sim \left(\mathbf{0},\sigma^2\left(\mathbf{X}'\mathbf{X}\right)\right)$, and $\boldsymbol{\eta}\sim \text{N}\left(\mathbf{0},\sigma^2\left(\mathbf{X}'\mathbf{X}\right)\right)$ if the noise distribution is multivariate normally distributed, as in section 2.2.1. Solving (18) for the noise vector yields

(19) $$\boldsymbol{\varepsilon} = \begin{bmatrix}\mathbf{P}'_{(\text{n-k})}\\ \mathbf{X}'\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{P}'_{(\text{n-k})}\hat{\boldsymbol{\varepsilon}}\\ \boldsymbol{\eta}\end{bmatrix} = \mathbf{c}+\mathbf{D}\boldsymbol{\eta}$$

where $\mathbf{c}$ is an appropriate $(n \times 1)$ vector defined by post-multiplying the first *n-k* columns of the inverse matrix $\begin{bmatrix} \mathbf{P}'_{(n\text{-}k)} \\ \mathbf{X}' \end{bmatrix}^{-1}$ by $\mathbf{P}'_{(n\text{-}k)}\hat{\boldsymbol{\varepsilon}}$, and $\mathbf{D}$ denotes the last k columns of the inverse matrix $\begin{bmatrix} \mathbf{P}'_{(n\text{-}k)} \\ \mathbf{X}' \end{bmatrix}^{-1}$.

The preceding representation of the noise vector given by (19) depicts an operational sampling mechanism for generating outcomes of $\boldsymbol{\varepsilon}_*$ from the degenerate probability distribution depicted in (12). Specifically, one could draw $\boldsymbol{\varepsilon}$'s iid from the distribution of the noise vector, form the vector $\boldsymbol{\eta} = \mathbf{X}'\boldsymbol{\varepsilon}$, and then calculate an outcome of $\boldsymbol{\varepsilon}_*$ by an application of relationship (19). As such, the sampling mechanism is conditional on the value of the noise term standard deviation σ.

Having conceptualized a method for sampling from the degenerate distribution (12), we now note that one could sample the $\boldsymbol{\varepsilon}$'s and insert them directly into (13) while still maintaining the appropriate stochastic characteristics implied by (19). To see this, let $\boldsymbol{\eta} = \mathbf{X}'\boldsymbol{\varepsilon}$ and pre-multiply (19) by $\mathbf{X}'$ to obtain

(20)
$$\mathbf{X}'\begin{bmatrix} \mathbf{P}'_{(n\text{-}k)} \\ \mathbf{X}' \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{P}'_{(n\text{-}k)}\hat{\boldsymbol{\varepsilon}} \\ \mathbf{X}'\boldsymbol{\varepsilon} \end{bmatrix} = \mathbf{X}'\begin{bmatrix} \mathbf{P}'_{(n\text{-}k)} \\ \mathbf{X}' \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{P}'_{(n\text{-}k)}\left(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})\mathbf{X}'\right) \\ \mathbf{X}' \end{bmatrix}\boldsymbol{\varepsilon}$$
$$= \mathbf{X}'\begin{bmatrix} \mathbf{P}'_{(n\text{-}k)} \\ \mathbf{X}' \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{P}'_{(n\text{-}k)} \\ \mathbf{X}' \end{bmatrix}\boldsymbol{\varepsilon} = \mathbf{X}'\boldsymbol{\varepsilon}$$

because $\mathbf{P}'_{(n\text{-}k)}\mathbf{X}(\mathbf{X}'\mathbf{X})\mathbf{X}' = \mathbf{0}$. Thus, (13) can be used directly to implement the BDIM for the conditional posterior distribution of the $\boldsymbol{\beta}$ parameter.

In order to be able to sample from the unconditional (on σ) posterior distribution for $\boldsymbol{\beta}$, the sampling outcomes in (19), or equivalently the probability distribution in (12), must be mixed over the marginal posterior distribution of the σ parameter. Unfortunately, while the linear model structure in (1) provides a direct mapping from the noise vector to the $\boldsymbol{\beta}$, there is no direct functional counterpart available for mapping the noise distribution to the scale parameter σ without conditioning on $\boldsymbol{\beta}$. In effect, for such an unconditional mapping to exist, an additional relationship connecting noise, data, and parameters needs to be obtained. In the parametric case, this is provided by the likelihood function. In the semi-parametric case, one must look elsewhere for linking the data to the noise variance.

To establish such a mapping, we now assume that the noise distribution is from a group family in which σ acts strictly as a *scale* parameter, i.e. does not shift the mean of the error distribution. This group includes a large collection of distributions found in empirical applications, including the normal, double exponential, Cauchy, logistic, and mean-zero exponential, gamma, and uniform distributions. Given this characteristic of the noise distribution, it follows that the transformed noise distribution

$$(21) \qquad \boldsymbol{\varepsilon}_o = \sigma^{-1}\boldsymbol{\varepsilon} \sim g(\mathbf{w})$$

is free of the parameter σ and continues to have mean zero. Now consider an alternative representation of the data generating process underlying outcomes of the linear model, given by

$$(22) \qquad \mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\varepsilon}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \sigma\left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)\boldsymbol{\varepsilon}_o,$$

which implies that

$$(23) \qquad \left[\left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)\boldsymbol{\varepsilon}_o\right]\sigma = \hat{\boldsymbol{\varepsilon}}.$$

Following an argument analogous to the one used to identify the conditional $\boldsymbol{\beta}$ BDIM in the previous section, the only solution for σ in (23) is given by

(24) $$\sigma = \left(\boldsymbol{\varepsilon}_o'\left(\mathbf{I}-\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)\boldsymbol{\varepsilon}_o\right)^{-1}\boldsymbol{\varepsilon}_o'\left(\mathbf{I}-\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)\hat{\boldsymbol{\varepsilon}} = \frac{\boldsymbol{\varepsilon}_o'\hat{\boldsymbol{\varepsilon}}}{\left(\boldsymbol{\varepsilon}_o'\left(\mathbf{I}-\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)\boldsymbol{\varepsilon}_o\right)}$$

which implies that

(25) $$\sigma^2 = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{\left(\boldsymbol{\varepsilon}_o'\left(\mathbf{I}-\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)\boldsymbol{\varepsilon}_o\right)}$$

based on the definition of $\boldsymbol{\varepsilon}_o$ in (21). Thus, outcomes from the BDIM can be generated in principle by fixing $\hat{\boldsymbol{\varepsilon}}$ at its observed value, sampling $\boldsymbol{\varepsilon}_o$ from its distribution in (21), and then calculating outcomes of σ^2 from (25).

To obtain outcomes from the unconditional BDIM for $\boldsymbol{\beta}$, the values of σ implied by (25) can be used to define the probability density function in (12). Then outcomes of $\boldsymbol{\varepsilon}_*$ can be generated from this density and used in (13) to generate an outcome of $\boldsymbol{\beta}$. This is exactly the algorithm derived in Heckelei and Mittelhammer (1996), where bootstrapped outcomes from the least squares residuals were subjected to this transformation. However, the algorithm in the previous work was motivated based on the sampling distributions of the least squares estimators, whereas here a more general motivation is used based on direct functional mappings from the error distribution to the parameters. It should be noted that the presented mappings automatically incorporate the standard ignorance prior σ^{-1}. For a proof refer to Heckelei and Mittelhammer, 1996.

Multivariate Regression Extensions

Heckelei and Mittelhammer 2002 extend the single equation case to the multivariate regression case using arguments based on sampling distributions of least squares estimators. A straightforward generalization of the preceding BDIM concept leads to the same computational algorithm. In order to conserve space we provide only a brief outline of the generalization here. We present additional computational details of this multivariate regression sampling procedure in the next section, where the procedure is embedded in the context of the semi-parametric analysis of the structural parameters of a simultaneous equation system.

Consider the system of regression functions given by

(26) $\mathbf{Y} = \mathbf{X}\Pi + \mathbf{V}$,

where $\mathbf{Y}$ is a (n×m) matrix of observations on m endogenous variables, $\mathbf{X}$ is a (n×k) matrix of observations on k exogenous variables, Π is a (k×m) matrix of regression coefficients, and $\mathbf{V}$ is a (n×m) matrix representing n iid outcomes of a 1×m disturbance vector having some joint density function $g(\mathbf{V}|\mathbf{0},\Sigma)$ with mean vector $\mathbf{0}$ and covariance matrix Σ, the latter acting as a scale parameter matrix for the disturbance vector distribution. A slightly modified version of (26), which is a multiple equation analogue to (6) , can be defined as

(27) $\mathbf{Y} = \mathbf{X}\,\Pi + \mathbf{UT}$,

where the rows of the (n×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×m) matrix T is a matrix for which $\Sigma = \mathbf{T'T}$, so that $\mathbf{U[i,.]}\mathrm{T} = \mathbf{V}_i = \mathbf{V[i,.]} \sim g(\mathbf{V}_i \mid \mathbf{0},\Sigma)\ \forall i$.

The multivariate analogue to the β BDIM in the single equation context of (13) is then given by

(28) $\mathbf{X}\Pi = \mathbf{Y} - \mathbf{UT} \;\Leftrightarrow\; \Pi = \hat{\Pi} - (\mathbf{X'X})^{-1}\mathbf{X'UT}$

while the analog to the Σ BDIM in (25) is given by

(29) $\Sigma = \mathbf{S}^{1/2}(\mathbf{U'MU})^{-1}\mathbf{S}^{1/2}$

where $\hat{\Pi} = \left(\mathbf{X'X}\right)^{-1}\mathbf{X'Y}$, $\mathbf{S} = \hat{\mathbf{V}}'\hat{\mathbf{V}}$, and $\hat{\mathbf{V}} = \mathbf{Y} - \mathbf{X}\hat{\Pi}$. Heckelei and Mittelhammer (2002) prove that the standard ignorance prior on the covariance matrix, $p(\Sigma) \propto |\Sigma|^{-(m+1)/2}$, as well as the standard constant ignorance prior on the Π vector, are automatically incorporated into the BDIM mappings in (28) and (29). Methods of simulating outcomes from these BDIM mappings are presented in the next section.

3 Single Equation Bayesian Analysis of Structural Parameters

Notation and the parametric results of this section rely heavily on Zellner, Bauwens, and van Dijk (1988) and Heckelei and Mittelhammer (2002), where additional details can be found. Let the unrestricted reduced form of a simultaneous equation system be represented as

$$(30) \qquad \left[\mathbf{y}_1\ \mathbf{Y}_1\ \mathbf{Y}_0\right] = \mathbf{X}\left[\boldsymbol{\pi}_1\ \boldsymbol{\Pi}_1\ \boldsymbol{\Pi}_0\right] + \left[\mathbf{v}_1\ \mathbf{V}_1\ \mathbf{V}_0\right]$$

where $\mathbf{Y} = \left[\mathbf{y}_1\ \mathbf{Y}_1\ \mathbf{Y}_0\right]$ is an n×m matrix of observations on m endogenous variables with $\mathbf{y}_1$, $\mathbf{Y}_1$, and $\mathbf{Y}_0$ being, respectively, the 'first' endogenous variable, the endogenous variables included in the first structural equation, and the endogenous variables excluded from the first structural equation. The matrix $\mathbf{X}$ is an n×k matrix of n observations on k predetermined variables and $\Pi=\left[\boldsymbol{\pi}_1\ \boldsymbol{\Pi}_1\ \boldsymbol{\Pi}_0\right]$ represents reduced form parameters corresponding to $\mathbf{y}_1$, $\mathbf{Y}_1$, and $\mathbf{Y}_{0,}$ respectively. The rows of the n×m disturbance matrix $\mathbf{V} = \left[\mathbf{v}_1\ \mathbf{V}_1\ \mathbf{V}_0\right]$ represent n iid random outcomes of a 1×m disturbance vector with joint density function $\mathbf{V}_i \sim g(\mathbf{V}_i|\mathbf{0},\Sigma)$ having mean vector $\mathbf{0}$ and covariance matrix Σ.

The first structural equation can be written as

$$(31) \qquad \left[\mathbf{y}_1\ \mathbf{Y}_1\ \mathbf{Y}_0\right]\begin{bmatrix} 1 \\ -\gamma_1 \\ \mathbf{0} \end{bmatrix} = \left[\mathbf{X}_1\ \mathbf{X}_0\right]\begin{bmatrix} \boldsymbol{\beta}_1 \\ \mathbf{0} \end{bmatrix} + \mathbf{u}_1\,,$$

where the $(m_1\text{-}1)\times 1$ vector γ_1 and the $k_1\times 1$ vector $\boldsymbol{\beta}_1$ are the structural parameters, and $\mathbf{u}_1$ is an n×1 vector of structural disturbance terms. Substituting $\mathbf{X}\boldsymbol{\Pi}_1 + \mathbf{V}_1$ for $\mathbf{Y}_1$ based on (30) and rearranging (31) yields

(32) $$\mathbf{y}_1 = \mathbf{X\Pi}_1\boldsymbol{\gamma}_1 + [\mathbf{X}_1\ \mathbf{X}_0]\begin{bmatrix}\boldsymbol{\beta}_1\\ \mathbf{0}\end{bmatrix} + \mathbf{u}_1 + \mathbf{V}_1\boldsymbol{\gamma}_1 = \mathbf{X}\left[\mathbf{\Pi}_1\boldsymbol{\gamma}_1 + \begin{bmatrix}\boldsymbol{\beta}_1\\ \mathbf{0}\end{bmatrix}\right] + \mathbf{u}_1 + \mathbf{V}_1\boldsymbol{\gamma}_1$$

Compatibility with (30) implies the parameter and disturbance restrictions

(33) $$\boldsymbol{\pi}_1 = \mathbf{\Pi}_1\boldsymbol{\gamma}_1 + \begin{bmatrix}\boldsymbol{\beta}_1\\ \mathbf{0}\end{bmatrix} \text{ and } \mathbf{v}_1 = \mathbf{u}_1 + \mathbf{V}_1\boldsymbol{\gamma}_1$$

which makes it clear that (32) is a representation of the reduced form for $\mathbf{y}_1$ expressed in terms of both the parameters of the first structural equation and the reduced form parameters in $\mathbf{\Pi}_1$. Eliminating parameters that are restricted to zero in (32) obtains the following representation of the reduced form equations for $\mathbf{y}_1$ and $\mathbf{Y}_1$ (see also equations 2.13a and b in Zellner, Bauwens, and van Dijk,1988):

(34) $$\mathbf{y}_1 = \mathbf{X\Pi}_1\boldsymbol{\gamma}_1 + \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{v}_1$$

(35) $$\mathbf{Y}_1 = \mathbf{X\Pi}_1 + \mathbf{V}_1$$

We call this system the *unrestricted error* (UE) representation of the system. It can be shown that applying a BDIM to (34) and (35), which maps from the unrestricted reduced form error distributions to the parameters given the data, results in the “two-stage least squares (2SLS) mapping” presented in Zellner, Bauwens, and van Dijk (1988) in their equation (2.36), and provides alternative motivation for the Bayesian bootstrap 2SLS mapping of Heckelei and Mittelhammer (2002). The 2SLS mapping does not assume that the identifying restrictions of the model must hold, and can thus be used to test hypotheses about the validity of these restrictions.

Utilizing the restrictions among error terms identified in (33), a representation of the first structural equation in terms of reduced form errors, together with the reduced form equations for the other endogenous variables appearing in the structural equation, is given by

(36) $$\mathbf{y}_1 = \mathbf{Y}_1\boldsymbol{\gamma}_1 + \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{v}_1 - \mathbf{V}_1\boldsymbol{\gamma}_1$$

(37) $$\mathbf{Y}_1 = \mathbf{X}\boldsymbol{\Pi}_1 + \mathbf{V}_1 .$$

We call this system the *restricted error* (RE) representation of the system. In this representation of the system of equations, the identifying restrictions on the parameters of the first structural equation are clearly in force. From the RE we first consider, as a benchmark reference point, the analytical posterior distribution of the structural parameters assuming a multivariate normal distribution for the reduced form errors in the system. We then identify a transformation of the system to which a BDIM can be applied in order to simulate the posterior distribution of the parameters of the system in a semi-parametric context.

Analytical Posterior

The joint posterior distribution, $h(\gamma_1,\boldsymbol{\beta}_1,\boldsymbol{\Pi}_1|\mathbf{y}_1,\mathbf{Y}_1)$, of the structural parameters and reduced form parameters in the RE system can be defined by first noting the fact that the joint posterior distribution of the coefficients in (34) and (35) can be represented as the posterior distribution of $(\gamma_1, \boldsymbol{\beta}_1)$ given Π_1, $h^1(\gamma_1,\boldsymbol{\beta}_1| \Pi_1,\mathbf{y}_1,\mathbf{Y}_1)$, times the marginal posterior distribution of Π_1, $h^2\left(\boldsymbol{\Pi}_1 \mid \mathbf{Y}_1\right)$, that is

(38) $$h\left(\boldsymbol{\gamma}_1,\boldsymbol{\beta}_1,\boldsymbol{\Pi}_1 \mid \mathbf{y}_1,\mathbf{Y}_1\right) = h^1\left(\boldsymbol{\gamma}_1,\boldsymbol{\beta}_1 \mid \boldsymbol{\Pi}_1,\mathbf{y}_1,\mathbf{Y}_1\right)h^2\left(\boldsymbol{\Pi}_1 \mid \mathbf{Y}_1\right).$$

If the error density $g(\mathbf{V}_1|\mathbf{0},\boldsymbol{\Sigma}_1)$ is of the multivariate normal type by virtue of the rows of $\mathbf{V}_1$ being independent and identically distributed multivariate normal random vectors, and if the standard ignorance prior $p(\Pi_1,\Sigma) = p(\Pi_1)p(\Sigma) \propto \left|\Sigma_1\right|^{-(m_1+1)/2}$ is employed, the marginal posterior distribution of the reduced form parameters Π_1, $h^2(\Pi_1|\mathbf{Y}_1)$, is a matrix student-t distribution denoted by $T(n\text{-}k,\hat{\boldsymbol{\Pi}}_1,\mathbf{S}_1)$ and defined by

(39) $$h^2\left(\boldsymbol{\Pi}_1 \mid \mathbf{Y}_1\right) = \left|\mathbf{S}_1 + \left(\boldsymbol{\Pi}_1 - \hat{\boldsymbol{\Pi}}_1\right)\mathbf{X}'\mathbf{X}\left(\boldsymbol{\Pi}_1 - \hat{\boldsymbol{\Pi}}_1\right)\right|^{-n/2}$$

where $\mathbf{S}_1 = \left(\mathbf{Y}_1 - \mathbf{X}\hat{\mathbf{\Pi}}_1\right)'\left(\mathbf{Y}_1 - \mathbf{X}\hat{\mathbf{\Pi}}_1\right) = \hat{\mathbf{V}}_1'\hat{\mathbf{V}}_1$ and $\hat{\mathbf{\Pi}}_1 = \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{Y}_1$. Analogously, if the reduced form errors associated with the first endogenous variable, $\mathbf{v}_1$, follow a normal density $g(\mathbf{v}_1|0,\sigma_1)$ and the ignorance prior $p(\gamma_1,\beta_1,\sigma_1) = p(\gamma_1,\beta_1)\ p(\sigma_1) = \sigma^{-1}$ is employed, then the conditional marginal posterior distribution of the parameters of the first structural equation, $h^1(\gamma_1,\beta_1|\ \Pi_1,\mathbf{y}_1,\mathbf{Y}_1)$, is a multivariate student-t distribution denoted by $t(n\text{-}k_1\text{-}m_1,\ \hat{\boldsymbol{\gamma}}_1,\hat{\boldsymbol{\beta}}_1,\beta_1\ ,s_1)$ and defined by

$$\text{(40)} \qquad h^1\left(\boldsymbol{\gamma}_1,\boldsymbol{\beta}_1 \mid \mathbf{\Pi}_1,\mathbf{y}_1,\mathbf{Y}_1\right) = \left| s_1^2 + \left(\begin{bmatrix}\boldsymbol{\gamma}_1\\ \boldsymbol{\beta}_1\end{bmatrix} - \begin{bmatrix}\hat{\boldsymbol{\gamma}}_1\\ \hat{\boldsymbol{\beta}}_1\end{bmatrix}\right)\left[\mathbf{Y}_1 \quad \mathbf{X}_1\right]'\mathbf{M}_{\mathbf{V}_1}\left[\mathbf{Y}_1 \quad \mathbf{X}_1\right]\left(\begin{bmatrix}\boldsymbol{\gamma}_1\\ \boldsymbol{\beta}_1\end{bmatrix} - \begin{bmatrix}\hat{\boldsymbol{\gamma}}_1\\ \hat{\boldsymbol{\beta}}_1\end{bmatrix}\right)\right|^{-n/2}$$

where

$$\text{(41)} \qquad \begin{bmatrix}\hat{\boldsymbol{\gamma}}_1\\ \hat{\boldsymbol{\beta}}_1\end{bmatrix} = \left(\left[\mathbf{Y}_1 \quad \mathbf{X}_1\right]'\mathbf{M}_{\mathbf{V}_1}\left[\mathbf{Y}_1 \quad \mathbf{X}_1\right]\right)^{-1}\left[\mathbf{Y}_1 \quad \mathbf{X}_1\right]'\mathbf{M}_{\mathbf{V}_1}\mathbf{y}_1,$$

$$s_1^2 = \left(\mathbf{y}_1 - \left[\mathbf{Y}_1 \quad \mathbf{X}_1\right]\begin{bmatrix}\boldsymbol{\gamma}_1\\ \boldsymbol{\beta}_1\end{bmatrix}\right)'\mathbf{M}_{\mathbf{V}_1}\left(\mathbf{y}_1 - \left[\mathbf{Y}_1 \quad \mathbf{X}_1\right]\begin{bmatrix}\boldsymbol{\gamma}_1\\ \boldsymbol{\beta}_1\end{bmatrix}\right) = \mathbf{v}_1'\mathbf{M}_{\mathbf{V}_1}\mathbf{v}_1, \text{ and } \mathbf{M}_{\mathbf{V}_1} = \mathbf{I} - \mathbf{V}_1\left(\mathbf{V}_1'\mathbf{V}_1\right)^{-1}\mathbf{V}_1'.$$

Consequently, the joint posterior of reduced form and structural parameters (38) can be written as

$$\text{(42)} \qquad h\left(\boldsymbol{\gamma}_1,\boldsymbol{\beta}_1,\mathbf{\Pi}_1 \mid \mathbf{y}_1,\mathbf{Y}_1\right) = t\left(n - m1 - k1, \hat{\boldsymbol{\gamma}}_1, \hat{\boldsymbol{\beta}}_1, s_1^2\right) T\left(n - k, \hat{\mathbf{\Pi}}_1, \mathbf{S}_1\right).$$

Monte Carlo integration based on outcomes of (42) or, in the case of informative priors (γ_1, β_1), based on prior-weighted outcomes of (42), provides a flexible approach for evaluating posterior expectations of general functions of the parameters of the model.

BDIM Mapping of the Projected RE System

The semi-parametric sampling approach we present now aims to generate outcomes of the joint posterior $h(\gamma_1,\beta_1,\Pi_1|\mathbf{y}_1,\mathbf{Y}_1)$ in the absence of parametric distributional assumptions. We contemplate the use of a BDIM applied to the RE system. In so doing, it is apparent from the outset that the error vector in (36), $(\mathbf{v}_1-\mathbf{V}_1\boldsymbol{\gamma}_1)$, depends explicitly on the structural parameter vector, $\boldsymbol{\gamma}_1$, and because each value of $\boldsymbol{\gamma}_1$ has its own associated error distribution, a BDIM for a parameter vector that also includes the parameter vector $\boldsymbol{\gamma}_1$ is undefined. We now show that the information contained in (36) can be projected into two component systems, one defining the component of the error distribution that can be mapped to the parameters using a BDIM, and the other which cannot be mapped.

The structural equation in (36) can be projected in such a way that the error distribution associated with the projection is independent of the $\boldsymbol{\gamma}_1$, which is a necessary condition for invoking a BDIM from an error vector distribution to parameters of the system. It is apparent that the projector must annihilate the vector $\mathbf{V}_1\boldsymbol{\gamma}_1$ in (36), and the obvious choice for such a projector is the matrix $\mathbf{M}_{\mathbf{V}_1}=\mathbf{I}-\mathbf{V}_1\left(\mathbf{V}_1'\mathbf{V}_1\right)^{-1}\mathbf{V}_1'=\mathbf{I}-\mathbf{P}_{\mathbf{V}_1}$, which forms a basis for the vector space orthogonal to $\mathbf{V}_1\boldsymbol{\gamma}_1$ and is such that $\mathbf{M}_{\mathbf{V}_1}(\mathbf{v}_1-\mathbf{V}_1\boldsymbol{\gamma}_1)=\mathbf{M}_{\mathbf{V}_1}\mathbf{v}_1$. Defining the error vector in (36) as $(\mathbf{v}_1-\mathbf{V}_1\boldsymbol{\gamma}_1)=\mathbf{M}_{\mathbf{V}_1}\mathbf{v}_1+\mathbf{P}_{\mathbf{V}_1}(\mathbf{v}_1-\mathbf{V}_1\boldsymbol{\gamma}_1)$ then identifies a decomposition of the error vector into a random disturbance component whose distribution is independent of $\boldsymbol{\gamma}_1$, namely $\mathbf{M}_{\mathbf{V}_1}\mathbf{v}_1$, plus a component whose distribution is dependent on $\boldsymbol{\gamma}_1$, given by $\mathbf{P}_{\mathbf{V}_1}(\mathbf{v}_1-\mathbf{V}_1\boldsymbol{\gamma}_1)$.

The projected version of the RE system for use in constructing the BDIM is then

(43) $$\mathbf{M}_{\mathbf{V}_1}\mathbf{y}_1=\mathbf{M}_{\mathbf{V}_1}\mathbf{Y}_1\boldsymbol{\gamma}_1+\mathbf{M}_{\mathbf{V}_1}\mathbf{X}_1\boldsymbol{\beta}_1+\mathbf{M}_{\mathbf{V}_1}\mathbf{v}_1$$

(44) $$\mathbf{Y}_1=\mathbf{X}\boldsymbol{\Pi}_1+\mathbf{V}_1.$$

Applying the BDIM approach to map the joint distribution of the error vectors $\mathbf{M}_{\mathbf{V}_1}\mathbf{v}_1$ and $\mathbf{V}_1$ into the parameters of the projected RE, the structural parameters in the first equation are given by

(45) $$\begin{bmatrix}\boldsymbol{\gamma}_1\\ \boldsymbol{\beta}_1\end{bmatrix}=\begin{bmatrix}\hat{\boldsymbol{\gamma}}_1\\ \hat{\boldsymbol{\beta}}_1\end{bmatrix}-\left(\left[\mathbf{Y}_1 \quad \mathbf{X}_1\right]'\mathbf{M}_{\mathbf{V}_1}\left[\mathbf{Y}_1 \quad \mathbf{X}_1\right]\right)^{-1}\left[\mathbf{Y}_1 \quad \mathbf{X}_1\right]'\mathbf{M}_{\mathbf{V}_1}\mathbf{v}_1$$

where $\begin{bmatrix}\hat{\boldsymbol{\gamma}}_1\\ \hat{\boldsymbol{\beta}}_1\end{bmatrix}$ is defined as it was in (41). Thus, the BDIM mapping implies precisely the same posterior distribution for the parameters of the structural equation as in the analytical derivation when the distribution of the reduced form errors is multivariate normal, but the BDIM mapping also applies more generally and does not rely on normality for its validity. The Bootstrap implementation of sampling from the BDIM induced posterior distribution for the structural parameters is defined ahead.

Bayesian Bootstrap Posterior

The computational algorithm for performing semi-parametric Bayesian posterior inference for the structural parameters based on the BDIM is as follows:

1. Draw a bootstrap sample $\tilde{\mathbf{V}}_1^*$ from $\hat{\mathbf{V}}_1$

2. Apply a transformation to calculate $\tilde{\mathbf{V}}_1^{**}=\mathbf{S}_1^{-1/2}(\mathbf{S}_1\mathbf{S}_1^{*-1}\mathbf{S}_1)^{1/2}\tilde{\mathbf{V}}_1^*$ where $\mathbf{S}_1^*=\tilde{\mathbf{V}}_1^*{}'\mathbf{M}_{\mathbf{X}}\tilde{\mathbf{V}}_1^*$ and $\mathbf{M}_{\mathbf{X}}=\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.[3]

3. Calculate an outcome from the marginal posterior of Π_1 as $\boldsymbol{\Pi}_1^*=\hat{\boldsymbol{\Pi}}-(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{V}}_1^{**}$

---

[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 $\mathbf{Q}=\mathbf{Q}^{1/2}{}'\mathbf{Q}^{1/2}$ for non-symmetric matrix square roots.

4. Transform the variables of the structural equation by pre-multiplication with

$\mathbf{M}^*_{\mathbf{V}_1} = \mathbf{I} - \mathbf{V}^*_1(\mathbf{V}^{*\prime}_1\mathbf{V}^*_1)^{-1}\mathbf{V}^{*\prime}_1$ where $\mathbf{V}^*_1 = \mathbf{Y}_1 - \mathbf{X}\boldsymbol{\Pi}^*_1$ to define $\mathbf{y}^*_1 = \mathbf{M}^*_{\mathbf{V}_1}\mathbf{y}_1$ and

$\mathbf{Z}^*_1 = \begin{bmatrix}\mathbf{Y}^*_1 & \mathbf{X}^*_1\end{bmatrix} = \mathbf{M}^*_{\mathbf{V}_1}\begin{bmatrix}\mathbf{Y}_1 & \mathbf{X}_1\end{bmatrix}$.

5. Calculate $\hat{\boldsymbol{\delta}}^*_1 = \begin{bmatrix}\hat{\boldsymbol{\gamma}}^*_1 \\ \hat{\boldsymbol{\beta}}^*_1\end{bmatrix} = \left(\mathbf{Z}^{*\prime}_1\mathbf{Z}^*_1\right)^{-1}\mathbf{Z}^{*\prime}_1\mathbf{y}^*_1$ and $\hat{\mathbf{v}}^*_1 = \mathbf{y}^*_1 - \mathbf{Z}^{*\prime}_1\hat{\boldsymbol{\delta}}^*_1$.

6. Draw a bootstrap sample $\tilde{\mathbf{v}}^*_1$ from $\hat{\mathbf{v}}^*_1$

7. Perform a transformation to obtain $\tilde{\mathbf{v}}^{**}_1 = \left(\dfrac{\hat{s}^2_1}{\tilde{s}^2_1}\right)^{1/2}\tilde{\mathbf{v}}^*_1$, where $s^{*2}_1 = \tilde{v}^{*\prime}_1\mathbf{M}_{\mathbf{Z}_1}\tilde{v}^*_1$, $\hat{s}^2_1 = \hat{\mathbf{v}}^{*\prime}_1\hat{\mathbf{v}}^*_1$, and

$\mathbf{M}^*_{\mathbf{Z}_1} = \mathbf{I} - \mathbf{Z}^*_1(\mathbf{Z}^{*\prime}_1\mathbf{Z}^*_1)^{-1}\mathbf{Z}^{*\prime}_1$.

8. Calculate an outcome from the marginal posterior of $\boldsymbol{\delta}_1 = \begin{bmatrix}\boldsymbol{\gamma}_1 \\ \boldsymbol{\beta}_1\end{bmatrix}$ as $\boldsymbol{\delta}^*_1 = \hat{\boldsymbol{\delta}}^*_1 - (\mathbf{Z}^{*\prime}_1\mathbf{Z}^*_1)^{-1}\mathbf{Z}^{*\prime}_1\tilde{\mathbf{v}}^{**}_1$

9. Repeat steps 1 to 9 n_b times to obtain n_b outcomes of the posterior distribution of $\boldsymbol{\delta}_1 = \begin{bmatrix}\boldsymbol{\gamma}_1 \\ \boldsymbol{\beta}_1\end{bmatrix}$ to use in performing posterior inference.

Steps 1 to 3 represent the computational algorithm for the multivariate extension to the single equation BDIM mapping presented above. It is fully equivalent to the Bayesian Bootstrap Multivariate Regression (BBMR) developed and motivated in detail by Heckelei and Mittelhammer (2002) and maps outcomes from a bootstrapped error density to the posterior outcomes of $\boldsymbol{\Pi}_1$. Steps 5 to 8 are exactly the single equation BDIM applied to the transformed structural equation that maps outcomes from a transformed bootstrapped error distribution to the posterior outcomes of the structural parameters (γ_1, β_1). Step 4 is the orthogonal projection procedure representing the Bootstrap analog to the equation (43) in the projected version of the RE system. We refer to the preceding algorithm as the *Limited Information Simultaneous Equations Bayesian Bootstrap* (LI-SEBB).

4 Some Monte Carlo Evidence

In order to evaluate how well the LI-SEBB reproduces posterior distributions of structural parameters we performed a Monte Carlo experiment comparing it with the parametric solution under normality. The simulations are based on Klein's Model I. Variable definitions and additional information about the model that is not reported here can be found 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 (1988) and Heckelei and Mittelhammer (2002) with results presented here.

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

Step 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 system). Specifically, the system takes the following form:

$$\text{(46)} \qquad \mathbf{Y}_t\boldsymbol{\Gamma} = \mathbf{X}_t\mathbf{B} + \mathbf{U}_t$$

where $\mathbf{Y}_t = \left\{ C_t \; I_t \; W_t^I \; X_t \; P_t \; D_t \right\}$, $\mathbf{X}_t = \left\{ 1 \quad t-1931 \quad W_t^G \quad T_t \quad G_t \quad P_{t-1} \quad K_{t-1} \quad X_{t-1} \right\}$,

$$\boldsymbol{\Gamma} = \begin{Bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 \\ -0.7901 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & -0.4005 & 1 & -1 & 0 \\ -0.1249 & 0.0131 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{Bmatrix}, \quad \mathbf{B} = \begin{Bmatrix} 16.44 & 28.18 & 1.8 & 0 & 0 & 0 \\ 0 & 0 & 0.1497 & 0 & 0 & 0 \\ 0.7901 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0.1631 & 0.7557 & 0 & 0 & 0 & 0 \\ 0 & 0.1948 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0.1813 & 0 & 0 & 0 \end{Bmatrix}.$$

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)

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

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 lagged endogenous variables in $\mathbf{X}_t$) using $\mathbf{Y}_t=\mathbf{X}_t\mathbf{B}\mathbf{\Gamma}^{-1}+\mathbf{U}_t\mathbf{\Gamma}^{-1}$. Having developed a limited information, i.e. single equation procedure above, we restrict the analysis to the structural parameters, δ_1^j for j=1,…,4, of the first model equation which is the consumption function.

Step 2: For each data sample k = 1,…,n_{rep}, n_{samp}=10000 outcomes are calculated from the marginal posterior distributions of the structural coefficients based on both the LI-SEBB procedure (${}_B\delta_1^{ji}$) and the parametric posterior under normality (${}_P\delta_1^{ji}$). The parametric sampling employs the posterior representation in (42) and mixes the multivariate t-distribution of δ_1 conditional on $\mathbf{\Pi}_1$ over the matrix T-marginal posterior of $\mathbf{\Pi}_1$. A method for generating matrix T-random numbers can be found in Zellner, Bauwens and van Dijk (1988).

Step 3: Several measures are calculated for each n_{samp} set of outcomes of the k-th data sample separately for the LI-SEBB and the parametric outcomes: Bootstrapped means and variances, as

$$ {}_{B}\bar{\delta}_{l}^{j} = \left(1/n_{samp}\right)\sum\nolimits_{i} {}_{B}\delta_{l}^{ji} \text{, and } \operatorname{Var}\left({}_{B}\delta_{l}^{j}\right) = \left(1/n_{samp}\right)\sum\nolimits_{i}\left({}_{B}\delta_{l}^{j} - {}_{B}\bar{\delta}_{l}^{j}\right)^{2}. $$

Their parametric counterparts are based on equivalent formulas applied to ${}_{P}\delta_{l}^{ji}$ outcomes. To better investigate the approximation of the entire distribution, values of δ_{l}^{j} corresponding to the 1st, 5th, 10th, 50th, 90th, 95th, and 99th percentile, denoted as ${}_{B}^{i^{th}}\delta_{l}^{j}$ and ${}_{P}^{i^{th}}\delta_{l}^{j}$, are computed by sorting the $n_{samp}\times 1$ vectors and choosing the element associated with the appropriate quantile of the sorted data.

Step 4: All measures calculated in step 3 are collected for all n_{rep} data samples to finally compute the following measures indicating the approximation accuracy of the LI-SEBB relative to the parametric posterior.

a. Bias of the posterior mean: $\text{BiasMean}^{j} = \left(1/n_{rep}\right)\sum\nolimits_{k}\left({}_{B}\bar{\delta}_{l}^{jk} - {}_{P}\bar{\delta}_{l}^{jk}\right)$

b. Relative Bias of the posterior variance:

$$ \text{BiasVar}^{j} = \left(1/n_{rep}\right)\sum\nolimits_{k}\left(\operatorname{Var}\left({}_{B}\delta_{l}^{jk}\right)\Big/\operatorname{Var}\left({}_{P}\delta_{l}^{jk}\right)\right) $$

c. Bias of posterior percentiles: $\text{BiasPerc}^{j} = \left(1/n_{rep}\right)\sum\nolimits_{k}\left({}_{B}^{i^{th}}\delta_{l}^{jk} - {}_{P}^{i^{th}}\delta_{l}^{jk}\right)$

We emphasize that the simulations are intended to evaluate the performance of LI-SEBB in representing various characteristics of the posterior distribution of the structural parameters. They do not evaluate the sampling properties of point estimators as those depend on the desired loss function and, in the context of empirical applications, on prior information. However, an accurate representation of the entire posterior distribution is certainly the key to an accurate representation of all possible point estimators based on the posterior.

Before proceeding to the simulation results it is important to note what type of approximation errors can be expected. First, the samples from the parametric as well as the LI-SEBB posterior distributions are limited to 10000, leaving a likely small but existent sampling noise which can be more pronounced for tail values of the distribution. Second, the empirical distribution function used by the bootstrap procedure will represent the true error distribution less than perfectly at the relatively small sample size of n=21. Experiments for a smaller number of data samples (n_{rep} = 100) but with a larger number of simulated samples from the posterior (n_{samp}=100,000) indicate that the EDF approximation error dominates the posterior distribution sampling noise with respect to estimated biases because the magnitude of the biases did not change significantly. Third, it is not currently known whether the mappings employed in the LI-SEBB algorithm will generate, *in general*, distributionally equivalent outcomes of the posterior distribution based on the true error distribution underlying the reduced form errors. Distributional equivalence will be obtained, however, in all cases where the rows of the reduced form matrix **V** are uncorrelated and the matrix of reduced form errors follows any elliptically contoured distribution (Ng, 2001). This includes the normal distribution as a special case, but includes a wide variety of other distributions 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).

Table 1 shows a very close approximation of the parametric marginal posteriors by the LI-SEBB approach. Generally, the accuracy for the intercept of the equation is slightly less than for the three slope coefficients, but the performance remains more than satisfactory. The posterior variance is slightly overestimated for all four coefficients, but again, the degree of inaccuracy is minimal. Although the values for the different percentiles are more accurately matched as one moves away from the tails of the distribution, all LI-SEBB results are, for all practical purposes, equivalent to the parametric values given that the reported variances indicate only very limited deviations of the bias measures from one sample to the other.

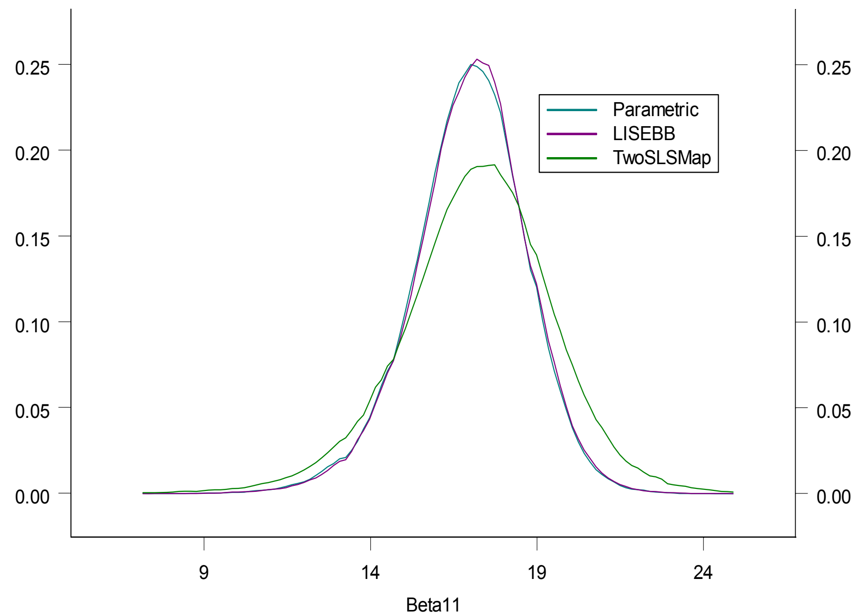
Table 1: LI-SEBB Performance under Normally Distributed Errors Relative to Parametric Posterior: Posterior Distribution Characteristics of Structural Coefficients under Ignorance Prior

Distance Measures	δ_1^1 (16.44)	δ_1^2 (0.1249)	δ_1^3 (0.7901)	δ_1^4 (0.1631)
	Structural Coefficients (True Value)			
BiasMean	2.40E-03	-3.77E-05	5.56E-05	-3.50E-05
	(1.44E-03)	(9.36E-07)	(2.58E-06)	(2.27E-06)
BiasVar	1.0057	1.0021	1.0012	1.0014
	(6.68E-03)	(7.51E-07)	(6.70E-06)	(7.25E-06)
BiasPerc				
1%	-8.22E-03	-4.25E-04	3.09E-04	1.36E-04
	(2.14E-02)	(4.36E-05)	(5.12E-05)	(8.60E-05)
5%	-6.63E-03	-1.98E-04	1.20E-04	1.37E-04
	(5.50E-03)	(7.21E-06)	(1.15E-05)	(1.57E-05)
10%	-3.53E-03	-9.08E-05	6.29E-05	8.01E-05
	(3.35E-03)	(3.48E-06)	(6.83E-06)	(8.54E-06)
50%	3.18E-03	4.12E-05	-1.24E-05	-1.75E-04
	(1.62E-03)	(1.20E-06)	(3.20E-06)	(3.46E-06)
90%	7.36E-03	-3.31E-05	6.96E-05	1.65E-05
	(2.33E-03)	(2.53E-06)	(9.46E-06)	(7.12E-06)
95%	7.71E-03	-1.56E-04	3.04E-04	1.97E-04
	(3.87E-03)	(4.26E-06)	(1.62E-05)	(1.49E-05)
99%	6.44E-03	-2.83E-04	2.64E-04	6.12E-04
	(1.40E-02)	(1.95E-05)	(9.45E-05)	(9.22E-05)

NOTE: n_{rep} = 1000; n_{samp} = 10000. Values in parenthesis below measures reflect the variances of the measures across data samples. The coefficients of government wages (W^G) and industry wages (W^I) are set equal in model estimation (δ_3).

In order to illustrate the global representation of the full parametric posterior by the LI-SEBB, we provide a graph of the posterior distributions for structural parameters in Figure 1. It is clear that the representation of the posteriors by the LI-SEBB is global, and is not limited to a few moment and quantile measures. For comparison purposes, we also plot the aforementioned 2SLS-mapping of the *unrestricted error* representation of the system of equations. It is apparent that the absence of enforcing the overidentifying restrictions implies a noticeably different posterior distribution on the parameters of the model.

Figure 1: Posterior Distributions of the Intercept Parameter in the Structural Equation



5 Conclusions and Outlook

The paper introduces a semi-parametric approach to Bayesian analysis of structural parameters in simultaneous equation systems by extending single and multivariate regression approaches of Heckelei and Mittelhammer (1996, 2002) to models with endogenous regressors. Moreover, the whole underlying idea of mapping the noise distribution to model parameters has been motivated in a more general fashion based on the concept of *Bayesian data information mapping* (BDIM). Monte Carlo evidence was provided that demonstrated the considerable accuracy of the LI-SEBB procedure in approximating posterior distributions under normally distributed errors, even for small sample sizes. However, further simulations are needed to evaluate the accuracy of the simulated posterior distributions under non-normal model noise. Also, the extension of the BDIM and Bayesian Bootstrapping procedures to a full information context is pending.

6 References

Graybill, F.A. *Matrices With Applications in Statistics*. Wadsworth, Inc, Belmont, California, 1983.

Heckelei, T. "Bayesian Bootstrap Analysis of Econometric Models: Theory and Application. Ph.D. Dissertation, Washington State University, 1995

Heckelei, T. and R.C. Mittelhammer. "Bayesian Bootstrap Multivariate Regression". Forthcoming in the *Journal of Econometrics*, 2002.

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

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

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

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

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

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

Ng, V.M. "Robust Bayesian Inference for Seemingly Unrelated Regressions with Elliptical Errors." *Journal of Multivariate Analysis*, doi:10.1006/jmva.2001.2054 :1-6, 2001.

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

Zellner, A. "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), 1996.

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

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