On Bayesian Structural Inference in a Simultaneous Equation Model

by Herman K. van Dijk*

March 21, 2002

Econometric Institute Report EI 2002-10

Abstract

Econometric issues that are considered fundamental in the development of Bayesian structural inference within a Simultaneous Equation Model are surveyed.

The difficulty of specifying prior information which is of interest to economists and which yields tractable posterior and predictive distributions has started this line of research. A major issue is the nonstandard shape of the likelihood due to reduced rank restrictions. It implies that existence of structural posterior moments under vague prior information is a nontrivial issue. The problem is illustrated through simple examples using artificially generated data in a so-called limited information framework where the connection with the problem of weak instruments in classical econometrics is also described.

A positive development is Bayesian inference of implied characteristics, in particular, dynamic features of a Simultaneous Equation Model. The potential of Bayesian structural inference, using a predictive approach for prior specification and using Monte Carlo simulation techniques for computational purposes, is illustrated by means of a prior and posterior analysis of the US business cycle in the period of the depression. A structural prior is elicited through investigation of the implied predictive features.

Some connections with modern time series econometrics are emphasized, in particular, the formal mathematical equivalence of overidentification in a SEM and cointegration in an vector autoregressive model.

It is argued that Bayesian structural inference is like a Phoenix. It was almost a dead topic in the late eighties and early nineties but it has become of renewed importance in models where reduced rank analysis occurs. These models include structural vector autoregressive models,

*I am indebted to Eric Zivot, Rodney Strachan and, in particular, to Frank Kleibergen for stimulating discussions on the topic of this paper. I am also grateful to Lennart Hoogerheide, Bernt Stigum, Tore Schwert and Nils Hjort for helpful comments on an earlier version of this paper. Responsibility for any errors is entirely mine.

asset price theory models, capital asset pricing models, factor models, dynamic panel models, state space models, consumer demand and factor demand systems, and error in variables models.

1 Introduction

A fundamental feature of economic systems based on the market mechanism is that prices are jointly determined by the laws of demand and supply. Another feature of a dynamic market system is the joint expansion and contraction of such economic variables as consumption, investment and gross national output during the business cycle. The econometric analysis of market price behavior and the nature of business cycles is in the twentieth century greatly advanced by the specification of the *Simultaneous Equation Model (SEM).* Important statistical implications of a linear SEM have been indicated by Haavelmo (1943). Seminal studies on the identification and the estimation of SEM parameters are contained in two Cowles Commission monographs (Koopmans (1950) and Hood and Koopmans (1953)). As classical inferential method use is made, in most cases, of the method of maximum likelihood, where an unknown 'true' model structure is assumed . The classical econometrician is 'condemned' to find this true structure. However, Drèze (1962) argues that such classical inference is not adequate since on the one hand available information on parameters is ignored - for instance the marginal propensity to consume is in the unit interval - while on the other hand too many exact zero restrictions are imposed - for instance zero restrictions due to the omission of certain variables in each equation. Drèze's paper has been a major stimulus for Bayesian structural inference of the SEM.

In this chapter we discuss some issues that have dominated the development of Bayesian inference of the SEM in the second half of the twentieth century. The first issue is the specification of prior information. As mentioned above, Drèze started this discussion in 1962 with the common sense arguments that many structural parameters are restricted to *a priori* plausible regions. This does not only hold for these parameters but also for dynamic characteristics of the SEM like correct signs of multipiers and plausible restrictions on the period of oscillation. It is, however, not trivial to incorporate these inequality restrictions in a class of analytically tractable priors like the natural conjugate class. Further, Rothenberg (1963) pointed out that this natural conjugate class is by itself highly restrictive for systems of equations. This result is known as 'Rothenberg's' problem. The difficulty to incorporate flexible prior information like inequality conditions or nonlinear restrictions and the restrictiveness of the natural conjugate class of priors by itself are the starting points of new research that started in the late sixties and the early seventies.

Two developments took place. First, in order to simplify the analysis, single equation inference was pursued by Drèze (1976) and Zellner (1971). In this approach a connection can be made with instrumental variable estimation. We discuss some motivating examples in Section 2 and analyze the limited information approach in detail in Section 3.

A second development has been to tackle the problem of finding numerical integration procedures that are computationally efficient and that allow for the use of flexible prior information which is of interest to economists. The use of Monte Carlo integration methods turned out to be revolutionary. The method of importance sampling was introduced in Bayesian inference by Kloek and Van Dijk (1978); see also Geweke (1989). This was followed by Markov Chain Monte Carlo methods as Metropolis Hastings and Gibbs sampling. An excellent survey is provided by Geweke (1999). We refer to that paper and the references cited there, in particular, Hastings (1970) and Tierney (1994). We will not focus on the description of Monte Carlo integration in this chapter.

The application of numerical procedures to the case of the SEM has not been trivial. Here we come to an important econometric issue in Bayesian inference of the SEM. The shape of the likelihood of a SEM is nonstandard. This holds for the case of a model with no restrictions on the parameters but also for a standard model of a market which is identified according to classical identification conditions. This is different from the case of the linear regression model where the marginal likelihood of parameters of interest belongs to the class of Student-t densities. The problem of a nonstandard shape of the likelihood has been analyzed by Van Dijk (1985), Bauwens and Van Dijk (1989), Kleibergen and Van Dijk (1992, 1994a, 1998). The basic reason is the presence of singularities in the parameter space due to reduced rank restrictions. We illustrate Bayesian structural inference in simultaneous equation analysis through simple examples in Section 3. Section 4 contains a condition which is sufficient for the existence of posterior moments in the full information case.

The flexibility of Monte Carlo integration procedures using simulation methods has greatly advanced Bayesian inference of dynamic features of a SEM. The elicitation of informative priors by investigation of the implications for short and long term prediction is rather trivial using simulation methods. The potential use of Bayesian structural inference using a simple predictive approach and using Monte Carlo as computational tool is illustrated in Section 5 through an analysis of the US business cycle in the period of the great depression.

The simultaneous equations model itself has been under attack by Liu (1960) and, in particular Sims (1980) because of its 'incredible' restrictions. The class of Vector AutoRegressive (VAR) models advocated by Sims is considered as more 'data driven' while the SEM is more 'theory driven'. In Section 6 we make some remarks on Bayesian inference in structural VAR models. Section 7 contains conclusions.

2 Motivation and Two Examples

Consider the stylized wage regression popular in empirical labor studies:

$$y_1 = \beta y_2 + x_1\gamma + u_1 \tag{1}$$

where y_1 is log of hourly wage, y_2 denotes education, which is measured as years of schooling, and x_1 equal to age - years of schooling - 6. This latter variable captures work experience. We note that all variables are given in deviations from their mean values. The structural parameter of interest, β ideally measures the rate of return to schooling. The variable y_2 (years of schooling) is potentially endogenous, however, and correlated with u_1 due to the omission of a variable measuring (unobservable) ability. The degree of endogeneity is potentially very high due to a high expected correlation between education and ability. Classical inference makes use of instrumental variable estimation methods but potential instruments for y_2 (= education) are hard to find since these variables must be correlated with education but uncorrelated with unobserved ability. Angrist and Krueger (1992) suggest using quarter of birth as dummy variables. They argue that quarter of birth is randomly distributed across the population and so is uncorrelated with ability and it affects years of schooling weakly, through a combination of the age at which a person begins school and the compulsory schooling laws in a person's state. Staiger and Stock (1997) give evidence that inference on the rate of return to schooling can be greatly affected by the weak quarter of birth instruments.

As a second example, consider the problem of determining the fraction of temporary income consumers spend in a permanent income/consumption model. Campbell and Mankiw (1989) use the simple regression equation

$$\Delta c = \beta \Delta y + u_1 \tag{2}$$

where c is log consumption, y is log income. In this simple model, β measures the fraction of temporary income consumed. Consumption and income are simultaneously determined and so Δy is potentially highly correlated with u_1. In the permanent income model c and y are cointegrated with cointegrating vector (1,-1) and the error correction model for Δy suggests using lagged values of Δy, Δc and the lagged error correction term, $c - y$, as instruments. However, the growth rate of income is not predicted very well from this error correction model so that the suggested instruments are expected to be fairly weak. Notice that in this example the quality of the instruments is determined by the short-run dynamics in the growth rate of income.

More formally, consider the following two-equation simultaneous equation model

$$y_1 = \beta y_2 + X_1 \gamma + u_1 \tag{3}$$

$$y_2 = X_1 \pi_{21} + X_2 \pi_{22} + v_2 \tag{4}$$

where y_1 is a $(T \times 1)$ vector of observations on the structural equation of interest, y_2 is a $(T \times 1)$ vector of observations on the included endogenous variable, β is the scalar structural parameter of interest, X_1 is a $(T \times k_1)$ matrix of included weakly exogenous variables, γ is a $(k_1 \times 1)$ vector of structural parameters (not of direct interest), X_2 is a $(T \times k_2)$ matrix of excluded weakly exogenous variables or instruments and π_{21} and π_{22} are $(k_1 \times 1)$ and $(k_2 \times 1)$ vectors of reduced form parameters, respectively. It is assumed that $k_2 \geq 1$ so that the necessary condition for identification of β is satisfied and we call $k_2 - 1$ the degree of overidentification. The variables in $X = [X_1 \vdots X_2]$, which may contain lagged predetermined variables, are assumed to be weakly exogenous for the structural parameters β and γ; see Engle, Hendry and Richard (1983). We assume that the rows of the error terms u_1 and v_2 are independently normally distributed with covariance matrix Σ, which has elements σ_{ij} $(j = 1, 2)$. We note that (3)-(4) is known as the INcomplete Simultaneous Equation Model (INSEM).[1]

The interpretation of the parameters is crucial in this context. The parameter $\rho = \sigma_{12}/(\sigma_{11}\sigma_{22})^{1/2}$ measures the degree of endogeneity of y_2 in (1) and π_{22} captures the quality of the instruments. The weak instrument problem occurs in (1)-(2) when $|\ \rho \ | \approx 1$ and $\pi_{22} \approx 0$ so that β is nearly nonidentified.[2] [3] The weak instrument problem in the INSEM has been examined from a Bayesian point of view by Kleibergen and Zivot (1998) and Gao and Lahiri (1999). In the next section we discuss the connection between weak instruments and Bayesian limited information.

3 Limited Information or Incomplete Simultaneous Equation Analysis

3.1 Parameterizations

There are several equivalent ways to parameterize the INSEM and Bayesian analysis is influenced by the adopted parameterization. The *structural form* of the INSEM is given in (3)- (4) and was the parameterization originally analyzed by Zellner (1971), see also Zellner, Bauwens and Van Dijk (1988). The multivariate regression representation of the structural form is

$$YB = X\Gamma + U \tag{5}$$

[1] There is a slight difference between this model and a complete SEM which is analyzed under limited information, as is done by Drèze (1976). For details see Bauwens and Van Dijk (1989). In this chapter we make use of the INSEM.

[2] Empirically, the weak instrument problem is often characterized by a low first stage R^2 or a low first stage F-statistic for testing $\pi_{22} = 0$.

[3] The effects of weak instruments on IV-based inference is discussed in Nelson and Startz (1990a,b), Bound, Jaeger and Baker (1995), Hall, Rudebusch and Wilcox (1996), Staiger and Stock (1997), Wang and Zivot (1997) and Zivot, Startz and Nelson (1997). In these papers it is shown that the 2SLS/IV estimator of β is biased in the same direction as the OLS estimator and the 2SLS/IV estimated standard errors are often spuriously too small which leads to seriously size distorted confidence intervals for β. Wang and Zivot (1997) and Zivot, Startz and Nelson (1997) show that inferences based on likelihood ratio and Lagrange multiplier statistics are more robust than inferences based on Wald statistics.

where $Y = [y_1, y_2]$, $X = [X_1, X_2]$ and $U = [u_1, v_2]$ with

$$B = \begin{bmatrix} 1 & 0 \\ -\beta & 1 \end{bmatrix}, \Gamma = \begin{bmatrix} \gamma & \pi_{21} \\ 0 & \pi_{22} \end{bmatrix}.$$

Since $\mid B \mid = 1$ the likelihood function for the INSEM is of the same form as a seemingly unrelated regressions (SUR) model.
The *unrestricted reduced form* of the model is

$$y_1 = X_1\pi_{11} + X_2\pi_{12} + v_1, \tag{6}$$

$$y_2 = X_1\pi_{21} + X_2\pi_{22} + v_2, \tag{7}$$

In systems form the unrestricted reduced form is

$$Y = X\Pi + V$$

where the rows of v are independently normally distributed with mean zero and covariance matrix Ω. Since (6)- (7) is simply a multivariate linear model all of the reduced form parameters are identified.

The identifying restrictions that tie the structural form to the reduced form are $\gamma = \pi_{11} - \pi_{21}\beta, \pi_{21} - \pi_{22}\beta = 0$, $\sigma_{11} = \omega_{11} - 2\beta\omega_{12} + \beta^2\omega_{22}$, $\sigma_{12} = \omega_{12} - \beta\omega_{22}$ and $\sigma_{22} = \omega_{22}$.

The *restricted reduced form* of the INSEM is obtained by imposing the identifying restrictions on the unrestricted reduced form and is given by

$$y_1 = X_1(\pi_{21}\beta + \gamma) + X_2\pi_{22}\beta + v_1, \tag{8}$$

$$y_2 = X_1\pi_{21} + X_2\pi_{22} + v_2, \tag{9}$$

where $v_1 = u_1 - v_2\beta$. In system form we solve (5) for Y and obtain

$$Y = -X\Gamma B^{-1} + UB^{-1}.$$

where $\Omega = B^{-1'}\Sigma B^{-1}$. The restricted reduced form is a multivariate regression model with nonlinear restrictions on the parameters. In the absence of restrictions on the covariance structure of Σ, β is identified if and only if $\pi_{22} \neq 0$ and $k_2 \geq 1$. If $\rho = 0$, however, then β is identified even if $\pi_{22} = 0$.

The *orthogonal structural* form (Zellner, Bauwens and Van Dijk (1988) reads

$$y_1 = y_2\beta + X_1\gamma_1 + v_2\phi + \eta_1, \tag{10}$$

$$y_2 = X_1\pi_{21} + X_2\pi_{22} + v_2. \tag{11}$$

where

$$\begin{pmatrix} \eta_1 \\ v_2 \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11.2} & 0 \\ 0 & \sigma_{22} \end{pmatrix}\right],$$

$$\sigma_{11.2} = \sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{21} \qquad \phi = \sigma_{22}^{-1}\sigma_{12}$$

The parameter ϕ measures the degree of endogeneity. Note that $\phi = \rho(\sigma_{11}/\sigma_{22})^{\frac{1}{2}}$

We are interested in cases where there are different degrees of identifiability. It is well known that the likelihood is flat when there are nonidentified parameters. It is of interest to note that different degrees of identification correspond to the case of different degrees of the quality of instrumental variables in classical IV estimation.

For convenience we consider the case of a *Just Identified Model*, where $k_2 = 1$. The *structural* representation is now

$$y_1 = y_2\beta + u_1, \tag{12}$$

$$y_2 = x\pi_{22} + v_2, \tag{13}$$

the *unrestricted reduced* form is

$$y_1 = x\pi_{12} + v_1, \tag{14}$$

$$y_2 = x\pi_{22} + v_2, \tag{15}$$

and the *restricted reduced* form is

$$y_1 = x\pi_{22}\beta + \beta v_2 + u_1, \tag{16}$$

$$y_2 = x\pi_{22} + v_2. \tag{17}$$

If $\rho \neq 0$, the structural parameter β is identified provided $\pi_{22} \neq 0$ and can be uniquely recovered from the reduced form via the transformation $\beta = \pi_{12}/\pi_{22}$. In addition, if β is identified then the structural correlation coefficient ρ can be recovered via the transformation $\rho = (\omega_{12} - \beta\omega_{22})/((\omega_{11} - 2\beta\omega_{12} + \beta^2\omega_{22})\omega_{22})^{1/2}$. Similarly, if ρ is identified then β can be recovered from the elements of Ω.

The *orthogonal structural* form is

$$y_1 = y_2\beta + v_2\phi + \eta_1, \tag{18}$$

$$y_2 = x\pi_{22} + v_2. \tag{19}$$

The case of nearly nonidentified parameters or the weak instrument context is when $\pi_{22} \approx 0$ and strong endogeneity occurs when $\mid \rho = \phi \mid \approx 1$.

3.2 Marginal Likelihood or Posterior Analysis with a Uniform Prior of the Structural Model

In this subsection we present a Bayesian analysis of the INSEM for the simple bivariate just identified model with no exogenous variables in the structural equation. We use flat priors for the parameters of interest and avoid getting into the complicated algebra that accompanies the general analysis. Our discussion draws on Zellner (1971), Drèze (1976), Bauwens and van Dijk (1989), Zellner, Bauwens and van Dijk (1988), Kleibergen and van Dijk (1992, 1994a, 1998) and Chao and Phillips (1998). A flat or diffuse prior for the structural parameters β, π_{22} and Σ is of the form

$$p(\beta, \pi_{22}, \Sigma) \propto \mid \Sigma \mid^{-\frac{1}{2}h}, \; h > 0. \tag{20}$$

The likelihood function of the structural model for a sample of size T is

$$L(\beta, \pi_{22}, \Sigma \mid y_1, y_2, x) \propto \mid \Sigma \mid^{-\frac{1}{2}T} \exp\{-\frac{1}{2}tr[\Sigma^{-1}U'U]\}. \tag{21}$$

and so the joint posterior based on the flat prior is

$$p(\beta, \pi_{22}, \Sigma \mid y_1, y_2, x) \propto \mid \Sigma \mid^{-\frac{1}{2}(T+h)} \exp\{-\frac{1}{2}tr[\Sigma^{-1}U'U]\}. \tag{22}$$

Using properties of the inverted Wishart distribution (see Zellner (1971) and Bauwens and Van Dijk (1989)), Σ^{-1} may be analytically integrated out of the joint posterior to give the following joint posterior for (β, π_{22}):

$$p(\beta, \pi_{22} \mid y_1, y_2, x) \propto \mid U'U \mid^{-\frac{1}{2}(T+h-3)} = \mid (u_1 v_2)'(u_1 v_2) \mid^{-\frac{1}{2}(T+h-3)} . \tag{23}$$

Note that $p(\beta, \pi_{22} \mid y_1, y_2, x)$ is of the same form as the concentrated likelihood function for (β, π_{22}) from the maximum likelihood analysis of the INSEM, apart from the degrees of freedom parameter h

The marginal posteriors for β and π_{22} can be determined from (21) using properties of the Student-t distribution and the decomposition

$$\mid (u_1 v_2)'(u_1 v_2) \mid = \mid u_1' u_1 \mid \mid v_2' M_{u1} v_2 \mid = \mid v_2' v_2 \mid \mid u_1' M_{v2} u_1 \mid, \tag{24}$$

where $M_A = I - P_A$, $P_A = A(A'A)^{-1}A'$ for any full rank matrix A. Straightforward calculations yield the following result. [4]

Theorem 1 Structural Inference, Drèze(1976).
The marginal posterior for β is a ratio of Student-t densities and is similar in form to the concentrated likelihood function.

For the exactly identified model one obtains

$$p(\beta \mid y_1, y_2, x) \propto \frac{\mid (y_1 - y_2\beta)' M_x (y_1 - y_2\beta) \mid^{\frac{1}{2}(T+h-5)}}{\mid (y_1 - y_2\beta)'(y_1 - y_2\beta) \mid^{\frac{1}{2}(T+h-4)}}, \tag{25}$$

[4] See Bauwens and Van Dijk (1989) and Kleibergen and van Dijk (1992, 1994a, 1998) for details.

The concentrated likelihood is given as

$$\ell(\beta \mid y_1, y_2, x) \propto \Big[\frac{(y_1 - y_2\beta)' M_x (y_1 - y_2\beta)}{(y_1 - y_2\beta)'(y_1 - y_2\beta)}\Big]^{\frac{1}{2}T}$$

The difference between the concentrated likelihood function and the marginal posterior reflects the difference between concentration and marginalization of the likelihood function (21). The concentrated likelihood function may be interpreted as the posterior density for β resulting from a flat prior *conditional* on the modal values of Σ and π_{22} whereas the marginal posterior for β is obtained after integrating out these parameters for the joint posterior.

For the just identified model we proceed as follows. Let G be the number of equations in the system. Choose $h = G + 1 = 3$ and rewrite (25) as

$$p(\beta \mid y_1, y_2, x) \propto c(\beta) \mid (y_1 - y_2\beta)' M_x (y_1 - y_2\beta) \mid^{-\frac{1}{2}(1+0)} \qquad (26)$$

where 1+0 refers to the dimension of β and the degrees of freedom parameter, respectively.

$$c(\beta) = \Big[\frac{\mid (y_1 - y_2\beta)' M_x (y_1 - y_2\beta) \mid}{\mid (y_1 - y_2\beta)'(y_1 - y_2\beta) \mid}\Big]^{\frac{1}{2}(T+h-4)}$$

It is easily seen that $c(\beta)$ is bounded from above and below by extreme eigenvalues of data matrices of moments of $(y_1 \; y_2)$ and $M_x(y_1 \; y_2)$. Thus (25) is bounded by a Student-t density $t(\beta \mid \hat{\beta}_{2SLS}, s^2 y_2' M_x y_2, 0)$, which is not proper (integrable) for $-\infty \leq \beta \leq \infty$. The nonintegrability or "fat-tails" of the posterior is the result of integrating the joint posterior for (β, π_{22}) across the infinite ridge in the β direction that occurs at $\pi_{22} = 0$. Consequently, unless the range of β is truncated a priori, posterior inference regarding β is not possible. For a related discussion on the consequences of using improper priors for the nonintegrability of posteriors, see Berger (1985) and Berger, Liseo and Wolpert (1999).

For overidentified models, it can be shown that the moments of the marginal posterior for β exist up to the order of overidentification. We note that the difference in the exponents of (25) becomes then larger; for details see Kleibergen and Van Dijk (1998). Hence, the marginal posterior of β is proper if there are at least two instruments. One consequence of this result is that the existence of posterior moments of β depends on the order condition for identification and not the rank condition. *Hence it is possible to sharpen inference about β by simply adding more instruments to the model.* This result was first pointed out by Maddala (1976) and was explained analytically by Kleibergen and Van Dijk (1998).

Finally, for large values of T, the marginal posterior for β may be approximated by

$$p(\beta \mid y_1, y_2, x) \propto \exp\{-\frac{1}{2}(y_1 - y_2\beta)' P_x (y_1 - y_2\beta)\}$$

which is a normal density centered at the 2SLS estimate of β. Thus, for large T the Bayesian results based on flat priors for the structural model will coincide with classical results.

A second result is obtained by marginalizing (25) with respect to β.

Theorem 2 Local Nonidentification and Implied Reduced Form Inference, Kleibergen and Van Dijk (1998)
The marginal posterior of π_{22} is equal to the kernel of a univariate Student-t centered at $\hat{\pi}_{22}^{OLS}$ multiplied by the factor $\mid \pi_{22} \mid^{-1}$.

$$p(\pi_{22} \mid y_1, y_2, x) \propto \mid \pi_{22} \mid^{-1} \mid (y_2 - x\pi_{22})'(y_2 - x\pi_{22}) \mid^{-\frac{1}{2}(T+h-4)} . \qquad (27)$$

The factor $\mid \pi_{22} \mid^{-1}$ creates a non-integrable asymptote since the posterior density is infinite at $\pi_{22} = 0$ and, therefore, the posterior is not a proper density. As a consequence forecasting and decision analysis are not feasible. The intuition behind this result stems from the non-identifiability of β when $\pi_{22} = 0$. When $\pi_{22} = 0$, β is not identified and the joint posterior of (β, π_{22}) is flat in the β direction along the axis where $\pi_{22} = 0$ (see Figures 1.1 - 3.3). As a result, the integral of the joint posterior over β is infinite which produces the asymptote at zero in the marginal posterior of π_{22}.

Kleibergen and Van Dijk (1994a) and Chao and Phillips (1998) point out that the non-integrability of (27) can be interpreted as a pathology that has been imposed on the model by a peculiar prior induced on the reduced form parameters (π_{12}, π_{22}) of (14)-(15). Since $\pi_{12} = \beta\pi_{22}$ there is a 1-1 relationship between (β, π_{22}) and (π_{12}, π_{22}) and a flat prior over (β, π_{22}) implies, by the change of variables formula, the non-flat prior for (π_{12}, π_{22}):

$$p(\pi_{12}, \pi_{22}) \propto p_F^S(\beta, \pi_{22}) \left| \frac{\partial(\beta, \pi_{22})}{\partial(\pi_{21}, \pi_{22}} \right| = \mid \pi_{22} \mid^{-1} .$$

The induced prior for (π_{12}, π_{22}) gives infinite density to the point $\pi_{22} = 0$ in the reduced form and thus the flat prior (20) is far from noninformative for π_{22} since it favors infinitely the point π_{22} in the posterior. Another way of explaining the result is that correlation between β and π_{22} which is present in the likelihood is not reflected in the flat prior. Kleibergen and van Dijk (1994a) point out the similarity between the discontinuity in the Bayesian results and the breakdown of standard classical asymptotic results when $\pi_{22} = 0$. They note that a classical solution is to not allow π_{22} to equal zero, as in the local-to-zero asymptotics of Staiger and Stock (1997), and they argue that a sensible Bayesian procedure should also make this type of restriction. A procedure for dealing with this problem is given by Kleibergen and van Dijk (1998) and discussed in the next subsection.

A simple example with artificially generated data

For illustrative purposes, posteriors are calculated from simulated data from (12)- (13) with $T = 100$, $\beta = 0$, $\sigma_{11} = \sigma_{22} = 1$, $x \sim N(0, 1)$ for three cases representing: strong identification/good instruments ($\pi_{22} = 1$); weak identification/weak instruments ($\pi_{22} = 0.1$) and non identification/irrelevant instruments ($\pi_{22} = 0$). These cases are combined with cases of: strong ($\rho = 0.99$), medium ($\rho = 0.5$) and no ($\rho = 0$) degree of endogeneity. The joint posterior (25) from the simulated data is plotted in Figures 1.1 - 3.3 for the three cases. [5] The bivariate and contour plots are highly informative and give three typical shapes of the marginal likelihood: *bell-shaped, multimodality, and elongated ridges.*

Table 1. The shape of the marginal likelihood/posterior with uniform prior

		Degree of Endogeneity: ρ		
		SE $\rho = .99$	ME $\rho = .5$	NE $\rho = 0$
Level of Identification/ Quality of Instruments	NI $\pi = 0$	elongated ridges (1,1)	(1,2)	(1,3)
	WI $\pi = 0.1$	(2,1)	multi-modality (2,2)	(2,3)
	SI $\pi = 1$	(3,1)	(3,2)	bell-shaped (3,3)

(I) Strong identification/good instruments give regular bell shaped posterior curves

Consider Figures 3.1-3.3. When $\rho = 0$, the bivariate posterior has the familiar bell-like shape of a bivariate normal distribution and is centered near the 2SLS estimate of β (which is also the LIML estimator since the model is just identified) and the OLS estimate of π_{22}. When $\rho = 0.99$, the bivariate posterior reduces to a thin single-peaked bell-like shape. We emphasize that the plots of $p_F^S(\pi_{22} \mid Y, x)$ do not always display the spike at $\pi_{22} = 0$ due to the fact that they are drawn for a finite grid of points that exclude $\pi_{22} = 0$.

[5]The author is heavily indebted to Eric Zivot for helpful discussion on the topic of these experiments and for providing the figures. The posteriors are normalized over the displayed range. All plots are computed using GAUSS 3.2

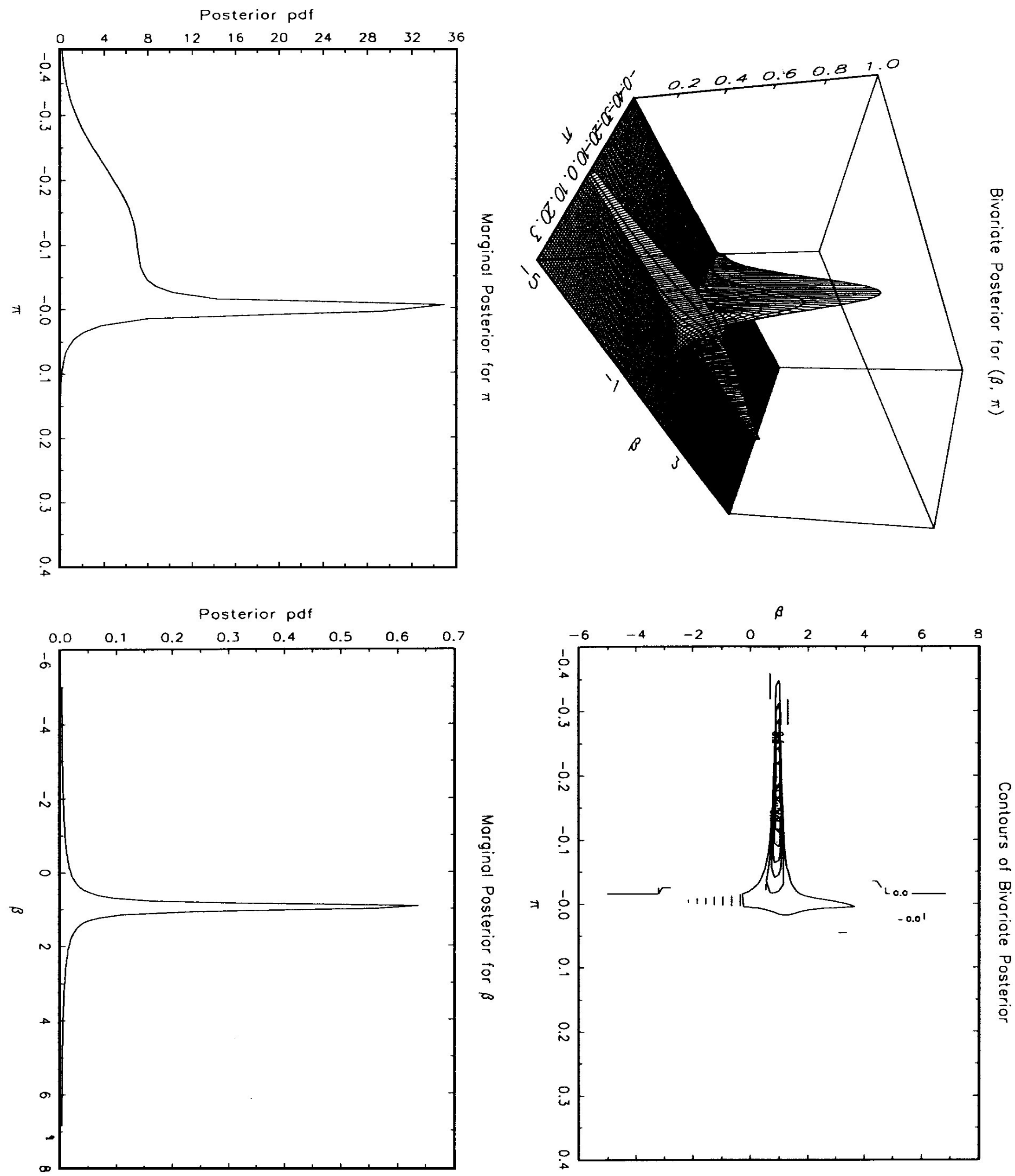

Figure 1.1: Unidentification/Irrelevant Instruments and Strong Endogeneity

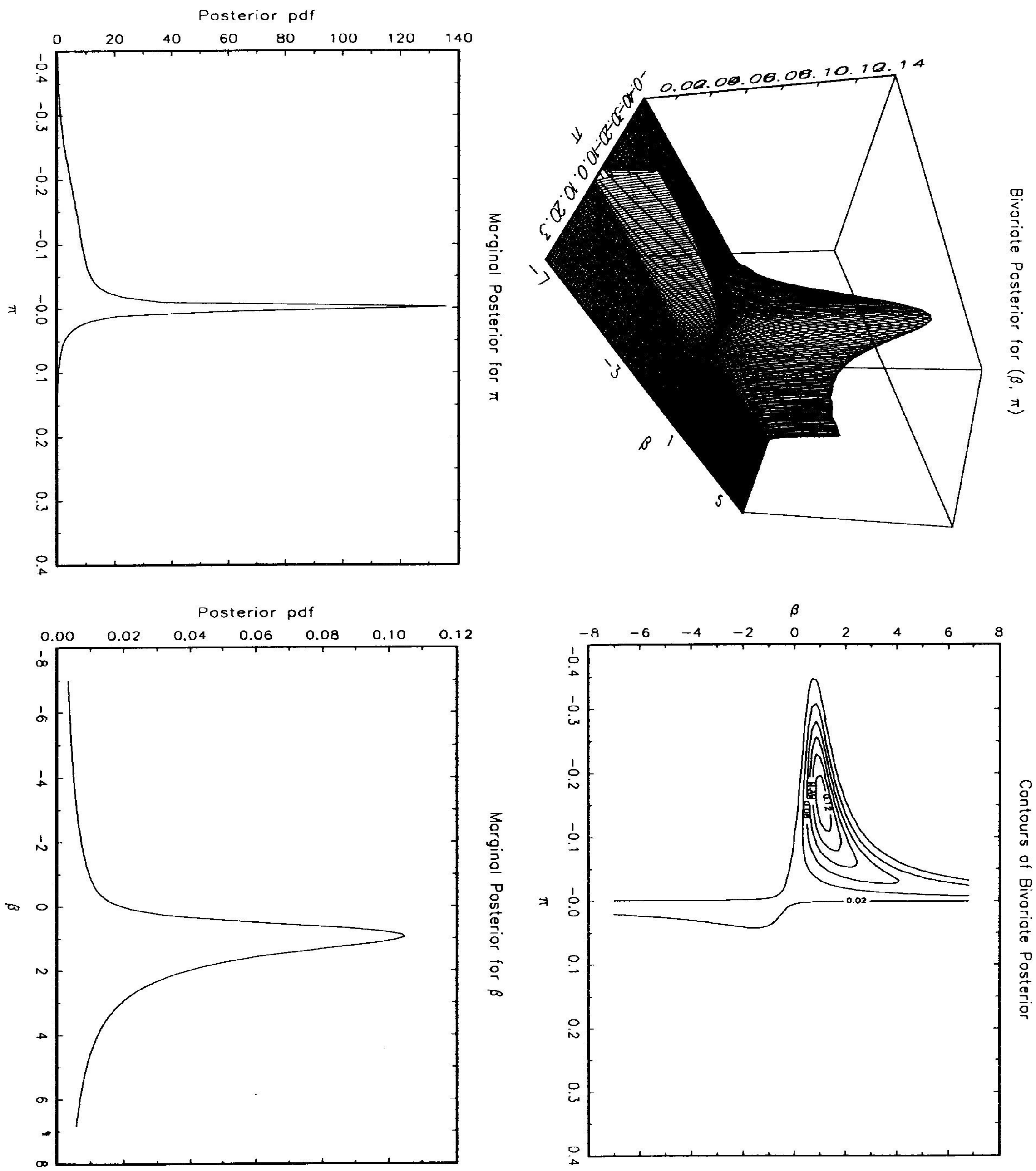

Figure 1.2: Unidentification/Irrelevant Instruments and Medium Endogeneity

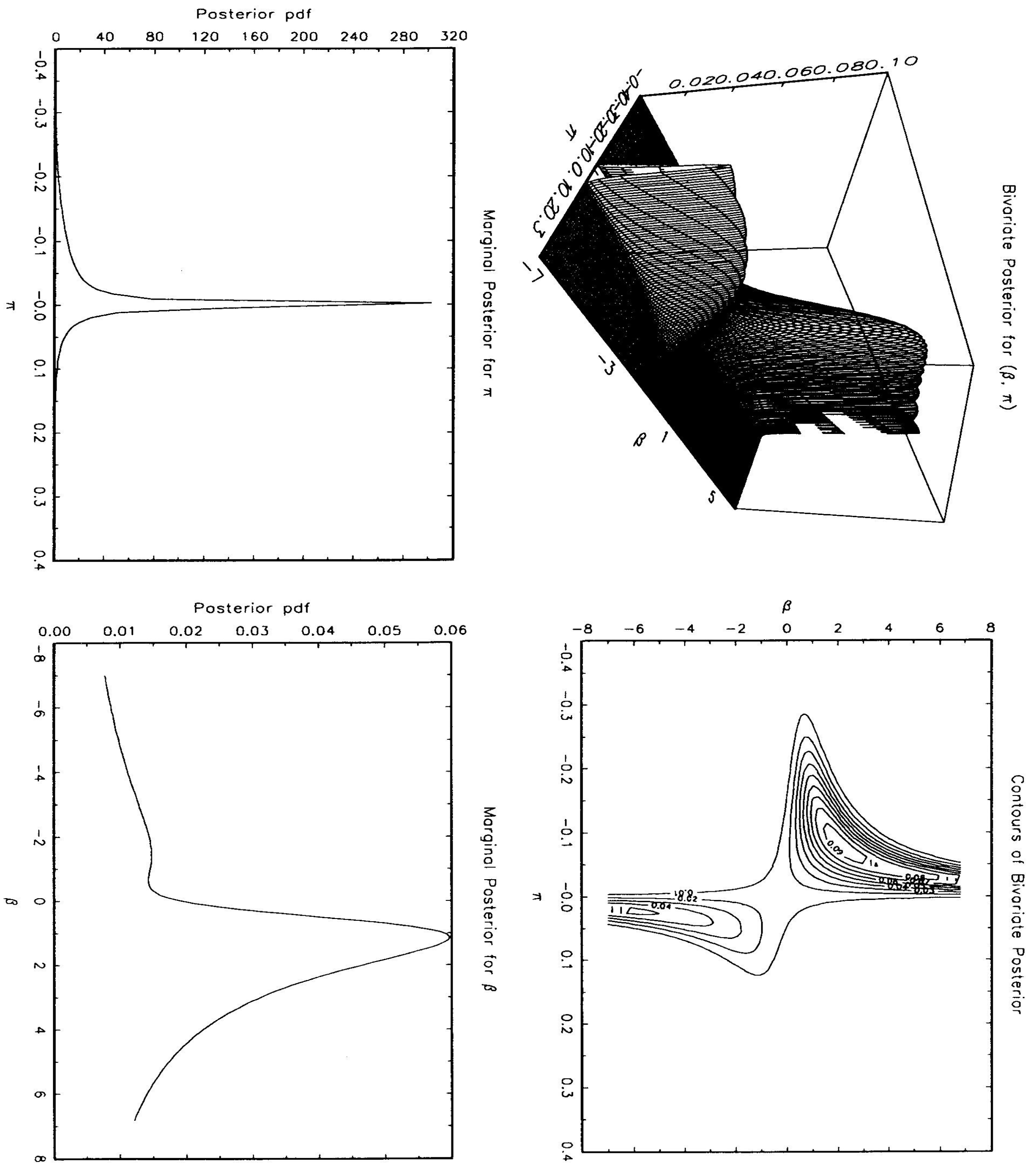

Figure 1.3: Unidentification/Irrelevant Instruments and No Endogeneity

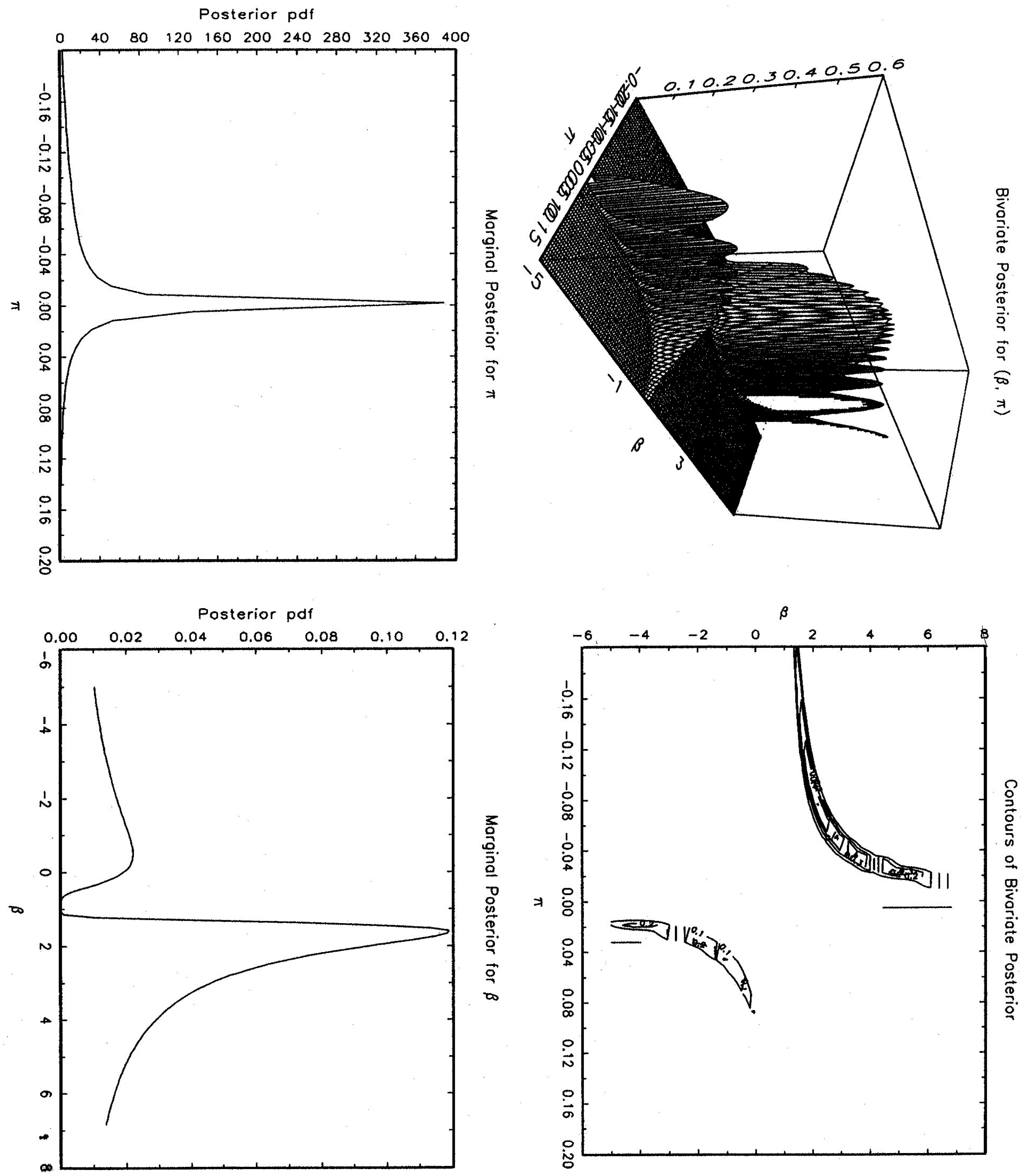

Figure 2.1: Weak Identification/Weak Instruments and Strong Endogeneity

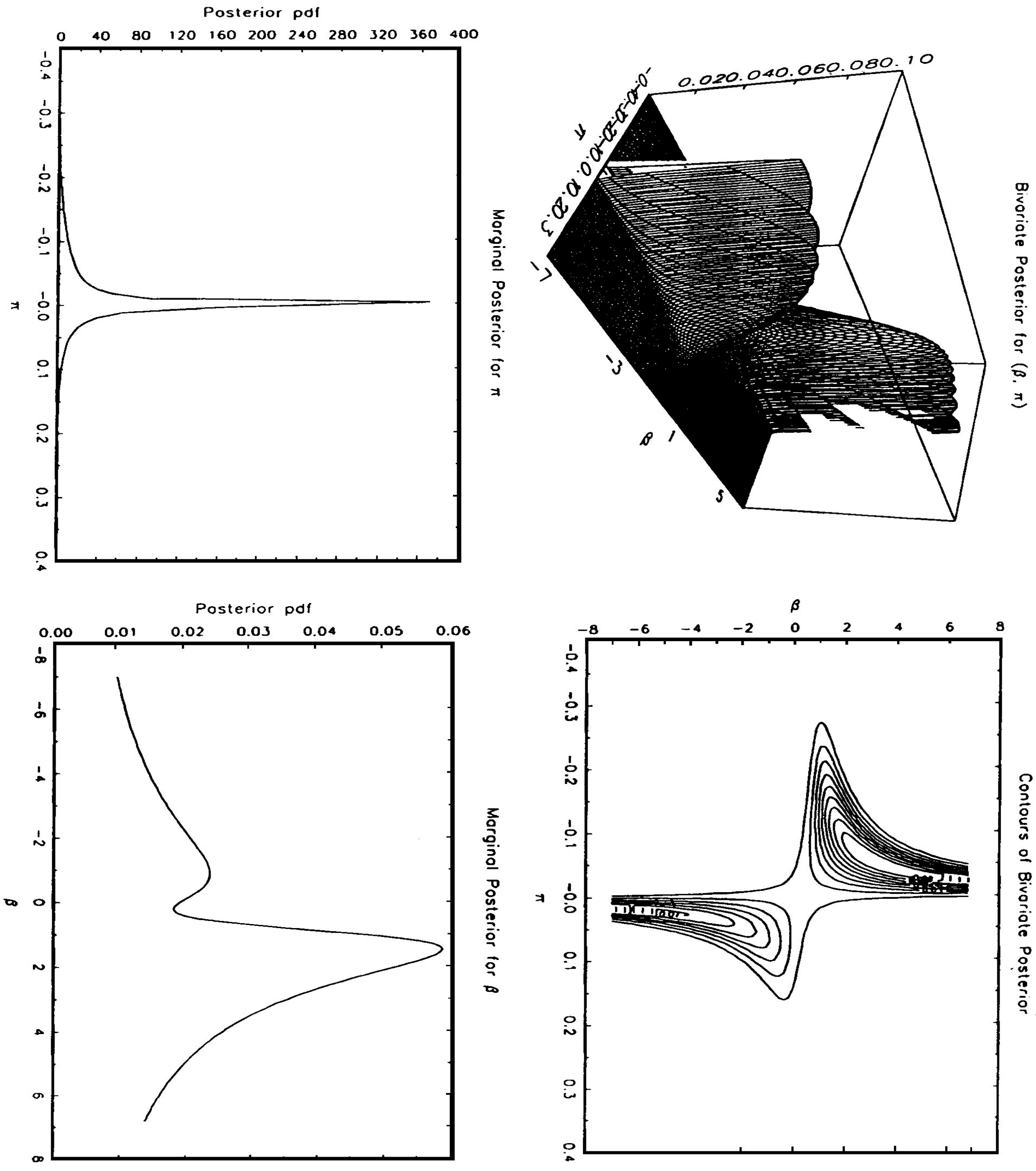

Figure 2.2: Weak Identification/Weak Instruments and Medium Endogeneity

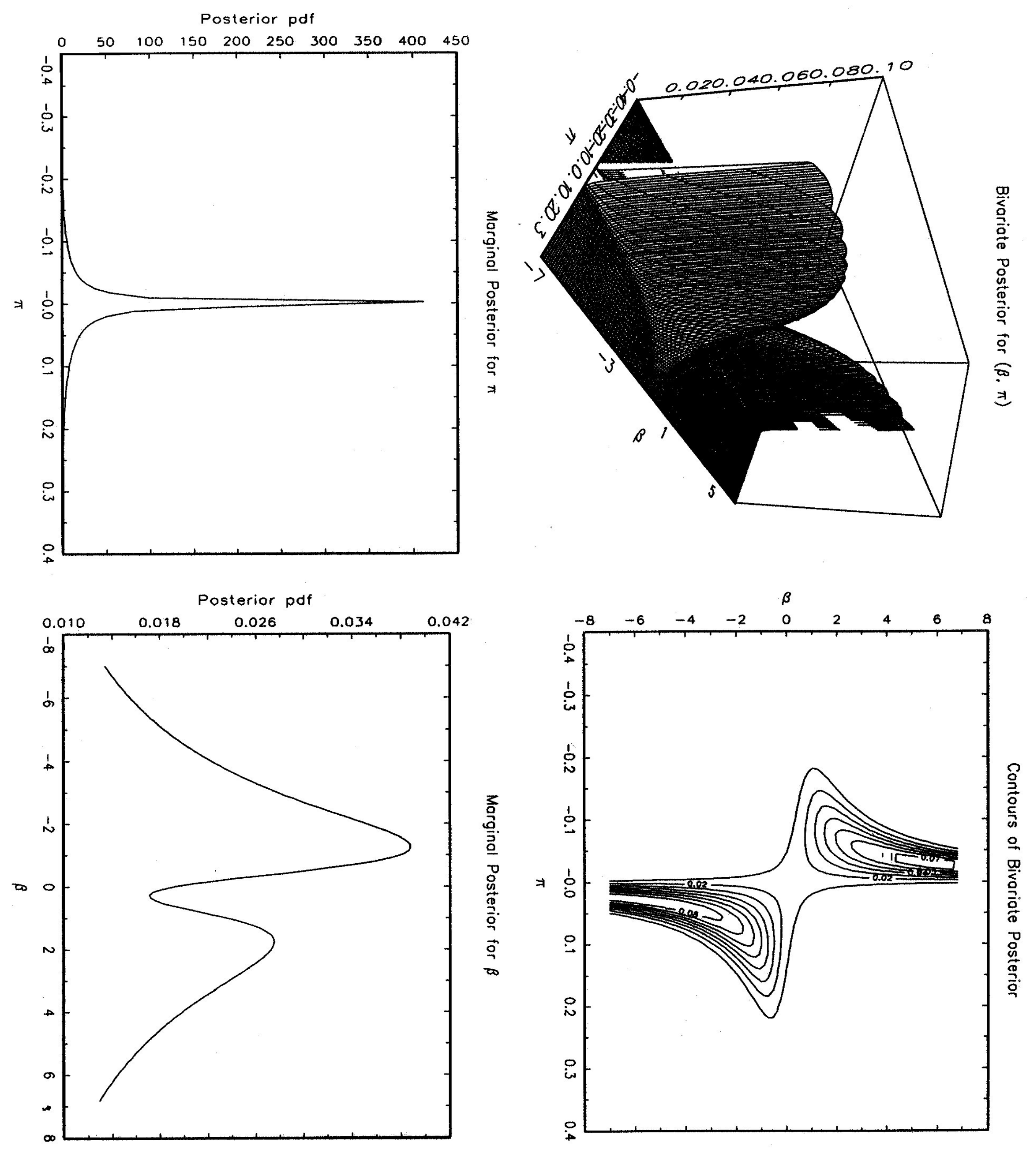

Figure 2.3: Weak Identification/Weak Instruments and No Endogeneity

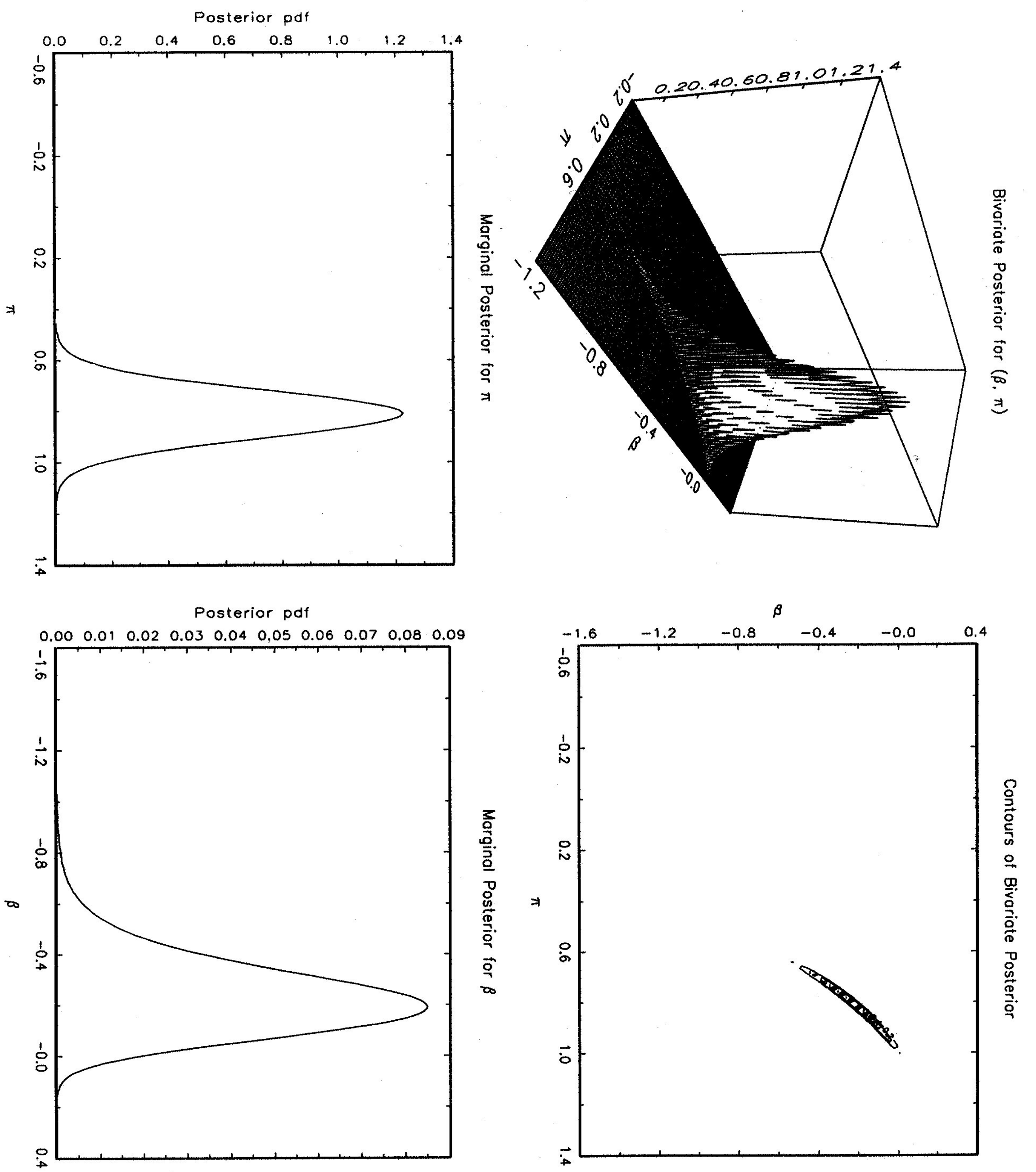

Figure 3.1: Strong Identification/Good Instruments and Strong Endogeneity

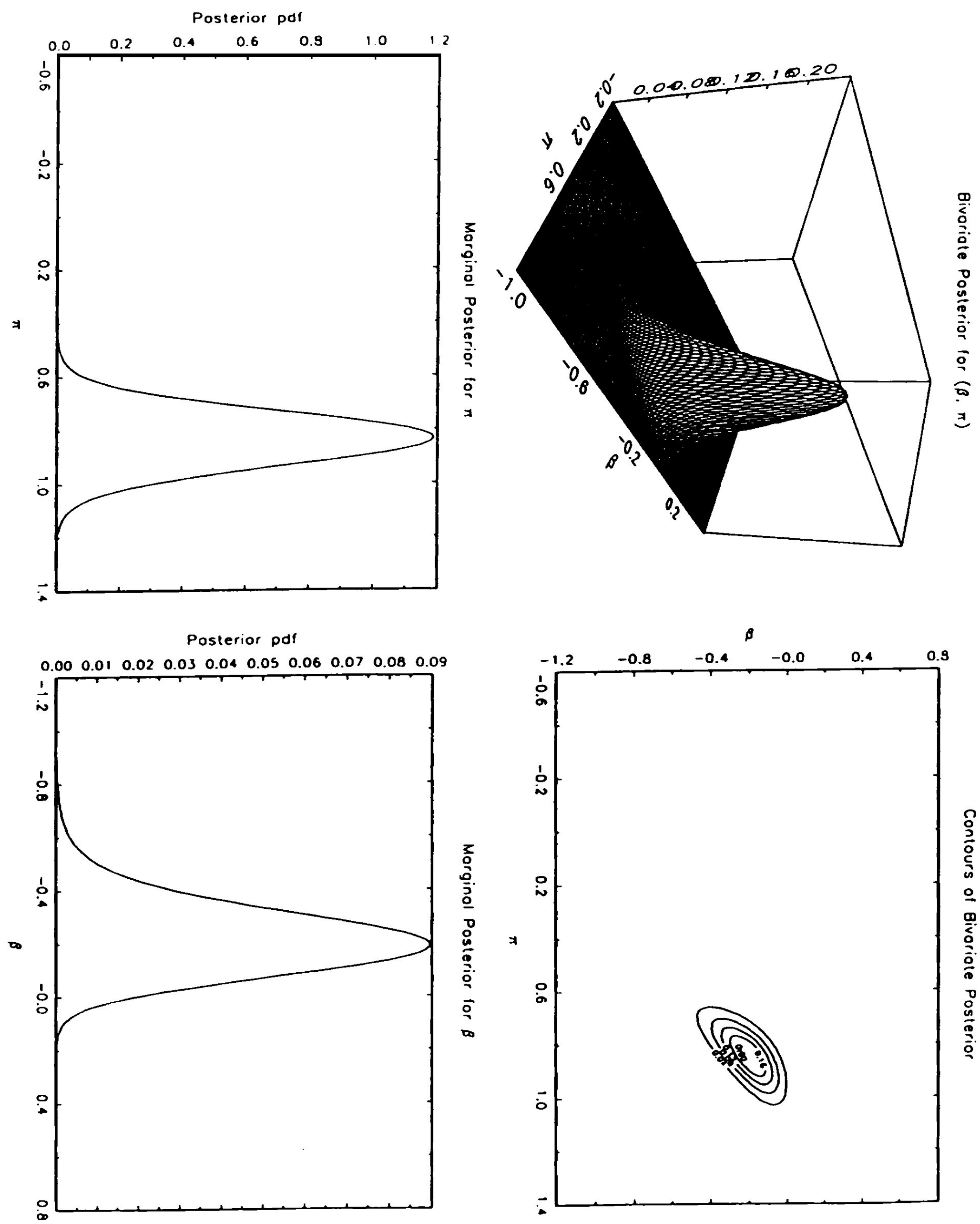

Figure 3.2: Strong Identification/Good Instruments and Medium Endogeneity

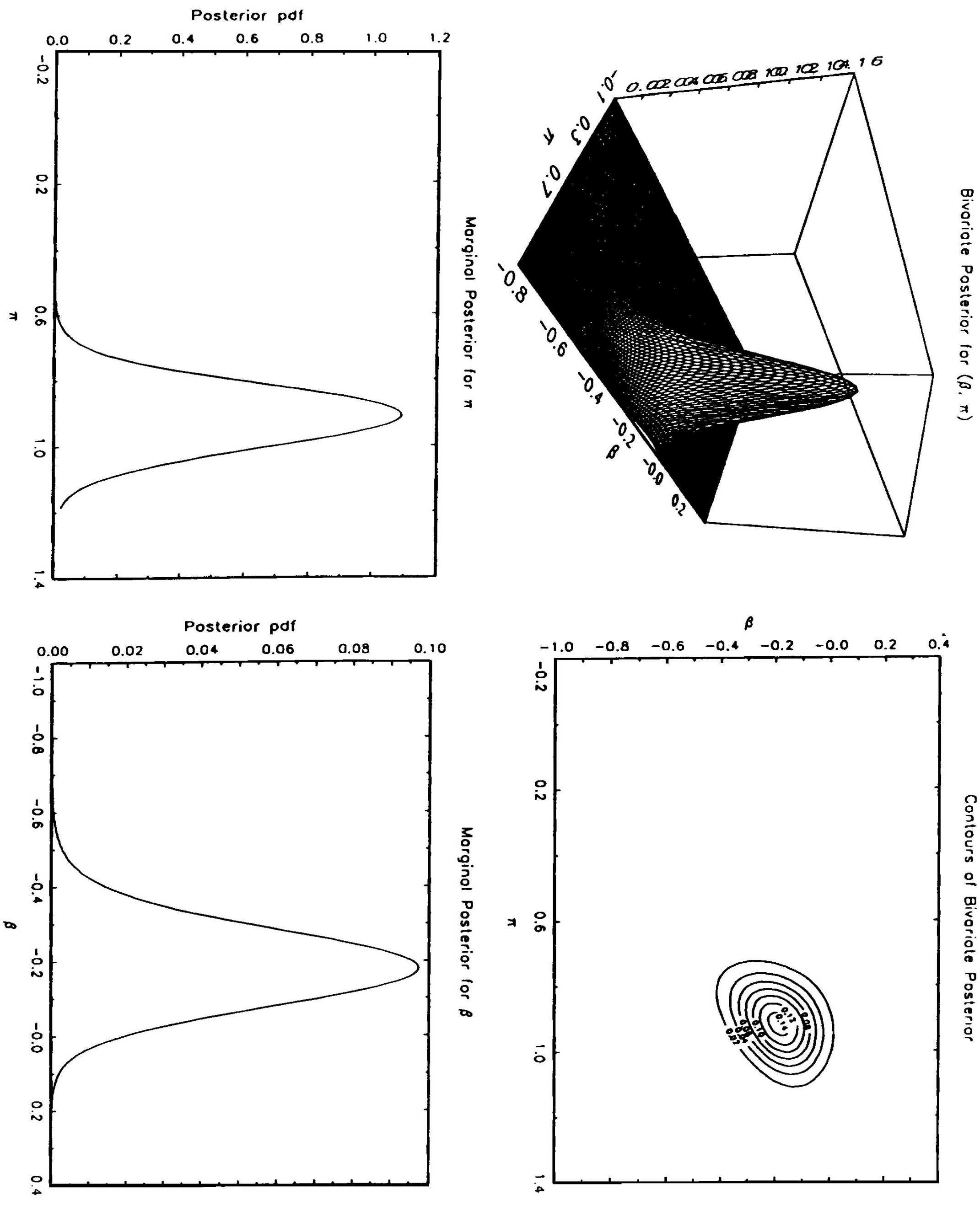

Figure 3.3: Strong Identification/Good Instruments and No Endogeneity

(II) Weak identification/weak instruments give multimodal posterior curves

Consider Figures 2.1-2.3. For the case with $\rho = 0.99$, the bivariate posterior for β, π_{22} is multimodal and actually separates into two L-shaped regions about the point $\pi_{22} = 0$ and $\beta = 0$. In addition, the bivariate posterior develops a long series of bell curves in the β direction at $\pi_{22} = 0$, suggesting that β is undetermined at that value of π_{22}. This "shelf" is the result of the lack of identification of β at $\pi_{22} = 0$.

(III) Nonidentification/irrelevant instruments give elongated ridges in posterior surface

Consider Figures 1.1-1.3. In the irrelevant instrument (unidentified) case, the bivariate posterior is again multimodal but now the "shelf" in the β direction at $\pi_{22} = 0$, is very prominent. The posterior does not separate this time and the contour plot indicates a wide range of possible values for β. This is what we expect with data generated from an unidentified model.

Given the generated data one can easily compute OLS and IV/2SLS estimates for β and π_{22}. For the good instrument case the OLS estimate of β has a strong bias and a small standard error. The IV/2SLS estimate of β is closer to the true value of zero. The 95% confidence interval easily covers zero. The OLS results for the reduced form give a 95% confidence interval which does not cover zero confirming the good quality of the instrument. The marginal posterior for β is very similar to the shape of a normal curve centered on the IV estimate with a standard deviation equal to the standard error of the IV estimate, so the classical and Bayesian analysis give very similar information. Note however the spike in the posterior of π_{22}.

For the weak instrument case, the OLS estimate of β is even more biased and the estimated standard error is smaller than in the good instrument case. The marginal posterior for β is multimodal and has extremely fat tails with secondary modes. The wide distribution of mass reflects a great deal of uncertainty about the true value of β and its convoluted shape reflects the width and prominence of the "shelf" in the bivariate posterior. Interestingly, the posterior degenerates near the OLS estimate which, in turn, is close to the point of concentration of the finite sample distribution of the IV estimate in the case of poor instruments.

In the irrelevant instrument case, the IV estimate of β has a Cauchy distribution centered at the point of concentration (see Phillips (1989)). The marginal posterior for β is unimodal, tightly peaked about the point of concentration but with long fat tails. The posterior is remarkably similar to the finite sample distribution of the IV estimate in the totally unidentified case determined by Phillips (1989). The tightness of the posterior about the point of concentration reflects the inability of the instrument to remove any of the bias due to simultaneity, while the fat tails of the distribution reflects the lack of information in the data about the true value of β.

3.3 Weakly informative priors

The embedding approach

The incomplete simultaneous equation model can be considered as a multivariate regression model with nonlinear restrictions on the parameters. In the embedding approach one starts from an unrestricted reduced form model and considers the transformation to the restricted reduced form. For the exact identified case Kleibergen and van Dijk (1998) start with a flat prior for the unrestricted reduced form parameters $(\pi_{12}, \pi_{22}, \Omega)$,

$$p(\pi_{12}, \pi_{22}, \Omega) \propto \mid \Omega \mid^{-\frac{1}{2}h}, \quad h > 0, \tag{28}$$

and use this prior to deduce a prior for the restricted reduced form parameters β, π_{22} and Ω via the change of variable formula. The absolute value of the Jacobian of the transformation from (π_{21}, π_{22}) to (β, π_{22}) is

$$\mid J \mid = \left| \frac{\partial(\pi_{12}, \pi_{22})}{\partial(\beta, \pi_{22})} \right| = \mid \pi_{22} \mid,$$

so the implied prior for the restricted reduced form parameters is

$$p(\beta, \pi_{22}, \Omega) \propto \mid \pi_{22} \mid\mid \Omega \mid^{-\frac{1}{2}h}, \quad h > 0. \tag{29}$$

Notice that the implied prior for the restricted reduced form parameters degenerates at $\pi_{22} = 0$ and so gives zero prior weight to a model for which β is not identified. Intuitively, one can think of the joint prior (29) as the product of three priors

$$p(\beta, \pi_{22}, \Omega) = p(\beta \mid \pi_{22}) p(\pi_{22}) p(\Omega)$$

where

$$\begin{aligned} p(\beta \mid \pi_{22}) &\propto \mid \pi_{22} \mid \\ p(\pi_{22}) &\propto \text{constant} \\ p(\Omega) &\propto \mid \Omega \mid^{-\frac{1}{2}h}, \; h > 0 \end{aligned}$$

The conditional prior, $p(\beta \mid \pi_{22})$, can be thought of as a limiting form of a normal prior with mean β_0 and variance σ_0^2/π_{22}^2. Such a normal prior has a variance that increases as π_{22} approaches zero.

For the overidentified case the transformation is nontrivial. Here one can make use of a singular value decomposition; see Kleibergen and van Dijk (1998) and Kleibergen (2001) for details.

The Information Matrix approach

The implied prior (29) is very similar to the Jeffreys prior derived from the structural model (12) - (13). The Jeffreys prior is invariant to smooth 1-1 transformations of the parameter space and is proportional to the square root of the determinant of the information matrix of the model. Kleibergen and van Dijk (1994a) derive the Jeffreys prior for the general structural model (4) - (5) and in the just identified case this prior reduces to

$$p(\beta, \pi_{22}, \Sigma) \propto \mid \pi_{22} \mid\mid \Sigma \mid^{-2} .$$

Combining the likelihood function (21) with the prior (29) gives the joint posterior

$$p(\beta, \pi_{22}, \Omega \mid y_1, y_2, x) \propto \mid \pi_{22} \mid\mid \Omega \mid^{-\frac{1}{2}(T+h)} \exp\{-\frac{1}{2} tr[\Omega^{-1} V'V]\}. \tag{30}$$

and integrating out Ω^{-1} one obtains

$$p(\beta, \pi_{22} \mid y_1, y_2, x) \propto \mid \pi_{22} \mid\mid U'U \mid^{-\frac{1}{2}(T+h-3)} \tag{31}$$

The result of the embedding and information matrix approach is that inference on the reduced form parameters is possible but structural inference with weakly or exactly identified models is not feasible for a finite sample. One solution in this respect is to use a penalty function as discussed in Kleibergen and Paap (1998) and Paap and Van Dijk (1999). Another approach is to limit the range of the structural parameters. This approach is presented in the next subsection

4 Full Information Analysis and Restrictions on the Range of Structural Parameters

Consider the linear simultaneous equation model (SEM), given in (5). The prior information with respect to the structural parameters (B, Γ, Σ) is given as follows. The diagonal elements of B are restricted to unity due to normalization and a certain number of zero elements of B and Γ follow from zero identifying restrictions. The unrestricted elements of B and Γ are denoted by the s-vector θ. The stochastic prior information on θ and Σ can be described by the prior density

$$p(\theta, \Sigma) \propto \mid \Sigma \mid^{-1/2h} \tag{32}$$

where h is interpreted as a p rior degrees of freedom parameter. So, we have a uniform prior on θ defined on the region of integration S, where S is equal to the s-dimensional real space.

The likelihood function of the parameters (θ, Σ) can be derived using the assumptions on the model. Combining the prior density (2) with the likelihood gives the joint posterior density of (θ, Σ). Marginalization with respect to Σ can be performed by making use of properties of the inverted-Wishart density. Then one obtains the kernel of the marginal posterior density of θ as

$$p(\theta \mid Y, X) \propto \parallel B \parallel^{T} \mid Q \mid^{-1/2(T+h-G-1)} \tag{33}$$

with

$$
\begin{aligned}
Q &= U'U \\
&= (YB + X\Gamma)'(YB + X\Gamma)
\end{aligned}
\tag{34}
$$

For details we refer to Drèze and Richard (1984), Zellner (1971, Chapter 9) or Bauwens and Van Dijk (1989). The right-hand side of (33) is similar to a concentrated likelihood function of θ as defined by Koopmans and Hood (1953, p. 191). The difference is the exponent of $\mid Q \mid$ which is $-\frac{1}{2T}$ for the concentrated likelihood function and which depends on the particular value of h in our case. For $h = G + 1$ the expressions are proportional.

One can derive an upper bound function to this marginal posterior. For convenience, define $ZA = YB + X\Gamma$, where $Z = (Y \quad X)$ and $A' = (B' \quad \Gamma')$. Let $R(Z)$ denote the rank of Z.

THEOREM 3. Given $R(Z) = G + K$, it follows that

$$
\parallel B \parallel^T \mid Q \mid^{-1/2(T+h-G-1)} \leq c \mid Q \mid^{-1/2(h-G-1)} \qquad (c > 0) \tag{35}
$$

if and only if $R\begin{pmatrix} B \\ \Gamma \end{pmatrix} = G$.

A proof is given in Van Dijk (1985).

We analyze the condition $R(B' \quad \Gamma')' = G$ and analyze a bound on the range of the degrees of freedom parameter h. Let q be a vector of constants, $q \neq 0$, then $q'A'Z'ZAq \geq \varepsilon q'q$, with $\varepsilon > 0$. It follows that

$$
q'A'Z'ZAq \geq (1/2)q'(A'Z'ZA + \varepsilon I)q > 0 \tag{36}
$$

Given that $\mid A'Z'ZA \mid = \prod_{i=1}^{G} \lambda_i$, where λ_i is the i-th characteristic root ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_G > 0$), and given that $\mid A'Z'ZA + \varepsilon I \mid = \prod_{i=1}^{G}(\lambda_i + \varepsilon)$, one obtains

$$
\mid A'Z'ZA \mid \geq (1/2)^G \mid A'Z'ZA + \varepsilon I \mid > 0 \tag{37}
$$

with $\lambda_G \geq \varepsilon > 0$. The determinant given at the right-hand side of the inequality (37) has the same functional form as the determinant of the inverse of a *matricvariate* Student t density [see, e.g., Dickey (1967), Zellner (1971, Appendix B5), and Drèze and Richard (1984, p. 589)]. The difference is, however, the presence of exact restrictions in the matrix A.

It is important to distinguish between the interpretation of the rank conditions on A in a classical and Bayesian framework. The classical rank condition for identification is derived for a set of unknown *constant* structural parameters. Given this rank condition, the event of underidentification of the structural parameters is in the classical approach a set of measure zero. In our Bayesian case the rank condition on A (or on U) has a an other interpretation. That is, since the unrestricted elements of A, given in the vector θ, are random variables

$R(A)$ is a random variable. The assumption $R(A) = G$ is interpreted as follows. The event that a nontrivial linear combination of vectors of parameters of the G equations is zero has prior probability zero and the event $R(A) = G$ has prior probability one. Given $R(A) = G$ with probability one, it follows that $\mid A'Z'ZA \mid > 0$ with probability one. But it is important to make a distinction between two cases of simultaneous equation models. The first case is one where $\mid A'Z'ZA \mid$ tends to zero in the prior parameter region of θ, which we denote by S. It follows that the right-hand side of (35) tends to infinity if $h > G + 1$. The assumption $P[R(A) = G] = 1$ implies, therefore, that the probability function of $R(A)$ has to be truncated at the value $R(A) = G$. In order to obtain a finite upper bound for the right-hand side of (35) one can make use of the following solution. Truncate the uniform prior density $p(\theta)$ in such a way that it is zero on an open subset of the region S where $\mid A'Z'ZA \mid < \varepsilon$, $\varepsilon > 0$ and $p(\theta)$ is equal to a positive constant elsewhere on S. This implies that $\mid Q \mid = \mid A'Z'ZA \mid \geq \varepsilon$ on a subset of S. One can investigate the sensitivity of the upper bound by varying ε. This approach implies that in the evaluation of posterior moments one replaces an infinite integral by a (truncated) finite value of the integral. This may be unattractive approach. Consider, for instance, the following simple two equation model that is interpreted as a market model for a particular commodity,

$$\begin{aligned} q + \beta_1 p + \gamma_1 I &= u_1 \\ q + \beta_2 p + \gamma_2 W &= u_2 \end{aligned} \tag{38}$$

where q represents the quantity traded of a certain commodity and p its price. Values of the endogenous variables p and q are jointly determined given a value of the exogenous variables I and W (e.g., I represents income and W represents weather conditions). If $h > G + 1$, and if γ_1 and γ_2 tend towards zero and β_1 tends towards β_2, then the upper bound function given at the right-hand side of (35) and the kernel $p(\theta \mid Y, Z)$, equation (34) tend both to infinity. The rate at which $p(\theta \mid Y, Z)$ tends to infinity depends on the particular value of h. Note that if $\gamma_1 = \gamma_2 = 0$ and $\beta_1 = \beta_2$, then $\mid A'Z'ZA \mid = 0$ **and** $\parallel B \parallel = \mid \beta_2 - \beta_1 \mid = 0$. So, in this case the upper bound function, mentioned above, is infinitely large, but $p(\theta \mid Y, Z)$ is zero; see Bauwens and Van Dijk (1989) for illustrative examples.

The second case of a linear SEM is one where the exact restrictions on the structural parameters are such that $\mid A'Z'ZA \mid \neq \eta$, $\eta > 0$ for all values of θ in the original region of S. In this case the discrete random variable $R(A)$ has degenerated to the constant G. A simple example is

$$\begin{aligned} C &= \gamma + \beta Y + u \\ Y &= C + Z \end{aligned} \tag{39}$$

where C, Y and Z represent T-vectors of observations on consumption expenditure (C), total expenditure (Y) and autonomous expenditure (Z). After substitution of the second equation into the first one, one can verify in a straightforward way that $\mid Q \mid = \mid u'u \mid \neq \eta$, $\eta > 0$ for all values of γ and β. We note that (39) is a simultaneous equation model with identities.

A third case is to use different normalization restrictions, in particular, a natural restriction is $A'A = I$. In recent work Strachan (2000), Villani (2000) and Strachan and Van Dijk (2001) use these type of restrictions on the parameters such that the condition of this section is fulfilled. As a consequence structural moments will exist.

5 Prior and posterior analysis of the US business cycle during the depression[6]

In this subsection we use an informal predictive approach to construct priors and illustrate Bayesian structural inference within the context of a well-known econometric model: Klein's model I; see Klein (1950).

This model describes the behavior of consumption, investment and wages over, roughly, the period 1920-1940. As such it describes the period of the great depression in the USA. The prior information on B, Γ, and Σ is specified as follows. The prior specification with respect to the nuisance parameters is taken from a noninformative approach. The exactly known parameter values of B and Γ are implied by the specification of Klein's Model I. Then there remain nine unrestricted structural parameters collected in the vector θ.

We specify a number of prior densities of θ and demonstrate how Monte Carlo may be used to investigate the implied prior information with respect to the reduced form parameters, the stability characteristics of the model, and the final form parameters (if these exist).

Our first and simplest prior for the vector θ is uniform on the nine-dimensional unit region minus the region where $\| B \| < .01$. We investigate the implications of our prior information for the multipliers and dynamic characteristics of the model. We obtained the implied prior means and standard deviations of these functions of θ by drawing θ vectors from the nine-dimensional standard uniform distribution. Each θ vector was checked with respect to the condition $\| B \| > .01$. In case this condition was not satisfied, the vector was rejected and replaced by a new vector.

As a next step we modified our first prior in several ways by adding sets of extra constraints, given as follows.

1. The system is assumed to be stable. For that reason we only accept vectors θ satisfying $| DRT | < 1$, where DRT is the dominant root of the characteristic polynomial.

2. The long-run effects in the structural equations are all assumed to be in the unit interval.

3. The short run multipliers are assumed to be less than 5 in absolute value and to have the correct sign.

[6]The material in this subsection is taken from Van Dijk (1986) and Van Dijk and Kloek (1978).

4. The same set of constraints as mentioned under 3 was applied to the long run multipliers.

5. The period of oscillation is assumed to be between 3 and 10 years. This is in accordance with the observed length of business cycles in the period 1890-1920 (see *Historical Statistics of the United States* 1975).

Eight different priors were obtained by combining the sets of extra constraints, 1 to 5, in several ways. We note that, due to space limitations, we present only results based on prior 2 (prior 1 with extra constraint 1) and prior 8 (prior 2 with extra constraints 1-5). Prior and posterior means and standard deviations of the period of oscillation and the dominant root are given, and we also present prior and posterior probabilities of the four states of the system.

Table 2. Means and Standard Deviations of Period of Oscillation and Dominant Root: Probability of States

	Period of Oscillation (years)	| DRT |	Damped Oscillatory	Damped Monotone	Explosive Oscillatory	Explosive Monotone
FIML (no prior)	34.83	.76	NA	NA	NA	NA
Prior 2	5.22 (4.74)	.78 (.17)	.96	.04	0	0
Posterior 2'	15.06 (2.90)	.84 (.08)	.9999	.0001	0	0
Prior 8	5.42 (1.57)	.72 (.18)	.98	.02	0	0
Posterior 8'	9.61 (0.37)	.77 (.08)	.9927	.0073	0	0

NA = not available

We observe that the prior constraints on the period of oscillation have rather large effects. The question arises whether this information is acceptable. The posterior mean and standard deviation of period of oscillation under prior 2 suggest that the hypothesis of a 10-year period is acceptable. Inspection of the prior and posterior densities of the period of oscillation in Figure 4 reveals that for the case of prior 2 the information from the likelihood function has modified the prior information substantially. The posterior probability that the period of oscillation is less than or equal to 10 years is less than .02. Further, the effect of constraint 5 is clearly reflected in the posterior density 8'. These results suggest

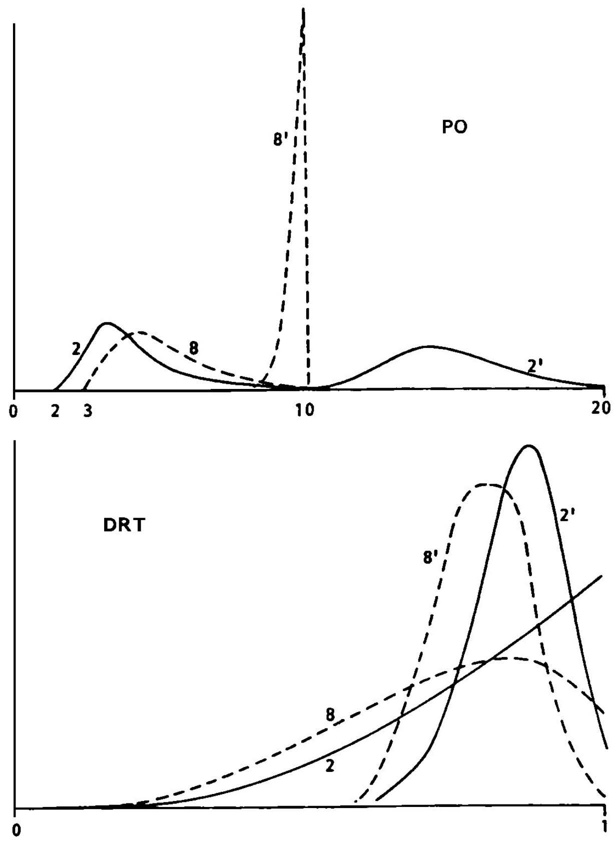

Figure 4: Prior and posterior densities of period of oscillation and dominant root

rejection of the constraint 5. When considering these results we investigated specification errors. Bayesian diagnostic results indicate that there are errors in the dynamic specification of the consumption function. So, instead of reducing the parameter space by making use of the sets of prior constraints 1-5, we have to enlarge the parameter space by including, e.g., lagged consumption in the consumption equation. Preliminary results obtained with an enlarged version of Klein's Model I confirm this.

6 Remarks on Structural Inference in Vector Autoregressive Models

The empirical validity of the large number of structurally identifying restrictions in large scale macro-econometric models was questioned by several researchers in the nineteen sixties and -seventies. These restrictions were considered to be 'incredible'. Further, due to the oil price shock in the seventies and due to the deregulation in several western economic systems in the eighties, many economic time series, in particular financial series, appear to be nonstationary. As a consequence research in the field of dynamic econometric modeling was redirected towards the analysis of systems of equations which have very few restrictions and which allow for nonstationarity.

As an alternative to the class of Simultaneous Equation Models the class of Vector AutoRegressive (VAR) Models was proposed; see Sims (1980). Within this class of models one may investigate the issue of stationary versus nonstationary behavior of linear combinations of economic variables, otherwise stated the issue of cointegration restrictions and the existence of error correction models; see Engle and Granger (1987) and Johansen (1991). Well known examples of cointegration models are permanent income hypotheses and other present value models relating to prices and dividends of stocks, and short and long term interest rates. It is of interest to observe that the mathematical structure of a cointegration model and a simultaneous equation model are the same; see Hoogerheide and Van Dijk (2001). As a consequence the same *local non identification* problem occurs as discussed in Section 3; see Kleibergen and Van Dijk (1994b).

In order to obtain meaningful inference in such models research should be directed towards the construction of flexible vector autoregressive models with some structural restrictions referring to the long run. Interesting results have been obtained in the field of macro-econometric modeling by Garrett et al (2001), Strachan and Van Dijk (2001), Sims and Zha (1999) and Paap and Van Dijk (1999).

7 Conclusions

Given the advances in computing methods and the analytical insight into the shape of the likelihood and the marginal posterior of a SEM, the Bayesian approach to a SEM can give useful insight in the structural effect of certain variables. However, using only vague prior information and relatively short data periods implies that structural inference is very often fragile. The conclusion on fragile finite sample inference is also reached in classical instrumental variable inference; see Dufour (1996), Staiger and Stock (1997) and Zivot, Nelson and Startz (1997). These authors show that traditional classical asymptotic inference, e.g. Wald tests on the structural parameters based on 2SLS break down when instruments are weak due to the near nonidentification of β. As a result, nonstandard methods are required for inference on structural parameters. Using the results of Kleibergen (2001) and Kleibergen and Van Dijk (1998), we have shown that similar results hold for Bayesian inference methods. Using information from a predictive approach strengthens inference considerably. Data based priors may also be useful in this respect; see Schweder and Hjort (2002). Flexible VAR models with some structural restrictions are an interesting topic for research, forecasting and policy analysis. We emphasize that the specification of informative proper priors on a large econometric model is not a trivial matter. This is the reason why the approaches discussed in this paper are explored.

It may be concluded that Bayesian structural inference is like a Phoenix. It was almost a dead topic in the late eighties and early nineties but it has become of renewed importance in models where reduced rank analysis occurs. These models include structural vector autoregressive models, asset price theory models, capital asset pricing models, factor models, dynamic panel models, state space models, consumer demand and factor demand systems, and error in variables models.

References

Angrist, J.D. and A.B. Krueger, 1991, Does compulsory school attendance affect schooling and earnings?, *Quarterly Journal of Economics*, CVI, 979-1014.

Bauwens, L. and H.K. van Dijk, 1989, Bayesian limited information analysis revisited. In: J.J. Gabszewicz et al. (eds.), *Economic Decision-Making: Games, Econometrics and Optimisation* (North-Holland, Amsterdam).

Berger, J.O., 1985, Statistical Decision Theory (Springer-Verlag, Berlin).

Berger, J.O., B. Liseo and R.L. Wolpert, 1999, Integrated likelihood methods for eliminating nuisance parameters (with discussion), *Statistical Science* 14, 1-28.

Bound, J., D.A. Jaeger, and R.M. Baker, 1995, Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak, *Journal of the American Statistical Association* 90, 443-450.

Campbell, J.Y. and G.N. Mankiw, 1989, Consumption, income, and interest rates: reinterpreting the time series evidence, *NBER Macroeconomics Annual 1989*.

Chao, J.C. and P.C.B. Phillips, 1998, Bayesian posterior distributions in limited information analysis of the simultaneous equation model using Jeffreys' prior. *Journal of Econometrics* 87, 49-86.

Dickey, J.M. 1967, Matricvariate generalizations of the multivariate t distribution and the inverted multivariate t distribution, *The Annals of Mathematical Statistics* 38, 511-518.

Drèze, J.H., 1962, The Bayesian approach to simultaneous equations estimation, ONR Research Memorandum 67, The Technological Institute (Northwestern University, IL).

Drèze, J.H., 1976, Bayesian limited information analysis of the simultaneous equations model, *Econometrica* 44, 1045-1075.

Drèze, J.H. and J.F. Richard, 1984, Bayesian analysis of simultaneous equations systems. In Z.Griliches and M.D. Intrilligator (eds.), *Handbook of Econometrics*, vol. 1, (North-Holland, Amsterdam).

Dufour, J-M, 1994, Some impossibility theorems in econometrics with applications to instrumental variables, dynamic models and cointegration, *Econometrica*.

Engle, R.F. and C.W.J. Granger, 1987, Co-integration and error correction: representation, estimation and testing, *Econometrica*, 55: 251-276.

Engle, R.F., D.F. Hendry and J.F. Richard, 1983, Exogeneity, *Econometrica* 51, 277-304.

Gao C. and K. Lahiri, 1999, A comparison of some recent Bayesina and Calssical procedures for simultaneous equations models with weak instruments, Manuscript.

Garratt, A., K. Lee, M.Hashem Pesaran, Y. Shin (2001), A long run structural macroeconometric model of the UK, Manuscript.

Geweke, J., 1989, Bayesian inference in econometric models using Monte-Carlo integration, *Econometrica* 57, 1317-1339.

Geweke, J., 1999, Using simulation methods for Bayesian econometric models: inference, development, and communication, *Econometric Reviews* 18(1), 1-74.

Haavelmo, T., 1943, The statistical implications of a system of simultaneous equations, *Econometrica* 11, 1-12.

Hall, A.R., G.D. Rudebusch and D.W. Wilcox, 1996, Judging instrument relevance in instrumental variables estimation, *International Economic Review*, 37, 283-289.

Hastings, W.K., 1970, Monte Carlo sampling using Markov chains and their applications, *Biometrika* 57, 97-109.

Hood, Wm.C. and T. C. Koopmans, eds., 1953, *Studies in Econometric Method*, Yale University Press, New Haven.

Hoogerheide, L.F. and H.K. van Dijk, 2001, Comparison of the Anderson-Rubin test for overidentification and the Johansen test for cointegration, Econometric Institute report 2001-04, 14 pages.

Johansen, S., 1991, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica*, 59: 1551-1580.

Kleibergen, F.R., 2001, How to overcome the Jeffreys-Lindleys paradox for invariant Bayesian inference in regression models, 22 pages.

Kleibergen, F.R. and H.K. van Dijk, 1992, Bayesian simultaneous equation models analysis: On the existence of structural posterior moments, Econometric Institute report 9269/A, Erasmus University Rotterdam.

Kleibergen, F.R. and H.K. van Dijk, 1994a, Bayesian analysis of simultaneous equation models using noninformative priors, Discussion paper TI 94-134, Tinbergen Institute.

Kleibergen, F.R. and H.K. van Dijk, 1994b, On the shape of the likelihood/posterior in cointegration models, *Econometric Theory* 10, 514-552.

Kleibergen, F.R. and H.K. van Dijk, 1998, Bayesian simultaneous equations analysis using reduced rank structures, *Econometric Theory* 14, 701-743.

Kleibergen, F.R. and E. Zivot, 1998, Bayesian and classical approaches to instrumental variable regression, Manuscript, 1998.

Klein, L.R., 1950, *Economic fluctuations in the United States*, 1921-1941 (Wiley, 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.

Koopmans, T.C., ed., 1950, *Statistical inference in dynamic economic models* (Wiley, New York).

Liu, T.C., 1960, Underidentification, structural estimation, and forecasting, *Econometrica* 28, 855-865.

Maddala, G.S., 1976, Weak priors and sharp posteriors in simultaneous equation models, *Econometrica* 44, 345-351.

Nelson, C.R. and R. Startz, 1990a, Some further results on the exact small sample properties of the instrumental variables estimator, *Econometrica*, 58, 967-976.

Nelson, C.R. and R. Startz (1990b), The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one, *Journal of Business*, 63, S125-S140.

Paap, R. and H.K. van Dijk, 1999, Bayes estimates of Markov trends in possibly cointegrated series: an application to US consumption and income, Econometric Institute report 9911, 36 pages.

Phillips, P.C.B., 1989, Partially identified econometric models, *Econometric Theory* 5, 181-240.

Rothenberg, T., 1963, A Bayesian analysis of simultaneous equation systems, Econometric Institute Report 6315, Erasmus University Rotterdam.

Schweder, T. and N.L. Hjort, 2001, Frequentist analogues of priors and posteriors, Chapter ?? of this book.

Sims, C., 1980, Macroeconomics and reality, *Econometrica* 48, 1-48.

Sims, Ch.A., and T. Zha, 1999, Error bands for impulse responses, *Econometrica*, 67(5), 1113-1155.

Staiger, D. and J.H. Stock, 1997, Instrumental variables regression with weak instruments, *Econometrica*, 65, 557-586.

Strachan, R.W., 2000, Valid Bayesian estimation of the cointegrating error correction model, Report 6/2000 of the Department of Econometrics and Business Statistics, Monash University.

Strachan, R.W. and H.K. van Dijk, 2001, The value of structural information in Vector Autoregressive models, Manuscript.

Tierney, L., 1994, Markov chains for exploring posterior distributions, *Annals of Statistics* 22, 1701-1762.

Van Dijk, H.K. and T. Kloek, 1978, Posterior analysis of Klein's model I, *Szigma, Hungarian Journal of Mathematical Economics*, 121-143.

Van Dijk, H.K., 1985, Existence conditions for posterior moments of simultaneous equation model parameters, report 8551 of the Econometric Institute, 31 pages.

Van Dijk, H.K., 1987, Some advances in Bayesian estimation methods using Monte Carlo integration. In: T.B. Fomby and G.F. Rhodes, eds., *Advances in econometrics* 6, (JAI Press, Greenwich, CT).

Villani, M., 2000, *Aspects of Bayesian cointegration*, thesis, Department of Statistics, Stockholm University.

Wang, J. and E. Zivot, 1998, Inference on structural parameters in instrumental variables regression with weak instruments, *Econometrica*, 66, 1389-1404.

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

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

Zivot, E., R. Startz and C.R. Nelson, 1998, Valid confidence intervals and inference in the presence of weak instruments, *International Economic Review*, 39, 1119-1246.