A PREDICTIVE LIKELIHOOD APPROACH TO POSSIBLE ENDOGENEITY – AN APPLICATION WITH US INCOME-EDUCATION DATA: Trade-off between Estimation Precision and the Necessity of Instruments

Nalan Baştürk* Lennart F. Hoogerheide† Herman K. van Dijk†

February 14, 2011

(PRELIMINARY DRAFT, PLEASE DO NOT CITE)

Keywords: *predictive likelihoods, IV models, returns to education.*

*Econometric Institute, Erasmus University Rotterdam, The Netherlands, corresponding author.

†Tinbergen Institute and Econometric Institute, Erasmus University Rotterdam, The Netherlands.

1 Introduction

A simple regression of earned income on years of education in order to measure the education-income effect neglects important issues. Although higher education levels are expected to increase an individual's earnings, the income-education relationship can be subject to omitted variables such as individuals' intellectual capabilities, measurement errors in reported earnings, or simultaneity as individuals can determine the amount of education they receive judging the possible monetary returns [Angrist et al., 1996].

A more formal way to state the issue is the following: The analysis of causal effects of economic variables brings difficulties when the explanatory variable is itself not exogenous. This *endogeneity* problem might arise as a result of the above mentioned phenomena of omitted variables, measurement errors, or simultaneity between the variables of interest [Goldberger, 1972]. The standard treatment of such relationships is the use of Instrumental Variable (IV) models[1]. IV models rely on a set of instruments (proxies) that can be used to make inference about the endogenous variable (see e.g. Sargan [1958], Bowden and Turkington [1990]).

One of the drawbacks of IV models is the loss of precision, stemming from the use of only an approximation to the explanatory variable. This issue is relatively more severe in case of *weak instruments*, where the instrument has a small amount of information regarding the endogenous variable. The main purpose of this paper is to define a general approach to assess alternative models' performance in analyzing the income-education relationship, and in general in IV models. We consider the performance of the proposed method in simulated datasets and Angrist and Krueger [1991] data on income and education, which includes quarter of birth as instruments[2]. This dataset has been analyzed in several studies, providing a common result that these instruments are quite weak in explaining education, at least in parts of the data [Bound et al., 1995, Hoogerheide and Van Dijk, 2006].

For assessing model performance, we rely on Bayesian methods, as they provide general probabilistic tools to explicitly account for parameter and model uncertainty. Regarding the latter, Bayesian treatment of model uncertainty relies on evaluating posterior model probabilities and the degree to which alternative models are suitable to the data [Clyde and George, 2004].

In the existence of model uncertainty, it is often very difficult or impossible to determine the true model structure, that is which model is correct (for an extensive discussion on model uncertainty and adequacy, see Geweke [2010]). For example, in the income-education analysis, one can define two models, one treating education as an endogenous explanatory variable (i.e. the IV model), and the second model treating education as an exogenous explanatory variable (i.e. the simple linear regression model). To choose one of these competing models can be problematic. In such situations, Bayesian Model Averaging (BMA) provides a theoretical motivation [Bates and Granger, 1969], and has been applied

[1] We note that there is a difference in terminology between IV *models* and IV *estimation technique*. We prefer to use the term 'IV model' to denote the multi-equation model, an incomplete simultaneous equations model, in which the regressor in the key equation of interest can be endogenous, and this problem is treated using instrumental variables.

[2] Angrist and Krueger [1991] quarter of birth as instruments based on schooling laws on compulsory education, which allow individuals to end their education only at a certain age.

mainly in the forecasting literature by success [Min and Zellner, 1993, Eklund and Karlsson, 2007, Geweke and Amisano, 2010]. The advantage of BMA over selecting one of the alternative models is to account for the uncertainty inherent in the model selection process by averaging over many different competing models. It appears more plausible and effective to weight the evidence of the alternative model structures and to take a weighted average of the two or more structures to predict the effect of the explanatory variable on the dependent variable.

Analyzing model performance (hence the implementation of BMA) brings challenges when one does not have strong prior information about the data [Bartlett, 1957]. Following [Hoogerheide and Van Dijk, 2006], we adopt uninformative priors for the income-education analysis and show that the problems associated with uninformative priors in the IV models can be avoided using a *predictive likelihood* approach, recently studied by Eklund and Karlsson [2007].

Predictive likelihoods can be interpreted as follows: the posterior results for a subsample of the data, the *training sample*, is treated as the prior for the analysis of the remaining observations, the *hold-out sample*. Hence the effect of the originally noninformative prior is eliminated (see Laud and Ibrahim [1995] for a discussion on predictive likelihood methods in the Bayesian context).

We apply this technique to simulated datasets with differing degrees of endogeneity problem and instrument strength, and the US data on the income-education relationship. We show that this method can be used to weight the evidence of different models and to assess some of the highly debated issues in IV models, such as the trade-off between choosing the correct model and precision.

Our empirical results show that the income-education relationship in the US states and divisions, using quarter of birth of individuals, is subject to significant heterogeneity in terms of the degree of endogeneity in education levels, as well as the strength of instruments. Hence the analysis of this data in the regional or divisional level can be quite misleading. Further, for this data we conclude that the precision losses from employing an IV model can be quite severe. We rather propose a model averaging approach to infer returns to education, for which the model weights take into account the degree of endogeneity and the instrument strength.

In sum, the methodological contribution of this study is the following: We show that the evidence for alternative IV models can be assessed using predictive likelihoods instead of marginal likelihoods, and BMA using predictive likelihoods provides a tool for efficient measurement of the marginal effects of regressors. On the other hand, the empirical contributions of this study are as follows: For the US data on the income-education relationship, we address the issues of weak instruments and degree of endogeneity, and show that for this dataset, the use of an IV model alone is not supported. A weighted average using an IV model and a simple regression model without instruments provides a more appropriate method to analyze the effects of education on income.

The remainder of this paper is organized as follows: Section 2 presents the IV model, the predictive likelihood approach, alternative models in case of possibly endogeneity and the posterior sampler. Section 3 applies the proposed method to simulated datasets with differing degrees of endogeneity and instrument strength. Section 4 applies the proposed method to the Angrist and Krueger [1991] income-education data. Section 5 concludes.

2 Standard IV model and the predictive likelihood approach

The IV model with one endogenous explanatory variable and p instruments is defined by [Bowden and Turkington, 1990]:

$$y_1 = y_2\beta + \epsilon, \tag{1}$$

$$y_2 = x\Pi + \nu, \tag{2}$$

where y_1 is the $N \times 1$ vector of the dependent variable, y_2 is the $N \times 1$ vector of the endogenous explanatory variable, x is the $N \times p$ matrix of instruments and all variables are demeaned i.e. x does not include a constant term. For $i \in 1, \ldots, N$, $\begin{pmatrix}\epsilon_i & \nu_i\end{pmatrix}' \sim NID\left(0, \Sigma\right)$ where $\Sigma = \begin{bmatrix}\sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22}\end{bmatrix}$ is a positive definite symmetric (PDS) matrix, and the disturbances are uncorrelated across observations i.

The matrix notation for the model is given by:

$$Y = x\Pi B + \upsilon, \tag{3}$$

where $Y = \begin{bmatrix}y_1 & y_2\end{bmatrix}$ is the $N \times 2$ matrix of endogenous variables, x is the $N \times p$ matrix of instruments, Π is $p \times 1$ vector of parameters, $B = \begin{bmatrix}\beta & 1\end{bmatrix}$, where β is a scalar, $\upsilon = \begin{bmatrix}\epsilon & \nu\end{bmatrix}\begin{bmatrix}1 & 0 \\ \beta & 1\end{bmatrix}$ is the $N \times 2$ matrix of error terms with $\text{vec}\left(\upsilon\right) \sim N\left(0, \Omega \otimes I_N\right)$, where $\Omega = \begin{bmatrix}1 & 0 \\ \beta & 1\end{bmatrix}' \Sigma \begin{bmatrix}1 & 0 \\ \beta & 1\end{bmatrix}$.

The likelihood function for the model in (3) is:

$$p\left(Y \mid x, B, \Pi, \Omega\right) \propto |\Omega|^{-\frac{N}{2}} \exp\left(-\frac{1}{2} tr\left(\Omega^{-1}\left(Y - x\Pi B\right)'\left(Y - x\Pi B\right)\right)\right). \tag{4}$$

Next, we briefly discuss some of the important issues in estimating the IV model.

Instruments' validity: The instruments x are said to be valid only if the condition $\text{cov}\left(x_i, \epsilon_i\right) = 0$ is satisfied. Possible endogeneity of instruments has been recently analyzed by Conley et al. [2008], allowing for a certain amount of variation in this condition.

Instrument strength: In the IV model, strength of the instruments is reflected by parameter Π. In the extreme case when $\Pi = 0$, the instruments have no explanatory power for the possibly endogenous explanatory variable y_2. This case is referred to as the case of *irrelevant* instruments.

The degree of endogeneity in the data: Assuming valid instruments, one can immediately see that the endogeneity problem in the model simply arises from the correlation between the structural errors:

$$\text{cov}\left(y_{2i}, \epsilon_i\right) = \text{cov}\left(x_i\Pi + \nu_i, \epsilon_i\right) \tag{5}$$

$$= \text{cov}\left(\nu_i, \epsilon_i\right). \tag{6}$$

Therefore the instruments are only necessary if the correlation between the structural errors, $\rho \equiv \text{corr}\left(\epsilon_i, \nu_i\right) = \sigma_{12}/\left(\sigma_{11}\sigma_{22}\right)^{1/2}$ is different from zero. It is important to note that when there is no endogeneity issue, standard estimation methods using a simple linear model instead of the IV model suffices.

The existence and the precision of the posterior β and Π: Under flat priors, the joint posterior density kernel of (Π, β) has a ridge at $\Pi = 0$ and the marginal posterior density kernel of Π has an asymptote at $\Pi = 0$. Further, in the case of exact identification (with as many instruments as explanatory endogenous variables), posterior of Π and β is not a proper density [Dréze, 1976, 1977, Kleibergen and Van Dijk, 1994a].

Kleibergen and Van Dijk [1998], Hoogerheide et al. [2007a] show that Jeffreys' prior eliminates some of these undesired properties. However, the precision in the posterior of β under Jeffreys priors also depends on the instrument strength. Kleibergen and Van Dijk [1998] show that the marginal posterior density of β under Jeffreys prior is a ratio of two t-densities, where the denominator depends on the explanatory power of the instruments. Intuitively, the posterior density is wider when one has weak instruments. In case of no endogeneity, this decrease in precision can be avoided by using a model without instruments, hence assessing the degree of endogeneity is important especially when instruments are rather weak in explaining the endogenous variable.

We next illustrate predictive model probabilities and how they can be employed to assess alternative IV model structures.

Predictive model probabilities and the Savage Dickey Density Ratio: Define two models M_0 and M_1, where M_0 is a nested model compared to M_1. Further, the nested model corresponds to M_1 with a parameter restriction: $\eta = 0$.

The posterior odds ratio, K_{01} for M_0 is the product of the Bayes factor and the prior odds ratio:

$$K_{01} = \frac{p(Y \mid M_0)}{p(Y \mid M_1)} \times \frac{p(M_0)}{p(M_1)}, \tag{7}$$

where Y is the observed data, and the prior model probabilities $(p(M_1), p(M_0)) \in (0,1) \times (0,1)$ and $p(M_1) + p(M_0) = 1$.

It is often difficult to compute K_{01} since the marginal likelihoods are given by the following integrals:

$$p(Y \mid M_0) = \int_{\theta_{-\eta}} \ell(\eta = 0, \theta_{-\eta}) \, p_0(\theta_{-\eta}) \, d(\theta_{-\eta}) \tag{8}$$

$$p(Y \mid M_1) = \int_{\theta_{-\eta}, \eta} \ell(\eta, \theta_{-\eta}) \, p(\eta, \theta_{-\eta}) \, d(\eta) \, d(\theta_{-\eta}), \tag{9}$$

where $\ell(\theta)$ is the likelihood function given parameters θ and $\theta_{-\eta}$ are the model parameters apart from η.

In order to calculate model probabilities, we make use of the Savage-Dickey Density Ratio (SDDR). Dickey [1971] shows that the Bayes factor can be calculated using a single model if the alternative models are nested and the prior densities satisfy the condition that the prior for $\theta_{-\eta}$ in the restricted model M_0 equals the conditional prior for $\theta_{-\eta}$ given $\eta = 0$ in the model M_1, i.e. $p_1(\theta_{-\eta} \mid \eta = 0) = p_0(\theta_{-\eta})$. In this case, (7) becomes:

$$K_{01} = \frac{p(\eta = 0 \mid Y, M_1)}{p(\eta = 0 \mid M_1)} \times \frac{p(M_0)}{p(M_1)}, \tag{10}$$

where $p(\eta \mid Y) = \int p(\eta, \theta_{-\eta} \mid Y) d\theta_{-\eta}$ and $p(\eta) = \int p(\eta, \theta_{-\eta}) d\theta_{-\eta}$[3].

One important consideration in model comparison is the effect of relatively uninformative priors. Choosing a prior $p(\eta, \theta_{-\eta})$ diffuse enough compared to $p(\theta_{-\eta})$, posterior odds ratio in (7) becomes larger independent of the data. Hence if we consider non-informative priors, the most restrictive model will typically be favored. This phenomenon is called Bartlett's paradox [Bartlett, 1957]. Specifically, the prior $p(\eta \mid \theta_{-\eta})$ must be proper for the Bayes factor to be well defined.

If one does not have informative priors for the model at hand, model probabilities can be computed using *predictive likelihoods* instead of the conventional *marginal likelihoods*. Eklund and Karlsson [2007] show that the sensitivity of model probabilities to the prior choice can be handled using predictive likelihoods and summarize alternative ways to calculate the predictive likelihood.

A predictive likelihood for the model M_1 is computed by splitting the dataset $Y = (y_1, \dots, y_N)$ into a training sample $y^* = (y_1, \dots, y_m)$ and a hold-out sample $\tilde{y} = (y_{m+1}, \dots, y_N)$. The predictive likelihood is given by:

$$p(\tilde{y} \mid y^*, M_1) = \int p(\tilde{y} \mid \theta_1, y^*, M_1) p(\theta_1 \mid y^*, M_1) d\theta_1, \tag{11}$$

where θ_1 are the model parameters for model M_1. Notice that equation (11) corresponds to the marginal likelihood for the hold-out sample $\tilde{y}$ times the exact posterior density after observing y^* as the prior. Therefore, model probabilities using this predictive measure does not provide the exact posterior model probability given the data, but rather a *predictive* model probability[4]. The exact posterior density $p(\theta_1 \mid y^*, M_1)$ is obtained by Bayes' rule:

$$p(\theta_1 \mid y^*, M_1) = \frac{p(y^* \mid \theta_1, M_1) p(\theta_1 \mid M_1)}{p(y^* \mid M_1)} \tag{12}$$

$$= \frac{p(y^* \mid \theta_1, M_1) p(\theta_1 \mid M_1)}{\int p(y^* \mid \theta_1, M_1) p(\theta_1 \mid M_1) d\theta_1}. \tag{13}$$

Substituting (13) into (11) leads to:

$$p(\tilde{y} \mid y^*, M_1) = \frac{\int p(\tilde{y} \mid \theta_1, y^*, M_1) p(y^* \mid \theta_1, M_1) p(\theta_1 \mid M_1) d\theta_1}{\int p(y^* \mid \theta_1, M_1) p(\theta_1 \mid M_1) d\theta_1}, \tag{14}$$

$$= \frac{\int p(y \mid \theta_1, M_1) p(\theta_1 \mid M_1) d\theta_1}{\int p(y^* \mid \theta_1, M_1) p(\theta_1 \mid M_1) d\theta_1}. \tag{15}$$

In case of predictive likelihoods, model probabilities are again calculated from the posterior odds ratio:

$$\frac{p(M_0 \mid y)}{p(M_1 \mid y)} = \frac{p(\tilde{y} \mid y^*, M_0)}{p(\tilde{y} \mid y^*, M_1)} \frac{p(M_0)}{p(M_1)}. \tag{16}$$

Combining the predictive likelihood formula in (16) and SDDR in (10), pos-

[3] As a generalization, Verdinelli and Wasserman [1995] show that K_{01} is equal to the Savage-Dickey density ratio in (10) times a correction factor when the prior condition fails.

[4] Equivalently, the term *predictive model weights* is used in the literature, see e.g. Eklund and Karlsson [2007].

terior odds ratio becomes:

$$K_{01} = \frac{p\left(M_0 \mid \tilde{y}, y^*\right)}{p\left(M_1 \mid \tilde{y}, y^*\right)} \tag{17}$$

$$= \frac{p(\eta = 0 \mid \tilde{y}, y^*, M_1)}{p(\eta = 0 \mid y^*, M_1)} \times \frac{p\left(M_0\right)}{p\left(M_1\right)}, \tag{18}$$

where $p(\eta \mid \tilde{y}, y^*) = \int p(\eta, \theta_{-\eta} \mid \tilde{y}, y^*) d\theta_{-\eta}$ and $p(\eta \mid y^*) = \int p(\eta, \theta_{-\eta} \mid y^*) d\theta_{-\eta}$ are the *exact* marginal posterior densities using the full data, and the training sample, respectively.

A final point concerning the calculation of predictive likelihoods is the size of the training sample. More stable results are achieved as the training sample size decreases, but the training sample should be large enough to provide a proper density given the originally diffuse/uninformative prior of parameters. Different training sample sizes have been proposed in the literature (see Gelfand and Dey [1994] for an overview of the forms of predictive likelihood under different training sample choices). More recently, Eklund and Karlsson [2007] suggests that the training sample should consist of around 20% of the data. Our choice of training sample size is based on this finding, however, we experiment with variations in the sample size considering training samples consisting of 5%, 10% and 25% of the observations.

Model averaging: Given the posterior odds ratio, it is possible to weight the evidence of alternative models using Bayesian Model Averaging (BMA). We consider the effect of model uncertainty on the estimation of the parameter β, as this parameter is the main focus in most cases. The information about β is summarized by the following posterior [Koop, 2003, ch. 11]:

$$p\left(\beta \mid Y\right) = p\left(\beta \mid Y, M_0\right) p\left(M_0 \mid Y\right) + p\left(\beta \mid Y, M_1\right) p\left(M_1 \mid Y\right). \tag{19}$$

Furthermore, functions of parameters, i.e. $g\left(\beta\right)$ in the IV model are estimated by:

$$E[g\left(\beta \mid Y\right)] = E[g\left(\beta \mid Y, M_0\right)] p\left(M_0 \mid Y\right) + E[g\left(\beta \mid Y, M_1\right)] p\left(M_1 \mid Y\right). \tag{20}$$

Hence both models under consideration should be estimated, and the inference on parameters is simply the weighted average of the results in both models. The weights in averaging the results are the predictive model probabilities.

Assessing the degree of endogeneity and instrument strength in IV models In this section we apply the predictive likelihood approach to weight the evidence of two alternative models: The first model under consideration is the IV model, that takes into account the endogeneity problem in the data at the expense of possible efficiency losses. The second model we consider is the *nested* model which assumes that there is no endogeneity problem in the data. This nested model corresponds the IV model in (1) and (2), with a parameter restriction $\rho = \sigma_{12} / \left(\sigma_{11}\sigma_{22}\right)^{1/2} = 0$.

Regarding the predictive model probabilities, we make use of the methods in Section 2, where calculation of predictive model probabilities is based on estimating the *general model*, which is the IV model in this case. Estimation of the *nested* model without endogeneity problem is straightforward: one can simply use the linear model in (1) to infer $p(\beta \mid Y, M_0)$.

The applications in this section focus on simulated data with different degrees of endogeneity and instrument strength, and Angrist and Krueger [1991] data on the income-education relationship. In the latter dataset, Hoogerheide et al. [2007a] show that this dataset suffers from weak instruments, and that some of the problems associated with weak instruments can be avoided using Jeffreys prior. Therefore in this section we consider the IV model under the Jeffreys prior.

For the IV model in (1) and (2), Jeffreys prior is given by:

$$p(\beta, \Pi, \Sigma) \propto \sigma_{11}^{(p-1)/2} |\Pi' X' X \Pi| \, |\Sigma|^{-\frac{p+3}{2}}. \tag{21}$$

The joint posterior is:

$$\begin{aligned} p(\beta, \Pi, \Sigma \mid Y, x) &\propto p(\Pi, \beta, \Sigma) \, p(Y \mid x, \beta, \Pi, \Sigma) & (22) \\ &\propto \sigma_{11}^{(p-1)/2} |\Pi' X' X \Pi| \, |\Sigma|^{-\frac{p+3}{2}} |\Sigma|^{-\frac{N}{2}} \\ &\quad \times \exp\left(-\frac{1}{2} tr\left(\Sigma^{-1} (Y - X\Pi B)' (Y - X\Pi B)\right)\right). & (23) \end{aligned}$$

Notice that conditional on a set of coefficients, the posterior density in (23) is rather standard, and it is straightforward to make inference relying on this density. However, the last term in (23) includes the product of parameters β and Π, and this issue is not easy to treat [Kleibergen and Van Dijk, 1994b, 1998, Hoogerheide et al., 2007b]. Therefore we explain the proposed posterior simulator in detail.

Posterior simulator: For the parameter restriction we consider, $\rho = 0$, the computation of the posterior odds ratio in (18) boils down to estimating the marginal posterior densities for the whole sample and the training sample, $p(\rho = 0 \mid y)$ and $p(\rho = 0 \mid y^*)$, respectively. However, to our knowledge, the marginal posterior density for the variance-covariance matrix for the IV model with Jeffreys' prior is not derived in the literature. Under flat priors, the conditional density for Π given β and the marginal density of Π under Jeffreys prior are derived in Kleibergen and Van Dijk [1994b] and Kleibergen and Van Dijk [1998]. Under flat priors, conditional density for β given Π and the marginal density of β are derived in Dréze [1976] and Dréze [1977]. These derivations rely on integrating out the variance-covariance matrix in the model.

Drawing from the posterior density is not straightforward because of the local non-identification issue: For $\Pi = 0$, parameter β drops out of the posterior density kernel in (23). We use the Metropolis Hastings algorithm (MH; Metropolis et al. [1953], Hastings [1970]) to get posterior draws from $p(\beta, \Pi, \sigma_{11}, \rho, \sigma_{22})$, using the candidate density:

$$p(\beta, \Pi, \Sigma) = p(\Pi) \, p(\beta \mid \Pi) \, p(\Sigma \mid \beta, \Pi), \tag{24}$$

where $p(\Pi)$ is a matrix-variate Student-t density with mean $\hat{\Pi} = (x'x)^{-1} x' y_2$ and the variance-covariance matrix $\hat{s}^2 (x'x)^{-1}$, where $\hat{s}^2$ is the OLS estimate for the error terms' variance (var(σ_{22})). The degrees of freedom is chosen low as 4, in order to cover all of the relevant range of Π values.

As a second step, β is drawn conditionally upon Π from its conditional posterior in the IV model under flat priors:

$$M_\nu y_1 = M_\nu y_2 \beta + \eta, \tag{25}$$

where $M_\nu = I - \nu(\nu'\nu)^{-1}\nu'\nu$ is the projection out of the space spanned by ν. Similar to $p(\Pi)$, the degrees of freedom for this density is kept small, to cover the relevant range (in the applications we use candidate densities with 4 degrees of freedom).

The *conditional* candidate density $p(\Sigma \mid \beta, \Pi)$ is the Inverted Wishart distribution with parameters matching the sample variance-covariance matrix of the error terms, and $N - k$ degrees of freedom. The true conditional posterior would result if we would choose $k = 0$, but we set $k = 10$ to cover a relatively wide range of values in the parameter space by the candidate draws.

In order to infer the posterior densities for ρ, we use posterior ρ draws from $p(\rho, \theta_{-\rho} \mid y)$. Then the marginal posterior density of ρ is approximated using non-parametric density estimation. This method is required since neither the conditional posterior nor the marginal posterior of ρ has a closed form solution[5].

In Section 3, we illustrate the performance of the predictive likelihood approach using simulated datasets with varying degrees of endogeneity and instrument strength. In Section 4, we implement this approach to assess the degree of endogeneity and instrument strength in the IV model for the income-education relationship in the US data.

3 Simulated data with varying degrees of endogeneity and instrument strength

We first check the performance of the model comparison approach using SDDR and predictive likelihoods on simulated data, and show that this comparison takes into account the necessity of instruments together with the degree to which instruments can explain the endogenous variable. The data are simulated according to the IV model in (1) and (2), with a single endogenous explanatory variable and a single instrument, where for $i = 1, ..., N$, $x_i \sim NID\,(0, 1)$, and the variance-covariance matrix for the error terms is $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$.

We consider 12 sets of simulated data with 1000 observations, different degrees of endogeneity and varying instrument strength. Regarding the degree of endogeneity, four specifications are considered: strong, moderate, weak endogeneity cases and finally the case of no endogeneity, where the IV model is clearly not necessary. The respective parameter settings are $\rho = 0.9$, $\rho = 0.5$, $\rho = 0.1$ and $\rho = 0$. In terms of the instrument strength, we consider strong,

[5] When the analytical expression for the conditional densities $p(\rho \mid \theta_{-\rho}, y)$ and $p(\rho \mid \theta_{-\rho}, y^*)$ are available, there are a number of methods to approximate the marginal posterior densities. See e.g. Silverman [1998] for kernel density estimation methods. Several methods such Chib's estimator [Chib, 1995], and the numerical integration method proposed by Verdinelli and Wasserman [1995] can be used.

weak and finally irrelevant instruments (where instruments have no explanatory power in explaining the possibly endogenous explanatory variable) with the corresponding parameter settings $\Pi = 1$, $\Pi = 0.1$ and $\Pi = 0$, respectively.

Next we calculate the predictive model probabilities using (18), where we set the prior model probabilities $p(M_0)$ and $p(M_1)$ to 0.5, i.e. the prior model probabilities favor both models equally. Regarding the training sample choice, we report results under three training samples, with 500, 250 and 125 observations. Furthermore, we eliminate the effect of random sampling by repeating each estimation with 20 different training samples. Notice that we consider cross-section data that do not have the natural ordering of time series data. For time series data, one can use the first subset of observations as training sample. We randomly choose the training sample, and consider 20 random choices to decrease the effect of a particular random sample selection[6]. Posterior results for all simulated datasets and training samples are reported in Table 1.

Table 1 about here

We first focus on the cases where there is the problem of endogeneity in the dataset, and instruments have some power as a proxy for the endogenous variable, i.e. cases where the degree of endogeneity and instrument strength are different from 'none'. Table 1 shows that given the same level of instrument strength, the predictive probability for the restricted model (in which β is estimated without making use of the instruments) is lower (higher) for simulated data with relatively strong (weak) degrees of endogeneity, regardless of the training sample choice. This result confirms that the unrestricted IV model has a higher predictive probability when there is a clear problem of endogeneity in the data.

On the other hand, given the same level of endogeneity, the model probability for the restricted model without instruments is lower (higher) for simulated data with relatively strong (weak) instruments. If the instruments are rather weak in explaining the endogenous explanatory variable, the gains from employing an IV model are questionable, and the IV model has a relatively lower model probability. We conclude that comparison of models using predictive likelihoods accounts for the trade-off between the correct model specification and precision considerations.

The cases with no endogeneity and irrelevant instruments should be considered separately, as these are the cases where the standard IV model assumptions do not hold. In case of no endogeneity, we expect the predictive model probabilities to favor the model without instruments. However, the assessment of endogeneity depends on the strength of the instruments. In the extreme case where instruments have no explanatory power, the IV model suffers from *local non-identification*: one cannot identify the effect of the explanatory variable (y_2) on the explained variable (y_1). Similarly, the degree of endogeneity is hard to assess as the test for endogeneity relies on the fit of the IV model in the first place. For this reason, in case of irrelevant instruments, the model probability of the IV model is not exactly zero.

[6]For a time series with independent observations, using a random training sample would not be fundamentally different from using a training sample of the first observations. In a certain sense, a cross section can be interpreted as a time series of independent observations. This makes it natural to apply the predictive likelihood concept in a cross section setting, using multiple selections for the training sample.

Model probabilities impled by predictive likelihoods are not only assessing instrument strength ($\Pi = 0$), but also the necessity of instruments. Intuitively, this comparison incorporates the degree to which instruments provide information for the data generating process, and efficiency losses from using instrumental variables when the degree of endogeneity is small. Further, when the level of endogeneity is not strong and the instruments are relatively weak, model probabilities are very close to 0.5. We therefore conclude that a *model choice* in these cases can be inappropriate, as none of the models are clearly favored.

We next estimate the posterior density of β for using BMA, using the predictive model probabilities achieved from training samples with 250 observations, reported in Table 1. Table 2 reports model averaging results for simulated data. M_0 denotes the nested model, M_1 denotes the IV model and 'average' denotes BMA estimates where average posterior means and variances for β are calculated from (20).

Table 2 about here

Model averaging leads to smaller variances than the IV estimate in the β parameter, except for the cases of strong instruments. For the cases where instruments have no explanatory power, the IV model leads to quite large posterior standard deviations, and mean β estimates are quite far from the true value. BMA in these cases leads to relatively more accurate results in posterior β density. However, as the predictive model probabilities are calculated *wrongly* by the IV model, the advantage of BMA is questionable. In the next section we apply the proposed method to Angrist and Krueger [1991] data on the income-education relationship.

4 Income-education relationship in Angrist and Krueger [1991] data

In this section, we apply the predictive likelihood approach for model comparison or combination to the Angrist and Krueger [1991] data on income and education. Angrist and Krueger [1991] data consists of men born in the US during the periods 1920-1929, 1930-1939 and 1940-1949, where the data for the first group is collected in 1970, and the data for the last two groups are collected in 1980. We use a subset of their data, consisting of men born during the period 1930-1939, including the data on weakly wages, number of completed years of education and instruments consisting of quarter of birth dummies. The data includes 51 states and 329.509 observations[7]. A summary of the data is given in Table 3.

Table 3 about here

The IV model for state i is:

$$\tilde{y}_i = \alpha_1 + \tilde{x}_i\beta + \sum_{t=1}^{9} D_{t,i}\delta_t + \tilde{\epsilon}_i \tag{26}$$

$$\tilde{x}_i = \alpha_2 + \sum_{q=2}^{4} D_{q,i}\Pi_q + \sum_{t=1}^{9} D_{t,i}\delta_t + \tilde{\nu}_i \tag{27}$$

[7]The source of the data is the 1980 Census, 5 percent public sample, also available from http://econ-www.mit.edu/faculty/angrist/data1/data/angkru1991.

where $\tilde{y}_i$ and $\tilde{x}_i$ are the natural logarithm of the weekly wage and the number of completed years of education for the person i in 1979, respectively.

In (26) and (27), $D_{t,i}$ for $t = \{1, \ldots, 9\}$ are the dummy variables for year of birth which take the value 1 if individual i was born in year $1929 + t$, and 0 otherwise. $D_{q,i}$ for $q = \{2, 3, 4\}$ are the quarter of birth dummy variables which take the value 1 if individual i was born in quarter q, and 0 otherwise. α_1 and α_2 are constants, and $\tilde{\epsilon}_i$ and $\tilde{\nu}_i$ are disturbances assumed to be normally distributed, and independent across individuals.

The model in (26) and (27) is similar to the model of Hoogerheide and Van Dijk [2006]. For simplicity, we do not consider interactions of year dummies and quarter of birth dummies as instruments. Furthermore, the model employed here does not include state dummies, as each state is analyzed separately. Similar to Hoogerheide and Van Dijk [2006], we simplify the IV model in (26) and (27) correcting for the constant terms and exogenous year of birth dummies. Using this simplification, the IV model becomes:

$$y_i = x_i\beta + \epsilon_i, \tag{28}$$
$$x_i = Z_i\Pi + \nu_i, \tag{29}$$

where y_i, x_i are the residuals from regressing the log weakly wage and years of education on a constant and year of birth dummies, respectively. Z_i is the 3×1 vector of instruments, obtained from regressing quarter of birth dummies on a constant and the year of birth dummies. ϵ_i and ν_i are the error terms that have a joint normal distribution, and are uncorrelated across individuals.

Posterior results for each state are reported in Table 4 and in Figures 1 and 2. The degree of instrument strength (indicated by posterior Π_1, Π_2, Π_3 densities) differs substantially across states, as reported in Hoogerheide and Van Dijk [2006]. A second source of heterogeneity across states is the degree of endogeneity (indicated by posterior ρ). For some states, such as Maine, Minnesota and Texas, 95% intervals for posterior ρ densities do not include point 0, while for the rest of the states posterior ρ values are centered around 0. Besides the finding of heterogeneity across states, we conclude that the use of instruments may not be necessary for most states.

Table 4 and Figures 1 and 2 about here

Posterior mean for the degree of endogeneity ρ in the IV model, and the predictive model probabilities for the model without instruments are given in Figure 3. For the predictive model probabilities, the training sample consists of 10% of the observations, prior model probabilities are chosen to be equal. Furthermore, the effect of the training sample choice is partially eliminated by averaging predictive model probabilities from 20 different random training samples. The results with 5% and 25% training sample sizes and a single training sample were similar, except for some states with very small number of observations, such as South Dakota.

Figure 3 about here

Model probabilities are quite close to 0.5 and do not indicate a specific model, except for a few states such as Texas and Tennessee. For Texas, model probabilities indicate that the IV model is necessary. For Tennessee on the other hand, we find strong evidence against the need for the IV model. We conclude that choosing one of the alternative models according to these probabilities can be quite inaccurate, and employ model averaging to infer the state-specific effects of income on education.

The *average* effects of education on income for the US states, i.e. the posterior distributions resulting from BMA, are summarized in Table 5. Model probabilities are achieved by using training sample with 10% of the observations, averaged over 20 repetitions. The main advantage of model averaging is the improved efficiency of the estimates. Standard deviations of posterior β draws are less than half of those achieved by the IV model only.

Table 5 about here

Regional patterns in income-education relationship - analysis of US divisions: We further analyze the income-education relationship in US divisions, reported in Table 3. We apply the IV model in (28) and (29) to 9 divisions for the Angrist and Krueger [1991] data according to the Census Bureau designated areas, as summarized in Table 3. The purpose of this analysis is to compare the results in terms of instrument strength with those of Hoogerheide and Van Dijk [2006], who show that quarter of birth dummies are strong instruments mainly in southern states. Furthermore, we document the effect of averaging the data within divisions or regions.

Table 6 reports posterior results of the IV model for US divisions. Similar to the state-specific results, census regions show heterogeneity both in terms of instrument strength and the degree of endogeneity. Posterior results for education effects on income are quite different across divisions. Especially for the West North Central division, the posterior standard deviation is quite high, indicating the relatively weak instruments in this division.

Table 6 about here

Figure 4 shows posterior mean ρ and model probabilities for the model without instruments for US divisions. The training sample consists of 25% of the observations[8]. Predictive model probability for the nested model without instruments is far from 0.5 only for two regions: East North Central Division, and West South Central Division. In East North Central division, the model without instruments is favored by model probabilities. Notice that the states within this region are quite heterogenous in terms of predictive model probabilities reported in Figure 3. The IV model is clearly necessary only for two states in this region, namely Arkansas (AR) and Texas (TX). Hence 'average' income-education relationship within this region is determined mainly by these two states. This problem is also seen in East North Central division. According to posterior model probabilities in Figure 3, this region consists of states where an IV model is clearly preferred, such as Minessota (MN) and South Dakota (SD), and also states where the IV model is not necessary, such as Iowa (IA). Hence the 'average' model probability for this region reported in Figure 4 is misleading.

[8]We experimented with the model using smaller training samples, and the results are quite insensitive to the training sample size.

Figure 4 about here

For the US data on the income-education relationship, we conclude that there is substantial heterogeneity in the income-education relationship across states. We document that differences between states are characterized by differing instrument strengths, as reported by Hoogerheide and Van Dijk [2006]. Despite this finding, our results show that the degree of endogeneity is also different across states and regions.

The dataset shows differing, and mostly weak power of quarter of birth in explaining education. This finding, in combination with the not so severe problem of endogeneity makes it hard to assess whether IV model should be preferred over a simpler and more parsimonious linear regression model without instruments. Hence we conclude that averaging over these alternative models is a reasonable way to deal with model uncertainty. Finally, we show that the combination of SDDR and predictive likelihoods provides a straightforward method to assess the degree of endogeneity and the relevance of the IV model.

5 Conclusion

We present a general framework for model comparison in the standard Bayesian treatment of IV models under non-informative priors, where the sensitivity of model probabilities to the prior choice can be avoided using predictive likelihoods. The method is applied to simulated datasets, and the US data on the income and educations. We show that this method can be used to assess a highly debated issue in IV models, namely the trade-off between choosing the correct model and precision.

The empirical contribution of this work is as follows: We show that the income-education relationship in the US states and divisions, using quarter of birth of individuals to form instruments, is subject to significant heterogeneity in terms of the degree of endogeneity in education levels, as well as the strength of instruments. For these data, we conclude that the precision losses from employing an IV model can be quite severe. We rather propose a model averaging approach to infer returns to education, for which the weights in model averaging take into account the degree of endogeneity and the instrument strength.

References

J. D. Angrist, G.W. Imbens, and D. B. Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434), 1996.

Joshua D Angrist and Alan B Krueger. Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4): 979–1014, November 1991.

M. S. Bartlett. A comment on dv lindley's statistical paradox. *Biometrika*, 44 (3–4):533, 1957.

J M Bates and C W J Granger. The combination of forecasts. *OR*, 20(4): 451–468, 1969.

J. Bound, D. A. Jaeger, and R. M. Baker. 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(430):443–450, 1995.

R. J. Bowden and D. A. Turkington. *Instrumental variables*. Cambridge Univ Press, 1990.

S. Chib. Marginal likelihood from the gibbs output. *Journal of the American Statistical Association*, 90(432), 1995.

M. Clyde and E. I. George. Model uncertainty. *Statistical Science*, 19(1):81–94, 2004.

T. G. Conley, C. B. Hansen, and P. E. Rossi. Plausibly exogenous. *unpublished working paper*, 2008.

J. M. Dickey. The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics*, 42(1):204–223, 1971.

J. H. Dréze. Bayesian limited information analysis of the simultaneous equations model. *Econometrica: Journal of the Econometric Society*, 44(5):1045–1075, 1976.

J. H. Dréze. Bayesian regression analysis using poly-t densities. *Journal of Econometrics*, 6(3):329–354, 1977.

J. Eklund and S. Karlsson. Forecast combination and model averaging using predictive measures. *Econometric Reviews*, 26(2):329–363, 2007.

A. E. Gelfand and D. K. Dey. Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(3):501–514, 1994.

J. Geweke. *Complete and incomplete econometric models*. Princeton Univ Pr, 2010.

J. Geweke and G. Amisano. Comparing and evaluating bayesian predictive distributions of asset returns. *International Journal of Forecasting*, 2010.

A. S. Goldberger. Structural equation methods in the social sciences. *Econometrica: Journal of the Econometric Society*, 40(6):979–1001, 1972.

W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

L. Hoogerheide, F. Kleibergen, and H. K. Van Dijk. Natural conjugate priors for the instrumental variables regression model applied to the angrist-krueger data. *Journal of Econometrics*, 138(1):63–103, 2007a.

L. F. Hoogerheide and H. K. Van Dijk. A reconsideration of the angrist-krueger analysis on returns to education. *Econometric Institute Report*, 2006.

L. F. Hoogerheide, J. F. Kaashoek, and H. K. Van Dijk. On the shape of posterior densities and credible sets in instrumental variable regression models with reduced rank: an application of flexible sampling methods using neural networks. *Journal of Econometrics*, 139(1):154–180, 2007b.

F. Kleibergen and H. K. Van Dijk. Bayesian analysis of simultaneous equation models using noninformative priors. *Tinbergen Institute Discussion Paper*, (94-134), 1994a.

F. Kleibergen and H. K. Van Dijk. On the shape of the likelihood/posterior in cointegration models. *Econometric Theory*, 10(3):514–551, 1994b.

F. Kleibergen and H. K. Van Dijk. Bayesian simultaneous equations analysis using reduced rank structures. *Econometric Theory*, 14(06):701–743, 1998.

G. Koop. *Bayesian econometrics*. Wiley, New York, 2003.

P. W. Laud and J. G. Ibrahim. Predictive model selection. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):247–262, 1995.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087, 1953.

C. Min and A. Zellner. Bayesian and non-bayesian methods for combining models and forecasts with applications to forecasting international growth rates. *Journal of Econometrics*, 56(1-2):89–118, 1993.

J. D. Sargan. The estimation of economic relationships using instrumental variables. *Econometrica: Journal of the Econometric Society*, 26(3):393–415, 1958.

B. W. Silverman. *Density estimation for statistics and data analysis*. Chapman & Hall/CRC, 1998.

I. Verdinelli and L. Wasserman. Computing bayes factors using a generalization of the savage-dickey density ratio. *Journal of the American Statistical Association*, 90(430), 1995.

Tables and Figures

Table 1: Simulated data: Model probabilities for the restricted model in which posterior mean of β is the ordinary least-squares (OLS) estimator in (1))

degree of endogeneity	degree of instrument strength	training sample size 500	250	125
strong	strong	0^*	0^*	0^*
strong	weak	0^*	0^*	0^*
strong	none	0.59	0.64	0.64
moderate	strong	0^*	0^*	0
moderate	weak	0.33	0.33	0.30
moderate	none	0.52	0.55	0.55
weak	strong	0.00	0.00	0.00
weak	weak	0.57	0.66	0.68
weak	none	0.29	0.26	0.24
none	strong	0.47	0.57	0.62
none	weak	0.64	0.71	0.78
none	none	0.44	0.46	0.45

* denotes the cases for which both densities in (18) (i.e. $p(\rho = 0 \mid \tilde{y}, y^*)$ and $p(\rho = 0 \mid y^*)$) are estimated as zero by the kernel density method, at least in one of the subsamples considered. If we generated a large enough set of random draws to estimate these densities as non-zero, which would typically have to be huge, then the predictive probability of the restricted model would be estimated as (almost) 0, since $p(\rho = 0 \mid \tilde{y}, y^*)$ would be much smaller than $p(\rho = 0 \mid y^*)$). Roughly stated, in the current setting the zero value of the former is much more zero than the latter. We set the predictive probability of M_0 to 0 for these cases.

Table 2: Simulated data: posterior β for the IV model and the nested model without instruments, and BMA

Degree of endogeneity	Degree of identification	Model	β Mean	Std. Dev.
strong	strong	M_0	1.47	0.00
		average	1.05	0.02
		M_1	1.05	0.02
strong	weak	M_0	1.89	0.00
		average	0.89	9.36
		M_1	0.89	9.36
strong	none	M_0	1.88	0.00
		average	1.97	1.87
		M_1	2.16	5.55
moderate	strong	M_0	1.24	0.00
		average	0.98	0.03
		M_1	0.98	0.03
moderate	weak	M_0	1.47	0.00
		average	1.08	0.20
		M_1	0.90	0.29
moderate	none	M_0	1.48	0.00
		average	1.54	4.11
		M_1	1.62	9.12
weak	strong	M_0	1.08	0.00
		average	0.99	0.03
		M_1	0.99	0.03
weak	weak	M_0	1.10	0.00
		average	1.18	0.22
		M_1	1.37	0.72
weak	none	M_0	1.08	0.00
		average	-1.78	24.86
		M_1	-2.83	33.95
none	strong	M_0	1.02	0.00
		average	1.01	0.02
		M_1	0.99	0.04
none	weak	M_0	1.03	0.00
		average	1.08	0.09
		M_1	1.21	0.38
none	none	M_0	1.01	0.00
		average	0.88	22.94
		M_1	0.78	41.07

Note: The table reports posterior means and standard deviations for posterior β draws under the IV model, the model without instruments and the average effect as a combination of both models, using weights achieved by the training sample with 250 observations, reported in Table 1.

Table 3: US regions and divisions

Division	States	Number of observations
	Northeast Region	
1. New England	Connecticut (CT), Maine (ME), Massachusetts (MA), New Hampshire (NH), Rhode Island and Vermont (RI)	20120
2. Middle Atlantic	New Jersey (NJ), New York (NY) and Pennsylvania (PA)	64364
	Midwest Region	
3. East North Central	Illinois (IL), Indiana (IN), Michigan (MI), Ohio (OH) and Wisconsin (WI)	67047
4. West North Central	Iowa (IA), Kansas (KS), Minnesota (MN), Missouri (MO), Nebraska (NE), North Dakota (ND) and South Dakota (SD)	35220
	South Region	
5. South Atlantic	Delaware (DE), District of Columbia (DC), Florida (FL), Georgia (GA), Maryland (MD), North Carolina (NC), South Carolina (SC), Virginia (VA) and West Virginia (WV)	48072
6. East South Central	Alabama (AL), Kentucky (KY), Mississippi (MS) and Tennessee (TN)	31668
7. West South Central	Arkansas (AR), Louisiana (LA), Oklahoma (OK) and Texas (TX)	34651
	West Region	
8. Mountain	Arizona (AZ), Colorado (CO), Idaho (ID), Montana (MT), Nevada (NV), New Mexico (NM), Utah (UT) and Wyoming (WY)	11228
9. Pacific	Alaska (AK), California (CA), Hawaii (HI), Oregon (OR) and Washington (WA)	17139

Note: The table reports US states included in the dataset, with the respective regions, divisions and the number of observations.

Table 4: Income-education effects in US states: Posterior results for parameters

	β	Π_2	Π_3	Π_4	ρ		β	Π_2	Π_3	Π_4	ρ
AK	0.1	1.54	1.5	-0.3	-0.2	MT	0.02	0.01	0.09	0.39	0.08
	(0.12)	(0.76)	(0.77)	(0.87)	(0.33)		(0.10)	(0.13)	(0.14)	(0.14)	(0.34)
AL	0.1	0.03	0.33	0.3	-0.2	NY	0.09	-0.04	0.13	0.24	-0.06
	(0.05)	(0.08)	(0.07)	(0.08)	(0.21)		(0.05)	(0.06)	(0.06)	(0.07)	(0.25)
AR	0.1	-0.18	0.1	0.4	-0.3	ND	0.13	-0.48	-0.19	-0.08	-0.25
	(0.04)	(0.09)	(0.08)	(0.08)	(0.17)		(0.09)	(0.13)	(0.14)	(0.15)	(0.30)
AZ	0.1	0.65	-0	-0.3	-0	NE	0	-0.13	-0.15	-0.12	0.20
	(0.05)	(0.21)	(0.20)	(0.22)	(0.27)		(0.16)	(0.09)	(0.10)	(0.08)	(0.44)
CA	0	0.26	0.22	0.1	0.08	NH	0.1	-0.14	0.1	0.2	-0.08
	(0.06)	(0.05)	(0.05)	(0.05)	(0.26)		(0.10)	(0.19)	(0.18)	(0.2)	(0.39)
CO	0.1	0.31	0.44	0.4	-0.1	NJ	0.13	0.06	0.07	0.21	-0.24
	(0.06)	(0.11)	(0.10)	(0.11)	(0.25)		(0.08)	(0.06)	(0.06)	(0.08)	(0.31)
CT	0.1	0.3	0.12	0.1	0.1	NM	0.04	0.23	0.11	0.39	0.06
	(0.09)	(0.09)	(0.10)	(0.10)	(0.37)		(0.10)	(0.18)	(0.19)	(0.17)	(0.41)
DC	-0	-0.44	-0.4	-0.5	0.43	NV	0	-0.03	0.07	0.85	0.17
	(0.09)	(0.16)	(0.17)	(0.16)	(0.27)		(0.11)	(0.38)	(0.37)	(0.35)	(0.35)
DE	0.1	0.56	0.69	0.2	-0.2	NV	0.09	0.1	0.05	-0.05	-0.05
	(0.08)	(0.22)	(0.22)	(0.23)	(0.31)		(0.07)	(0.04)	(0.04)	(0.04)	(0.30)
FL	0.2	0.32	0.24	0.3	-0.3	OH	0.11	-0.04	0.05	0.05	-0.17
	(0.07)	(0.10)	(0.09)	(0.09)	(0.26)		(0.10)	(0.06)	(0.05)	(0.05)	(0.37)
GA	0.2	-0.25	0.05	0	-0.4	OK	0.01	-0.04	0.16	0.23	0.21
	(0.05)	(0.05)	(0.06)	(0.06)	(0.22)		(0.07)	(0.08)	(0.07)	(0.07)	(0.28)
HI	0.1	0.09	1.55	0.8	-0.1	OR	0.05	0.12	0.11	0.06	0.01
	(0.07)	(0.41)	(0.39)	(0.35)	(0.28)		(0.15)	(0.12)	(0.12)	(0.12)	(0.47)
IA	0	-0.04	-0	0.1	0.11	PA	0.15	0.02	0.01	0.05	-0.33
	(0.10)	(0.06)	(0.06)	(0.06)	(0.36)		(0.07)	(0.03)	(0.03)	(0.03)	(0.26)
ID	0.1	0.16	-0	0.1	0.02	RI	0.07	-0.39	0.11	0.12	0.04
	(0.13)	(0.16)	(0.17)	(0.14)	(0.45)		(0.07)	(0.15)	(0.17)	(0.18)	(0.3)
IL	0	0.07	-0.1	0.1	0.24	SC	0.17	-0.11	-0.05	0.31	-0.39
	(0.08)	(0.03)	(0.04)	(0.04)	(0.29)		(0.07)	(0.09)	(0.07)	(0.08)	(0.25)
IN	0.2	0.04	0.08	0	-0.3	SD	0.17	0.35	0.3	0.56	-0.42
	(0.15)	(0.05)	(0.05)	(0.05)	(0.44)		(0.08)	(0.14)	(0.13)	(0.15)	(0.24)
KS	0.1	0.3	0.34	0.2	-0.2	TN	0.06	-0.06	0.19	0.47	0.07
	(0.07)	(0.07)	(0.08)	(0.08)	(0.26)		(0.03)	(0.08)	(0.07)	(0.07)	(0.17)
KY	0.1	0.08	0.35	0.6	-0.2	TX	0.16	-0.04	0.23	0.26	-0.43
	(0.03)	(0.07)	(0.07)	(0.07)	(0.17)		(0.06)	(0.05)	(0.05)	(0.07)	(0.19)
LA	0.1	0.1	0.26	0.3	-0.2	UT	0.11	-0.02	-0.15	-0.25	-0.20
	(0.10)	(0.09)	(0.09)	(0.10)	(0.36)		(0.13)	(0.14)	(0.15)	(0.16)	(0.45)
MA	0.1	0.13	0.17	0.3	-0.2	VA	0.06	0.08	0.3	0.32	0.05
	(0.07)	(0.06)	(0.06)	(0.07)	(0.32)		(0.07)	(0.08)	(0.09)	(0.08)	(0.31)
MD	0	0.38	0.43	0.3	0.17	VT	0.09	0.22	0.47	0.33	-0.06
	(0.06)	(0.10)	(0.10)	(0.09)	(0.24)		(0.10)	(0.19)	(0.22)	(0.21)	(0.39)
ME	0.3	0.01	0.28	0	-0.6	WA	0.13	0.14	0.12	0	-0.22
	(0.13)	(0.09)	(0.10)	(0.11)	(0.26)		(0.15)	(0.09)	(0.08)	(0.11)	(0.47)
MI	0.1	0.15	0.11	0.1	-0.2	WI	0.07	0.21	0.01	0.1	-0.01
	(0.07)	(0.03)	(0.04)	(0.04)	(0.28)		(0.08)	(0.06)	(0.08)	(0.07)	(0.28)
MN	-0.1	-0.2	-0.2	-0.1	0.55	WV	0.05	-0.04	0.09	0.27	0.04
	(0.10)	(0.06)	(0.06)	(0.05)	(0.21)		(0.06)	(0.09)	(0.07)	(0.07)	(0.26)
MO	0.1	-0.08	0.09	0	-0	WY	0.03	0.14	0.38	-0.13	0.09
	(0.08)	(0.06)	(0.05)	(0.05)	(0.30)		(0.11)	(0.23)	(0.22)	(0.25)	(0.38)
MS	0.1	0.07	0.22	0.3	-0.2						
	(0.08)	(0.08)	(0.09)	(0.08)	(0.36)						

Note: The table reports posterior means for the parameters and posterior standard deviations (in parentheses) for each state. Π_2, Π_3 and Π_4 are the coefficients for the 2nd, 3rd and 4th quarter of birth dummies, respectively. Posterior results are achieved by 30000 draws (3000 burn-in).

Table 5: Income-education effects in US states: *average* effects of education on income in US states

State	Mean	Std. Dev.	State	Mean	Std. Dev.	State	Mean	Std. Dev.
AL	0.11	0.03	LA	0.09	0.04	OH	0.08	0.04
AZ	0.11	0.03	ME	0.09	0.03	OK	0.04	0.03
AR	0.07	0.02	MD	0.05	0.03	OR	0.05	0.10
CA	0.05	0.01	MA	0.21	0.09	PA	0.11	0.03
CO	0.07	0.02	MI	0.08	0.03	RI	0.07	0.03
CT	0.06	0.04	MN	-0.06	0.08	SC	0.12	0.03
DE	0.02	0.05	MO	0.07	0.03	SD	0.16	0.07
DC	0.10	0.04	MS	0.09	0.04	TN	0.07	0.01
FL	0.13	0.05	MT	0.04	0.04	TX	0.16	0.06
GA	0.12	0.02	NC	0.08	0.02	UT	0.09	0.07
HI	0.08	0.04	NC	0.09	0.04	VT	0.06	0.03
ID	0.05	0.06	NE	0.03	0.09	VA	0.08	0.04
IL	0.05	0.08	NH	0.09	0.04	WA	0.10	0.09
IN	0.04	0.03	NJ	0.09	0.03	WV	0.06	0.03
IA	0.15	0.12	NM	0.05	0.05	WI	0.05	0.03
KS	0.08	0.03	NV	0.03	0.06	WY	0.04	0.06
KY	0.07	0.01	NY	0.08	0.03			

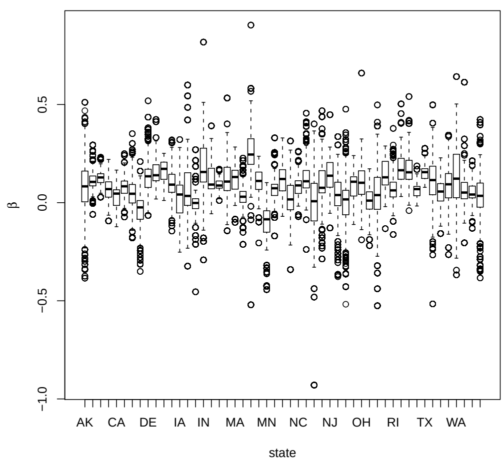
Note: The table reports means and standard deviations of effect of education on income, resulting from BMA, for the US states, achieved by 30000 draws (3000 burn-in).

Table 6: Income-education effects in US divisions: parameter estimates

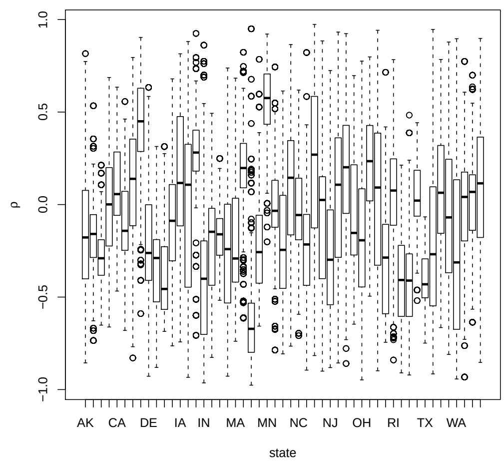
	β	Π_2	Π_3	Π_4	ρ
New England Division	0.11	0.09	0.17	0.21	-0.16
	(0.05)	(0.04)	(0.04)	(0.04)	(0.23)
Middle Atlantic Division	0.07	0.07	0.03	0.03	0.03
	(0.07)	(0.02)	(0.02)	(0.03)	(0.31)
East North Central Division	-0.03	0.07	0.02	0.08	0.36
	(0.08)	(0.02)	(0.02)	(0.02)	(0.25)
West North Central Division	0.02	-0.06	0.01	0.02	0.15
	(0.13)	(0.04)	(0.04)	(0.04)	(0.40)
South Atlantic Division	0.11	-0.01	0.14	0.22	-0.18
	(0.03)	(0.03)	(0.03)	(0.03)	(0.16)
East South Central Division	0.09	0.03	0.27	0.41	-0.13
	(0.02)	(0.04)	(0.04)	(0.04)	(0.12)
West South Central Division	0.12	-0.04	0.20	0.30	-0.29
	(0.02)	(0.04)	(0.04)	(0.03)	(0.11)
Mountain Division	0.01	0.20	0.14	0.18	0.21
	(0.08)	(0.05)	(0.06)	(0.06)	(0.30)
Pacific Division	0.04	0.23	0.21	0.11	0.08
	(0.05)	(0.04)	(0.04)	(0.04)	(0.23)

Note: The table reports posterior means for 9 US divisions, achieved by 30000 draws (3000 burn-in). Estimated standard errors are reported in parentheses.

Figure 1: Income-education effects in US states: Boxplots for income effects and degree of endogeneity

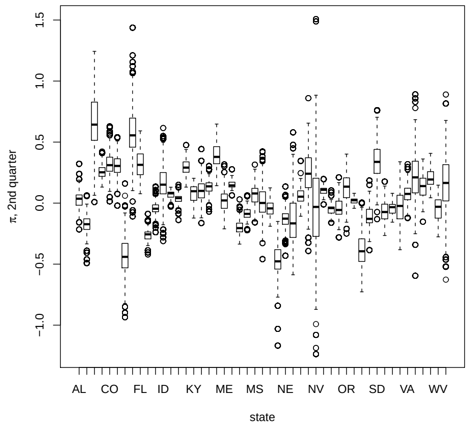


Posterior β values

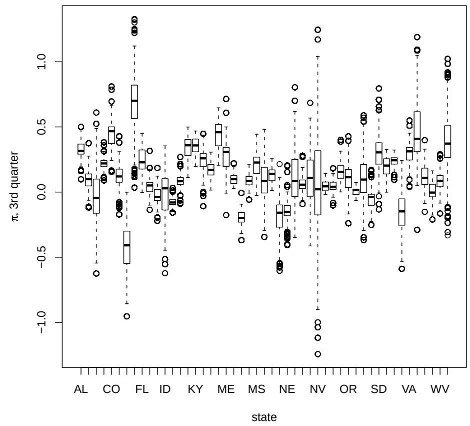


Posterior ρ values

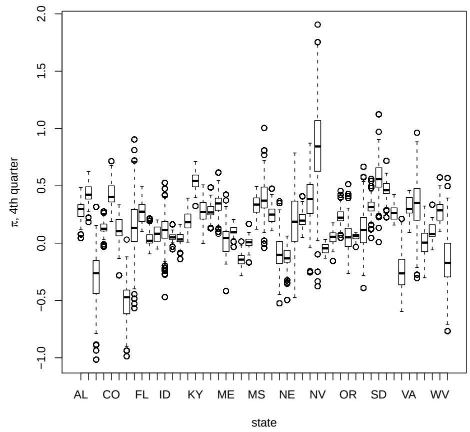
Figure 2: Income-education effects in US states: Boxplots for the quarter of birth effects on education



Posterior Π_2 values

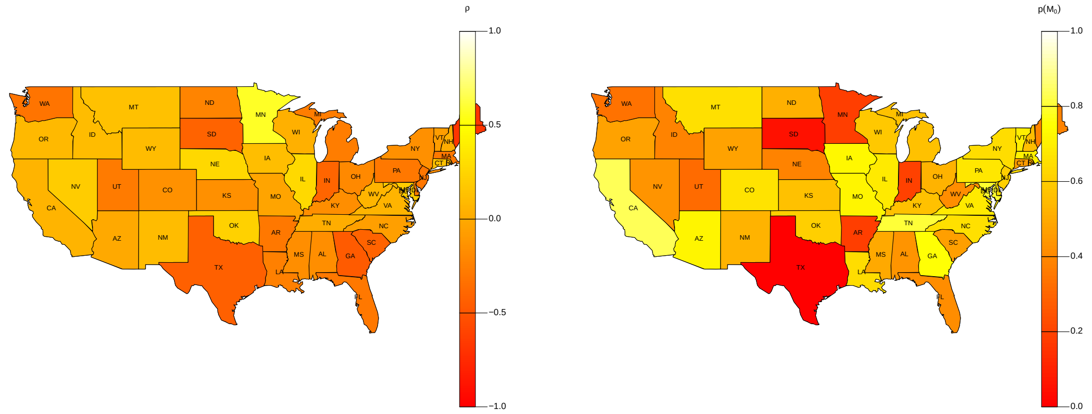


Posterior Π_3 values



Posterior Π_4 values

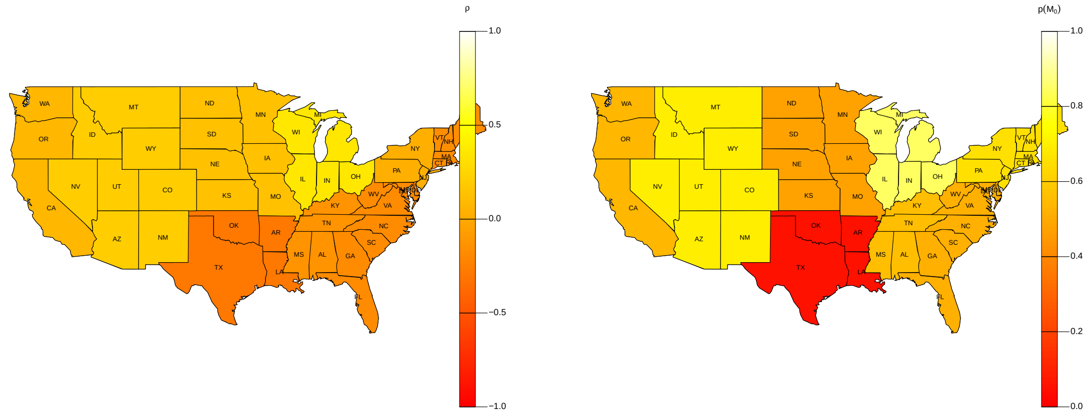
Figure 3: Degree of endogeneity in the US states and predictive model probabilities



Posterior mean ρ for US states

Model probabilities for the model without instruments (M_0) for US states

Figure 4: Degree of endogeneity in the US regions and predictive model probabilities



Posterior mean ρ for US divisions

Model probabilities for the model without instruments (M_0) for US divisions