Bayesian Treatment of Model Uncertainty Under Endogeneity

by

Anne Miloschewski

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Statistics

University of Warwick, Department of Statistics
November 2019

Contents

List of Figures 3

List of Tables 4

Abbreviations 13

1 Introduction 14

2 Dealing with Model Uncertainty 16

3 Literature Review - Model Uncertainty and the Standard Linear Regression Model 21

3.1 Priors for the Model Parameters . . . 22

3.2 Priors on Model Space . . . 24

4 Introducing Endogeneity 28

4.1 The Instrumental Variables Regression Model . . . 29

4.2 Effects of Ignoring Endogeneity Within the Standard Linear Regression BMA-Framework . . . 31

4.3 Existing Approaches for Incorporating Model Uncertainty in the IV Model Case . . . 36

5 The Karl and Lenkoski (2012) Approach and Some Extensions 41

5.1 Derivation of the Full Conditionals and the Conditional Bayes Factors . . . 41

5.2 Extensions to the Prior Structure . . . 46

5.3 Savage-Dickey Density Ratio for Endogeneity . . . 58

5.3.1 Marginal Prior Simulations . . . 63

5.3.2 Marginal Posterior Simulations . . . 66

5.4 Empirical Application - Drivers of Corruption . . . 72

6 An Alternative Modelling Approach 82

6.1 Bayesian Methods for IV-models . . . 82

6.2 The Hoogerheide et al. (2007) Approach and some Extensions . . . 86

7 Conclusion 112

Bibliography 114

A Some Distributions **118**

B Additional Results: Simulation Study - Ignoring Endogeneity **119**

C Additional Results: Simulation Study - Different Prior Settings **121**

D Additional Results: SD-Density Ratio for Endogeneity **137**

E Additional Results: Empirical Example **142**

List of Figures

1 Influence of g on marginal prior of β 94

2 Influence of μ_0 on marginal prior of β. 94

3 Influence of $\delta_0 = \binom{\delta_{11,0}}{\delta_{21,0}}$ on marginal prior of β. 95

4 Influence of $\delta_0 = \binom{\delta_{11,0}}{0.5}$ on marginal prior of β. 95

5 Influence of $\delta_0 = \binom{0.5}{\delta_{21,0}}$ on marginal prior of β. 96

6 Influence of $\omega_{21,0}$ on marginal prior of β. 96

7 Influence of ω_{11} on marginal prior of β. 97

8 Influence of ω_{22} on marginal prior of β. 97

9 Marginal posterior of β with informative prior under $\beta_0 = 0$. 100

10 Marginal posterior of β with uninformative prior under $\beta_0 = 0$. 101

11 Marginal posterior of β with informative prior under $\beta_0 = 0.5$. 101

12 Marginal posterior of β with uninformative prior under $\beta_0 = 0.5$. 102

13 Marginal posterior of β with informative prior under $\beta_0 = -0.5$. 102

14 Marginal posterior of β with uninformative prior under $\beta_0 = -0.5$. 103

List of Tables

1 Simulation study ignoring endogeneity - first version. 33

2 Simulation study ignoring endogeneity - second version. 34

3 Simulation study ignoring endogeneity - third version. 35

4 Simulation study different priors - PIPs and Means correctly specified system. 50

5 Simulation study different priors - quantiles correctly specified system. . . 51

6 Simulation study different priors - covariances correctly specified system. 51

7 Simulation study different priors - PIPs and Means incorrectly specified system. 52

8 Simulation study different priors - quantiles incorrectly specified system. . 53

9 Simulation study different priors - covariances incorrectly specified system. 53

10 Simulation study different priors - PIPs and Means correctly specified system and scaled data. 54

11 Simulation study different priors - quantiles correctly specified system and scaled data. 55

12 Simulation study different priors - covariances correctly specified system and scaled data. 55

13 Simulation study different priors - PIPs and Means incorrectly specified system and scaled data. 56

14 Simulation study different priors - quantiles incorrectly specified system and scaled data. 57

15 Simulation study different priors - covariances incorrectly specified system and scaled data. 57

16 Non-parametric density estimates at 0 of marginal prior of σ_{21} - rule of thumb bandwidth. 61

17 Non-parametric density estimates at 0 of marginal prior of σ_{21} - cv-criterion bandwidth. 61

18 Density estimates at 0 of marginal prior σ_{21} - via algorithm. 63

19 Density estimates at 0 of marginal prior of σ_{2k} where $\Sigma \sim IW_k(I_k, \nu)$ - via algorithm. 64

20 Density estimates at 0 of marginal prior of σ_{2k} where $\Sigma \sim IW_k(Q_{k,a}, \nu)$ - via algorithm. 65

21 Density estimates at 0 of marginal prior of σ_{2k} where $\Sigma \sim IW_k(Q_{k,b}, \nu)$ - via algorithm. 65

22 Density estimates at 0 of marginal prior of σ_{2k} where $\Sigma \sim IW_k(Q_{k,c}, \nu)$ - via algorithm. 66
23 Density value of the marginal posterior of σ_{13} at zero and SD-ratio - with $\rho_{12} = 0.7$ and $\rho_{13} = 0.6$ 68
24 Density value of the marginal posterior of σ_{13} at zero and SD-ratio - with $\rho_{12} = 0.7$ and $\rho_{13} = 0.3$ 68
25 Density value of the marginal posterior of σ_{13} at zero and SD-ratio - with $\rho_{12} = 0.7$ and $\rho_{13} = 0.3$, $N = 500$ 69
26 Density value of the marginal posterior of σ_{13} at zero and SD-ratio - with $\rho_{12} = 0.4$ and $\rho_{13} = 0$ 69
27 Density value of the marginal posterior of σ_{13} at zero and SD-ratio - with $\rho_{12} = 0.4$ and $\rho_{13} = 0$, $N = 500$ 70
28 True and estimated posterior mean of the covariance matrices from first run of testing-down setup. 70
29 Posterior estimation results testing down example - first run 71
30 True and estimated covariance matrices from second run of testing-down setup. 71
31 Posterior estimation results testing down example - second run 72
32 Posterior estimation results under $\Sigma \sim IW(I_{m+1}, m+2)$ - empirical example 77
33 CV- results under $\Sigma \sim IW(I_{m+1}, m+2)$ - empirical example 78
34 Posterior estimation results under $\Sigma \sim IW(Q_{m+1,c}, m+2)$ - empirical example 80
35 CV-results under $\Sigma \sim IW(Q_{m+1,c}, m+2)$ - empirical example 80
A.1 Estimation results for sample sizes $N = 100$ and $N = 10000$ for version one - ignoring endogeneity. 119
A.2 Estimation results for sample sizes $N = 100$ and $N = 10000$ for version two - ignoring endogeneity. 120
A.3 Estimation results for sample sizes $N = 100$ and $N = 10000$ for version three - ignoring endogeneity. 120
A.4 Simulation study different priors - PIPs and Means for correctly specified system. (full table) 121
A.5 Simulation study different priors - quantiles for correctly specified system. (full table) 122

A.6 Simulation study different priors - PIPs and Means for incorrectly specified system. (full table) . . . 122
A.7 Simulation study different priors - quantiles for incorrectly specified system. (full table) . . . 123
A.8 Simulation study different priors - PIPs and Means for correctly specified system and scaled data. (full table) . . . 123
A.9 Simulation study different priors - quantiles for correctly specified system and scaled data. (full table) . . . 124
A.10 Simulation study different priors - PIPs and Means for incorrectly specified system and scaled data. (full table) . . . 124
A.11 Simulation study different priors - quantiles for incorrectly specified system and scaled data. (full table) . . . 125
A.12 Simulation study different priors - PIPs and Medians for correctly specified system. . . . 126
A.13 Simulation study different priors - PIPs and Medians for incorrectly specified system. . . . 126
A.14 Simulation study different priors - PIPs and Medians for correctly specified system and scaled data. . . . 127
A.15 Simulation study different priors - PIPs and Medians for incorrectly specified system and scaled data. . . . 127
A.16 Simulation study different priors - PIPs and Means for correctly specified system and $N = 500$. (full table) . . . 128
A.17 Simulation study different priors - quantiles for correctly specified system and $N = 500$. (full table) . . . 129
A.18 Simulation study different priors - covariances for correctly specified system and $N = 500$. (full table) . . . 129
A.19 Simulation study different priors - PIPs and Means for incorrectly specified system and $N = 500$. (full table) . . . 129
A.20 Simulation study different priors - quantiles for incorrectly specified system and $N = 500$. (full table) . . . 130
A.21 Simulation study different priors - covariances for incorrectly specified system and $N = 500$. (full table) . . . 130
A.22 Simulation study different priors - PIPs and Means for correctly specified system and scaled data and $N = 500$. (full table) . . . 131

A.23 Simulation study different priors - quantiles for correctly specified system and scaled data and $N = 500$. (full table) 131
A.24 Simulation study different priors - covariances for correctly specified system and scaled data and $N = 500$. (full table) 132
A.25 Simulation study different priors - PIPs and Means for incorrectly specified system and scaled data and $N = 500$. (full table) 132
A.26 Simulation study different priors - quantiles for incorrectly specified system and scaled data and $N = 500$. (full table) 133
A.27 Simulation study different priors - covariances for incorrectly specified system and scaled data and $N = 500$. (full table) 133
A.28 Simulation study different priors - PIPs and Means for correctly specified system and $\nu = 8$. (full table) 134
A.29 Simulation study different priors - quantiles for correctly specified system and $\nu = 8$. (full table) 135
A.30 Simulation study different priors - covariances for correctly specified system and $\nu = 8$. (full table) 135
A.31 Simulation study different priors - PIPs and Means for incorrectly specified system and $\nu = 8$. (full table) 135
A.32 Simulation study different priors - quantiles for incorrectly specified system and $\nu = 8$. (full table) 136
A.33 Simulation study different priors - covariances for incorrectly specified system and $\nu = 8$. (full table) 136
A.34 Density value at 0 for the marginal posterior density of σ_{12} from a two-equation system with $N = 100$ observations and I-I prior. 137
A.35 SD-ratio value for a two-equation system with $N = 100$ observations and I-I prior. 137
A.36 Density value at 0 for the marginal posterior density of σ_{12} for a two-equation system with $N = 100$ observations and g-g prior. 138
A.37 SD-ratio value for a two-equation system with $N = 100$ observations and g-g prior. 138
A.38 Density value at 0 for the marginal posterior density of σ_{12} for a two-equation system with $N = 100$ observations and gprvar-g prior. 138
A.39 SD-ratio value for a two-equation system with $N = 100$ observations and gprvar-g prior. 138

A.40 Density value at 0 for the marginal posterior density of σ_{12} for a two-equation system with $N = 1000$ observations and I-I prior. 138
A.41 SD-ratio value for a two-equation system with $N = 1000$ observations and I-I prior. 139
A.42 Density value at 0 for the marginal posterior density of σ_{12} from a two-equation system with $N = 1000$ observations and g-g prior. 139
A.43 SD-ratio value for the marginal posterior density of σ_{12} from a two-equation system with $N = 1000$ observations and g-g prior. 139
A.44 Density value at 0 for the marginal posterior density of σ_{12} from a two-equation system with $N = 1000$ observations and gprvar-g prior. 139
A.45 SD-ratio value for the marginal posterior density of σ_{12} from a two-equation system with $N = 1000$ observations and gprvar-g prior. 139
A.46 Density value marginal posterior of σ_{13} at zero and SD-ratio for three-equation system, $N = 100$ and $\rho_{12} = -0.7$ and $\rho_{13} = -0.6$. 140
A.47 Density value of marginal posterior of σ_{13} at zero and SD-ratio for three-equation system, $N = 100$ and $\rho_{12} = -0.4$ and $\rho_{13} = 0$. 140
A.48 Density value of marginal posterior of σ_{13} at zero and SD-ratio for three-equation system, $N = 100$ and $\rho_{12} = -0.7$ and $\rho_{13} = -0.3$. 141
A.49 Density value of marginal posterior of σ_{13} at zero and SD-ratio for three-equation system, $N = 500$ and $\rho_{12} = 0.7$ and $\rho_{13} = 0.6$. 141
A.50 Density value of marginal posterior of σ_{13} at zero and SD-ratio for three-equation system, $N = 500$ and $\rho_{12} = -0.7$ and $\rho_{13} = -0.6$. 141
A.51 Density value of marginal posterior of σ_{13} at zero and SD-ratio for three-equation system, $N = 500$ and $\rho_{12} = -0.7$ and $\rho_{13} = -0.3$. 141
A.52 Description of variables used in the empirical application based on Table A1 in Jetter and Parmeter (2018). 142
A.53 List of countries and their corresponding CPI based on Table A4 in Jetter and Parmeter (2018). 143
A.54 Posterior estimation results empirical example full system, using the original I-I prior for model parameters and $\Sigma \sim IW(I_{13}, 14)$. 144
A.55 Posterior estimation results empirical example tested-down system, using the original I-I prior for model parameters and $\Sigma \sim IW(I_7, 8)$. 145
A.56 Posterior estimation results empirical example full system, using the original I-I prior for model parameters and $\Sigma \sim IW(Q_{13,c}, 15)$. 146

A.57 Posterior estimation results empirical example tested-down system, using the original I-I prior for model parameters and $\Sigma \sim IW(Q_{3,c}, 5)$. 147
A.58 Posterior estimation results empirical example full system, using the g-g prior for model parameters and $\Sigma \sim IW(I_{13}, 14)$. 148
A.59 Posterior estimation results empirical example down-tested system, using the g-g prior for model parameters and $\Sigma \sim IW(I_5, 6)$. 149
A.60 Posterior estimation results empirical example full system, using the g-g prior for model parameters and $\Sigma \sim IW(Q_{13,c}, 15)$. 150
A.61 Posterior estimation results empirical example down-tested system, using the g-g prior for model parameters and $\Sigma \sim IW(Q_{2,c}, 4)$. 151
A.62 Posterior estimation results empirical example full system, using the gprvar-g prior for model parameters and $\Sigma \sim IW(I_{13}, 14)$. 152
A.63 Posterior estimation results empirical example down-tested system, using the gprvar-g prior for model parameters and $\Sigma \sim IW(I_4, 5)$. 153
A.64 Posterior estimation results empirical example full system, using the gprvar-g prior for model parameters and $\Sigma \sim IW(Q_{13,c}, 15)$. 154
A.65 Posterior estimation results empirical example down-tested system, using the gprvar-g prior for model parameters and $\Sigma \sim IW(Q_{8,c}, 10)$. 155

Acknowledgements

I owe a great deal of debt to many people whose help enabled me to write and, most importantly, finish this thesis. First and foremost, I am very grateful to my supervisor Mark Steel for the outstanding mentoring and support he has given me in the last four years. My scientific work has benefited vastly from the enjoyable and fruitful atmosphere here at the Department of Statistics at the University of Warwick. Amongst others, I want to thank Rohan Lourdelet, Simon Parkes, Phillip Harvey-Smith, Ric Crossman and Matt Moores, who helped me in various aspects, be it administrative support or helping to provide computational resources. I also want to express gratitude to my panel members Jane Hutton, Simon Spencer and Martyn Plummer for constructive comments, valuable suggestions and stimulating discussions. I want to thank Alex Lenkoski for his openness regarding this work and for taking the time for detailed discussions. Further, I would like to mention my appreciation towards Michael Jetter for sharing the data I was able to use in the empirical application in this thesis. In addition, I would like to thank the Department of Statistics, the anonymous donor and the EPSRC for their financial support that eventually made this work possible. I also want to stress my deep gratitude towards my friends Thomas Kampf, Ifan Dafydd Johnston, Rémy Gavard, Lewis Rendell, Kenneth Lim, Nora Haunert and Robert Röhle who supported me throughout the years in a myriad of different ways. Finally, I owe more than can be said to the support and encouragement offered to me by my family and, particularly, by my partner Peter Malec, who has been the most important personal supporter throughout the entire process.

Declaration

I, Anne Miloschewski, declare that the PhD thesis entitled ’Bayesian Treatment of Model Uncertainty Under Endogeneity’ contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work.

Abstract

The main goal of this thesis is to extend BMA methodology under endogeneity, specifically for IV models. It is comprised of five main sections. Section 2 gives an overview of the general methodology for dealing with model uncertainty. In Section 3, we review the work that has been done in this context for the standard linear regression model. From that, the reader gains a better understanding of the factors that are influential within BMA frameworks and need to be considered carefully. In Section 4, the model we base the analysis on is introduced and the effects of ignoring endogeneity within a model uncertainty framework are investigated. We now know that, not only do we obtain biased posterior point estimates (false posterior means), but also wrong posterior inclusion probabilities, if endogeneity is not accounted for properly. We then continue in Section 5 by examining a particular approach to this problem more closely. In this context, we introduce adapted prior structures in a more generalised setting for the Karl and Lenkoski (2012) approach and investigate their influence on the outcome of the estimation procedure by means of simulation studies. Additionally, we develop a tool for the classification of variables with regards to being endogenous or exogenous. To achieve that, we extended the code base of the ivbma R package. Within an empirical application example, we demonstrate that the choices of the priors do matter. Furthermore, we can point out that using our proposed prior structures and endogeneity classification approach helps to obtain improved results in terms of out-of-sample-prediction accuracy. In Section 6, we first give a short overview of Bayesian frameworks for the IV model that do not incorporate model uncertainty. Subsequently we analyse and extend the Hoogerheide et al. (2007) approach, conducting the first analytical steps towards the latter.

Abbreviations

BMA	Bayesian model averaging
SEM	simultaneous equation model
2SBMA	instrumental variables two-stage Bayesian model averaging
2SLS	two-stage least-squares
AE	absolute error
AIC	Akaike information criterion
BACE	Bayesian averaging of classical estimators
BAMLE	Bayesian averaging of maximum likelihood estimators
BF	Bayes factor
BIC	Bayesian information criterion
CBF	conditional Bayes factor
CRPS	continuous rank predictive score
FMA	frequentist model averaging
GMM	generalised method of moments
IV model	instrumental variables model
JMA	jack-knife model averaging
LIBIC	limited information Baysian information criterion
LIBMA	limited information Bayesian model averaging
LIML	limited information maximum likelihood
LPS	log-predictive score
MCMC	Markov chain Monte Carlo
MMA	Mallows model averaging
PIP	posterior inclusion probability
PR	posterior odds ratio
RJMCMC	reversible jump Markov chain Monte Carlo
RRF	resctricted reduced form
SD-ratio	Savage-Dickey density ratio
SE	squared error
SF	structural form
UN	United Nations
URF	unrestricted reduced form
VAR	predictive variance

1 Introduction

Model uncertainty is a problem in various fields of research like economics (e.g. Cuaresma and Doppelhofer, 2007; Moral-Benito, 2015; Steel, 2019), political and social sciences (e.g. Montgomery and Nyhan, 2010; Raftery, 1995), biostatistics (e.g. Hoeting et al., 1999; Yeung et al., 2005) as well as astrophysics (e.g. Parkinson and Liddle, 2013; Trotta, 2008). Very often model uncertainty is associated with the uncertainty regarding which covariates to include in the model and, if so, in which (functional) form. Therefore, the choice of one particular model is often difficult and rather ad-hoc, leading to underreporting of the variability and a lack of robustness of the results.

First attempts at dealing with this kind of uncertainty can be found in Leamer (1985), who conducts an extreme bounds analysis. According to the corresponding test, a variable is 'fragile' when its coefficient changes sign in only one of the possible regressions. However, Sala-i-Martin (1997) argues that Leamer's extreme bounds test is too harsh for most of the variables. In particular, it is not plausible to treat all the regressions equally. Sala-i-Martin (1997) therefore alters the above test to overcome some of the flaws by examining the distribution of the parameters and deciding according to a level of confidence whether a variable should be included or not. However, these methods are still rather ad-hoc since they lack a sound statistical justification.

To overcome the problem of a lacking statistical underpinning, Bayesian methods were introduced. In particular, Bayesian model averaging (BMA) has become popular in the literature, since it has proven to be a robust method to incorporate the aforementioned uncertainty and to provide models with superior predictive abilities (see, e.g. Fernández et al., 2001b). Especially due to the fact that computational challenges nowadays are solvable in a reasonable amount of time, this kind of approach is being used more and more often in applied research. In this context, e.g., Clyde and George (2004) argue that BMA is particularly useful for cases where no single model stands out in terms of probability.

In the field of applied economics, the BMA methodology is particularly useful. Researchers often face the problem of lacking an accurate underlying theory of what drives certain quantities that they are interested in, one example being *economic growth*. One specific model class that is used quite frequently in this context is the standard linear regression model. For the latter, the influence of prior choices for the parameters and models on the results of BMA is relatively well-understood. Furthermore, there do exist good software packages like the well-established *BMS* R package developed by Martin Feldkircher and Stefan Zeugner.

However, the linear regression model does not always suffice to accurately represent relations between the quantities of interest. One particular complication which can arise is the endogeneity problem.

A possible reason for why this problem may occur is that the causality between the target variable and a regressor is basically bidirectional at the same point in time. An existing class of models for dealing with endogeneity are simultaneous equation models (SEMs), of which the instrumental variables model (IV model) is a special case. The latter is frequently used and well-known in the applied economic research community in a frequentist setting. However, Bayesian frameworks that also incorporate model uncertainty for this model class are not as well-established and understood, yet. The aim of this thesis is therefore to investigate BMA methodology within the IV model setting.

Accordingly, the remainder of this thesis is composed of the following six sections. Section 2 gives an overview of the general methodology for dealing with model uncertainty. In Section 3, we review the work that has been done in this context for the standard linear regression model in order to provide the reader with a better understanding of the factors that have been found to be influential regarding the results. It is also important to discuss these insights for the case of the standard linear regression framework as they will serve as guidance for the extensions under endogeneity. In Section 4, we introduce the model we are basing the extensions on, investigate what the effects of ignoring endogeneity are within a model uncertainty framework and briefly discuss the existing approaches for dealing with model uncertainty in the IV model case. We then continue in Section 5 by examining a particular approach to this problem more closely, analysing it in a more generalized setting, introducing some extensions to the existing prior set-up and investigating their influence by means of a simulation study. We additionally propose a tool that makes it easier to decide which variables should be assumed to be endogenous. The section closes with an empirical application example. In Section 6, we first give a short overview of Bayesian frameworks for the IV model that do not incorporate model uncertainty. We then examine one particular approach that could lead to a different approach for the underlying problem when model uncertainty is incorporated. Section 7 concludes by summarizing the most important findings of this thesis and sketching how they could be used and extended in further research.

2 Dealing with Model Uncertainty

Bayesian Treatment of Model Uncertainty

An introduction to and summary of the foundations of Bayesian statistics and econometrics can be found in Bernardo and Smith (1994) or Koop (2003). The following notation is borrowed from the latter. Assume we have a model for a vector or matrix of data Y, which is parameterised by θ. By Bayes' rule:

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)}, \tag{2.1}$$

where $p(\theta|Y)$ is the posterior distribution of a random variable θ, which can be a parameter vector (or matrix). $p(Y|\theta)$ denotes the likelihood function and $p(\theta)$ is the prior density on that quantity θ, which reflects the prior beliefs and knowledge of the researcher about the quantity. In contrast to frequentist statistics, θ is not a fixed and unknown parameter, but a random variable with a distribution reflecting its uncertainty. The marginal likelihood, $p(Y)$, does not depend on the parameters θ and is therefore a constant. Hence, the posterior distribution of θ can also be expressed as being proportional to the product of likelihood and prior:

$$p(\theta|Y) \propto p(Y|\theta)p(\theta). \tag{2.2}$$

Raftery et al. (1997) show how model uncertainty can be expressed using simple rules of probability. A special case is model uncertainty implied by the uncertainty as to which variables should be included in the model. In this thesis we focus on the latter. We can introduce an index i that corresponds to the distinct subset of parameters in θ that model i contains. The entire model space is denoted by $\mathscr{M}$ and its elements are models $M_i, i = 1, ..., I$, where $I = 2^k$ with k being the total number of variables (full model). We treat the model itself as a random quantity. Hence, the random quantity θ_i depends upon the model M_i. Therefore, the posterior distribution of θ_i can be formulated conditionally on the model:

$$p(\theta_i|Y, M_i) = \frac{p(Y|\theta_i, M_i)p(\theta_i|M_i)}{p(Y|M_i)}. \tag{2.3}$$

Applying the rules of probability, one can calculate posterior model probabilities:

$$p(M_i|Y) = \frac{p(Y|M_i)p(M_i)}{p(Y)}, \tag{2.4}$$

where $p(M_i)$ is the prior probability of model M_i, or

$$p(M_i|Y) \propto p(Y|M_i)p(M_i), \tag{2.5}$$

with the marginal likelihood for model M_i being the integrating constant in (2.3):

$$p(Y|M_i) = \int p(Y|\theta_i, M_i)p(\theta_i|M_i)\,\mathrm{d}\theta_i. \tag{2.6}$$

The overall marginal likelihood is given by $p(Y) = \sum_{i=1}^{I} p(Y|M_i)p(M_i)$.

Now, if we are interested in some quantity δ that is not model specific, e.g., the effect of a covariate or some predictive quantity, we can express the posterior distribution of δ as a weighted average of posterior distributions of δ in a specific model, weighted by the posterior model probability of that specific model:

$$p(\delta|Y) = \sum_{i=1}^{I} p(\delta|Y, M_i)p(M_i|Y). \tag{2.7}$$

As in the general Bayesian framework, we can calculate the variance and the expectation of that random quantity δ:

$$E(\delta|Y) = \sum_{i=1}^{I} E(\delta|Y, M_i)p(M_i|Y), \tag{2.8}$$

where the expected value is a weighted average of the expected values of δ in a specific model M_i with the weights given by the posterior model probability of model M_i and

$$\mathrm{Var}(\delta|Y) = \sum_{i=1}^{I} \mathrm{Var}(\delta|Y, M_i)p(M_i|Y) + \sum_{i=1}^{I} \left(E(\delta|Y, M_i) - E(\delta|Y)\right)^2 p(M_i|Y). \tag{2.9}$$

The first part is a weighted average of the variances and the second part reflects the variability of the expectation of δ between the specific models in comparison to the average expected value of δ.

Similarly, an averaged forecast or predictive density can be defined as

$$p(y^*|y) = \sum_{i=1}^{I} p(y^*|y, M_i)p(M_i|y), \tag{2.10}$$

with

$$p(y^*|y, M_i) = \int p(y^*|y, \theta_i, M_i)p(\theta_i|y, M_i)\,\mathrm{d}\theta_i. \tag{2.11}$$

In order to compare two models, say M_1 and M_2, directly to one another, one can calculate the posterior odds ratio (PR), which is defined as the product of the Bayes factor (BF) and the prior odds ratio:

$$PR_{1,2} = \frac{p(M_1|Y)}{p(M_2|Y)} = \text{BF}_{1,2} \times \text{prior odds ratio}_{1,2} = \frac{p(Y|M_1)}{p(Y|M_2)} \times \frac{p(M_1)}{p(M_2)}. \tag{2.12}$$

If models M_1 and M_2 are equally likely a priori, the PR is equal to the BF:

$$PR_{1,2} = BF_{1,2} = \frac{p(M_1|Y)}{p(M_2|Y)} = \frac{p(Y|M_1)}{p(Y|M_2)}. \tag{2.13}$$

If the $\text{PR}_{1,2}$ is larger than one, model M_1 is preferable to model M_2.

In the following, we will assume that the 'true' model is part of the model space we consider ($\mathscr{M}$-closed assumption).

Hybrid Approaches and Approximations

A hybrid approach combines frequentist as well as Bayesian elements to deal with model uncertainty. Often, the parameter estimates result from a frequentist estimation procedure and model weights are derived in a Bayesian fashion. One example is the Bayesian Averaging of Maximum Likelihood Estimators (BAMLE) approach. Here, the weights are proportional to a Bayesian information criterion (BIC) approximation[1] (see Moral-Benito, 2012).

Another example is the Bayesian Averaging of Classical Estimators (BACE) approach by Sala-i-Martin et al. (2004). The estimates of the parameters are constructed as weighted averages of OLS estimates. The authors argue that this approach is a limiting case of a standard Bayesian analysis as the prior information becomes dominated by the data information. OLS estimates result from the assumption of diffuse priors when using the g-prior (see section 3) and taking the limit as the information contained in the data becomes large. The prior model probabilities are specified by choosing a prior mean model size $\bar{k}$ such that each variable is included with probability $\frac{\bar{k}}{k}$ (k being the total number of regressors). Hence, the model weights are proportional to the BIC resulting from the BIC approximation. Using this approach, the authors only need to specify one hyper-parameter, namely the expected model size. The latter is easy to interpret and the overall approach is easy to implement.

[1] The derivation of the BIC approximation can be found in Raftery (1995).

However, Ley and Steel (2009) show that using the BACE approach implicitly corresponds to making certain prior assumptions in the linear regression model. Further, there are three major prerequisites for the BIC approximation to be a good approximation of the Bayes factors. First, even though it does not rely on the data being independent and identically distributed, a large number of observations is needed. In practice, this is often problematic with the meaning of 'large' depending on the concrete setting. Second, a Laplace approximation is used during the derivation, which requires the underlying distribution to be uni-modal, otherwise the approximation can be bad as well. Third, the approximation error gets larger and does not vanish if we do not use a multivariate normal prior for the parameters. Hence, if the model becomes more complicated or the researcher does not have a considerable amount of data at his disposal, it is questionable whether this approximation is admissible.

Therefore, even though the BACE approach is very attractive in terms of simplicity and interpretability, in most situations it is advisable to specify a fully Bayesian model.

Frequentist Approaches

Hjort and Claeskens (2003) argue that the ad hoc choice of the prior and the mixing of potentially conflicting prior opinions regarding the parameters of interest are problematic in the Bayesian approach. Further, Wang et al. (2009) state that the corresponding estimators in the frequentist setting are totally driven by data, implying that the researchers' subjective opinion about the parameters would have little or no influence on the results. An overview of some frequentist methods to incorporate model uncertainty can be found in Hjort and Claeskens (2003), Wang et al. (2009) and Moral-Benito (2015).

Wang et al. (2009) define the frequentist model averaging (FMA) estimator in the following way:

$$\hat{\theta}_{FMA} = \sum_{i=1}^{I} \omega_{M_i} \hat{\theta}_{M_i},$$

where, again, $I = 2^k$ is the total number of models with k being the total number of parameters, $\hat{\theta}_{M_i}$ is the parameter estimate for model M_i and ω_{M_i} is the corresponding weight. $\sum_{i=1}^{I} \omega_{M_i} = 1$ and $0 \leq \omega_{M_i} \leq 1$.

The weights can be chosen in various ways with one being to base them on an information criterion. Let IC_{M_i} be the information criterion with

$$IC_{M_i} = -2\log(L_{M_i}) - l,$$

where L_{M_i} is the maximised likelihood under model M_i and l is a penalty function depending on the number of parameters and the number of observations. The weights can be constructed using IC_{M_i}

$$\omega_{M_i} = \frac{\exp(-\frac{1}{2}\mathrm{IC}_{M_i})}{\sum_{m=1}^{I}\exp(-\frac{1}{2}\mathrm{IC}_{M_m})}.$$

If $l = 2k$, the resulting weights lead to the smoothed Akaike-information criterion (AIC) estimator. Another possibility for a standard linear regression model is to use Mallows' criterion (Mallows Model Averaging (MMA)), which is asymptotically optimal if the weight vector used is constrained to a discrete set. In the case of a more general linear model, e.g. with heteroskedastic errors, the model weights should be based on the cross-validation criterion also known as jack-knife model averaging (JMA). JMA is more efficient in this setting and the models are allowed to be non-nested.

However, the argument that the frequentist approach is free from ad hoc decisions made by the researcher is ultimately not valid. Hjort and Claeskens (2003) show in an example that, e.g., different choices of information criteria can lead to different choices of models, while using the same data.

Different weights lead to different asymptotic properties of the corresponding FMA estimator, which is equally problematic as the fact that a Bayesian researcher has to choose a prior distribution for the models and the parameters (see Moral-Benito, 2015) . Further, Wang et al. (2009) point out that many issues concerning the FMA approach in more complex models, such as models with endogeneity issues, panel data models or models for censored data, are not solved or analysed, yet.

An additional problem with frequentist methods arises due to the fact that both the theory of optimal weight choice and the (knowledge of the) distribution of the FME depend on the availability of large sample sizes. In practice, however, the latter often are not available to the researcher. Finally, the 'finite-sample problem' is not something a Bayesian researcher has to worry about, which is a big advantage of the Bayesian approach (see Zellner, 1970).

Given the above reasons, it is perfectly valid to focus on Bayesian approaches.

3 Literature Review - Model Uncertainty and the Standard Linear Regression Model

A major part of the BMA literature focuses on the analysis of the standard linear regression model. Prior assumptions on the parameter and model space as well as their implications on posterior inference are well-studied and understood in this context. The standard linear regression model is defined in the following way:

$$y = \alpha\iota + X\gamma + \epsilon, \qquad \epsilon \sim N(0, \sigma^2 I_N), \tag{3.1}$$

where y is a $N \times 1$ vector, α is a scalar, ι is a $N \times 1$ vector of ones, X is a $N \times k$ matrix of covariates which are assumed to be exogenous, i.e. $E(X'\epsilon) = 0$, and γ is the corresponding $k \times 1$ vector of parameters. The error process consists of independent, identically distributed errors with $\epsilon_j \sim N(0, \sigma^2), j = 1, ..., N$. Further, it is assumed that $\iota' X = 0$ and that X is of full column rank.

Given the above model, one can formulate the likelihood conditional on model M_i:

$$p(y|\alpha, \gamma, \sigma^2, M_i) = \frac{1}{\sqrt{\pi^N \sigma^{2N}}} \exp\left(-\tfrac{1}{2\sigma^2}(y - \alpha\iota - X_i\gamma_i)'(y - \alpha\iota - X_i\gamma_i)\right), \tag{3.2}$$

where the model index i denotes the subset of variables which are included in model M_i.

Raftery et al. (1997) describe an approach for the linear regression model with generalized error covariance using a standard conjugate class of priors. Further, they discuss Occam's window and a Markov chain Monte Carlo algorithm, MC^3, originally developed by Madigan and York (1995). Additionally Raftery et al. compare the predictive performance between BMA and 'just using the best model' and establish that the BMA methodology posseses better predictive abilities.

Closely related to Raftery et al. (1997) is the study by Hoeting et al. (1999) who provide a general overview of the use of BMA and its implementation details for different model classes (normal linear regression model, generalized linear models, survival analysis and graphical models). Further, Clyde and George (2004) describe BMA as a natural and general framework that simultaneously treats both model and parameter uncertainty. They argue that BMA is particularly useful for cases where no single model stands out in terms of probability. In addition, it is described (in a fairly general manner) how the BMA framework can be implemented for different kinds of model classes (non-parametric regression, generalized linear models, tree models and graphical models).

Moral-Benito (2015) gives a comprehensive summary of the literature dealing with model uncertainty in a Bayesian and frequentist manner. However, in contrast to the aforementioned authors, he also includes an overview of prior choices for the parameter and model space as well as their implications in terms of posterior analysis. Additionally, he summarizes some of the advances which have been made in the BMA literature in the context of dealing with more complex structures like models involving endogeneity or panel data models.

3.1 Priors for the Model Parameters

One possible prior choice is the natural conjugate prior structure, which implies a Normal distribution for γ and an Inverted-Gamma distribution for σ^2 (see Raftery, 1995). However, Fernández et al. (2001a) show that, especially if there is only little prior knowledge about the parameters, it might be beneficial to consider a combination of a Jeffreys prior and a g-prior (for the latter see Zellner, 1986). In that case, the only hyperparameter which has influence on the Bayes factor and which must be determined in advance is g. This prior is probably the most studied and popular one for the coefficients of the above model (e.g. Fernández et al., 2001a,b; Liang et al., 2008; Ley and Steel, 2012). The priors for the model parameters hence are

$$p(\sigma^2) \propto \frac{1}{\sigma^2}, \tag{3.3}$$

$$p(\alpha) \propto 1, \tag{3.4}$$

$$p(\gamma|\sigma^2, M_i) = f_N^i\left(0, \sigma^2 g (X_i^{'} X_i)^{-1}\right). \tag{3.5}$$

The priors for α and σ^2 are improper, but since these parameters appear in every model and have the same interpretation, this fact does not cause problems related to the Bayes factors.

We can calculate the marginal likelihood as

$$p(y|M_i) = \frac{\Gamma\left(\frac{N-1}{2}\right)}{\sqrt{N}\pi^{\frac{N-1}{2}}} \left(\tfrac{1}{1+g}\right)^{\frac{k_i}{2}} \left(y^{'} y - N\bar{y} - \tfrac{g}{1+g} y^{'} X_i (X_i^{'} X_i)^{-1} X_i^{'} y\right)^{-\frac{N-1}{2}}. \tag{3.6}$$

This expression can also be written in terms of Z_i, where $Z_i = (\iota : X_i)$:

$$p(y|M_i) = \frac{\Gamma\left(\frac{N-1}{2}\right)}{\sqrt{N}\pi^{\frac{N-1}{2}}} \left(\tfrac{1}{1+g}\right)^{\frac{k_i}{2}} \left(\tfrac{1}{1+g}(y - \bar{y}\iota)'(y - \bar{y}\iota) - \tfrac{g}{1+g} y^{'} Z_i (Z_i^{'} Z_i)^{-1} Z_i^{'} y\right)^{-\frac{N-1}{2}}. \tag{3.7}$$

Assuming that g is a fixed value, Fernández et al. (2001a) analyse various choices for g in a simulation study and measure the resulting predictive performance using the log predictive score. They suggest to set $g = \max\{N, k^2\}$, where k is the maximum number of parameters (full model). An application of this methodology can be found in Fernández et al. (2001b).

Fernández et al. (2001a) argue that, under a uniform prior for the model space, g acts like a dimensionality penalty and effectively controls model selection. Large values of g tend to concentrate the prior on small models with large coefficients, while small values of g tend to concentrate the prior on saturated models with small coefficient values. The authors show that for $g = N$, the BF asymptotically behaves like the BIC criterion and, for $g = k^2$, the BF asymptotically behaves like the risk inflation criterion of Foster and George (1994).

Liang et al. (2008) review the g-prior and analyse its influence on the resulting Bayes factors. In their analysis, the authors show that the BF, for fixed choices of g and a uniform prior on the model space, exhibits some undesirable features. First, if $g \to \infty$ while N and the number of parameters are fixed, the Bayes factor (null model based) goes to zero. Thus, the large spread of the prior induced by the non-informative choice of g forces the BF to favour the null model regardless of the information in the data. This is called 'Bartlett's paradox'. Second, suppose there is overwhelming information supporting M_i, meaning that $R_i^2 \to 1$, and we were to compare the same to the null model[2]. While N and the parameters are fixed, the usual F-statistic will go to infinity. However, the BF will converge to a constant in this setting, such that the limiting behaviour will not reflect the information in the data. This phenomenon is called 'information paradox'. There also exist empirical Bayes methods to determine g, which avoid the information paradox. However, one has to use the data in advance to determine the hyper -parameter, which is conflicting with the overall Bayesian idea.

A solution to the above problem is to treat g not as a constant, but to put a prior on g as well, leading to a hierarchical prior structure. Liang et al. (2008) discuss two major groups for the latter, i.e., the Zellner-Siow and the hyper-g priors. The former prior can be represented as a mixture of g-priors with an Inverse Gamma prior on g. If the two models under consideration are nested, a flat prior is put on the common parameters and a Cauchy prior (which can be expressed as a scale mixture of normals) on the remaining parameters. The authors show that the Zellner-Siow prior provides

[2] R_i^2 is the coefficient of determination of model i.

robustness to misspecifications of g, solves the information paradox, has asymptotic consistent model posterior probabilities[3] if the true model is not the null model and is asymptotic consistent for prediction under squared error loss. The disadvantage of this prior family is that there exists no closed form for the marginal likelihoods. Hence, they have to be approximated.

Hyper-g priors are priors on g of the form $p(g) = \frac{a-2}{2}(1+g)^{-\frac{a}{2}}$ for $g > 0$. This is a proper distribution for $a > 2$. The authors state that one could also consider the corresponding prior on the shrinkage factor $\frac{g}{1+g} \sim \text{Beta}(1, \frac{a}{2}-1)$, such that, for $a = 4$, the prior on the shrinkage factor is uniform. In contrast to the standard g prior, the hyper-g prior leads to nonlinear data-dependent shrinkage. It has similar properties to the Zellner-Siow prior. However, the hyper-g-prior does not lead to asymptotically consistent model posterior probabilities since, in contrast to the Zellner-Siow prior, it does not depend on N. For that reason, the authors suggest a modification, which they refer to as hyper-g/n prior, defined as $p(g) = \frac{a-2}{2N}\left(1+\frac{g}{N}\right)^{-\frac{a}{2}}$. Here, no analytical expressions are available for the quantities of interest, such that an approximation, as for example the Laplace-approximation, needs to be used. The latter implies, as the Zellner-Siow prior, consistent model selection under 0-1 loss for any assumed true model.

There exists a generalized version of the g-prior by Maruyama and George (2011), which puts more shrinkage on higher variance estimates. It can also be used in the case where the number of parameters is larger than the number of observations which, especially for applications in economics, can be very useful.

3.2 Priors on Model Space

Ley and Steel (2009) investigate the influence of prior assumptions on the model space in the standard linear regression framework regarding the posterior results when $g = k^2$ or $g = N$. In previous studies, the prior on the model space is often specified as $p(M_i) = p^{k_i}(1-p)^{k-k_i}$, assuming that each regressor enters the model independently with probability p. In particular, for $p = 0.5$, which corresponds to the benchmark choice, this expression collapses to a uniform prior with $p(M_i) = 2^{-k}$ with expected model size $\frac{k}{2}$. In a more general way, this can be formulated by defining a random variable ϑ_i, which is equal to one if the i-th covariate is included in the model and zero otherwise. Then, $\vartheta_i \sim \text{Bernoulli}(p)$. Hence, the model size $W = \sum_{i+1}^{k} \vartheta_i$ with $W \sim \text{Bin}(k, p)$. For a fixed p, the expected model size is $E(W) = pk = m$ and the variance is $\text{Var}(W) = p(1-p)k$.

[3]If M_s is the true model which generated the data, then $\text{plim}_{N\to\infty}\ p(M_s|y) = 1$ and $\text{plim}_{N\to\infty}\ p(M_j|y) = 0\ \forall M_j \neq M_s$ (see, e.g., Fernández et al., 2001a).

If p is treated as a random quantity and a prior distribution is put on p, such as a Beta(a,b) distribution, then the expected model size and variance of the model size will be functions of a, b and k. The prior model size distribution becomes a Binomial-Beta distribution:

$$P(W=w) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)\Gamma(a+b+k)} \begin{pmatrix} k \\ w \end{pmatrix} \Gamma(a+w)\Gamma(b+k-w), \quad w = 0, ..., k.$$

If $a = 1$, one only needs to determine the prior mean model size m with $0 < m < k$. The resulting expected value and variance are $E(W) = \frac{k}{1+b} = m$, such that $b = \frac{k-m}{m}$ and $p = \frac{m}{k}$. Ley and Steel (2009) show that treating p as random implies a substantial increase in prior uncertainty.

Since the posterior odds between any two models in $\mathscr{M}$, say M_1 and M_2, are defined as

$$\frac{P(M_1|y)}{P(M_2|y)} = \frac{P(M_1)}{P(M_2)} \frac{l_y(M_1)}{l_y(M_2)},$$

one can see that the prior distribution on the model space affects posterior inference only through the prior odds ratio $\frac{P(M_1)}{P(M_2)}$. For fixed p, the prior odds ratio in terms of the corresponding mean model size is $\frac{P(M_1)}{P(M_2)} = \left(\frac{m}{k-m}\right)^{k_1-k_2}$. Thus, if the prior mean model size satisfies $m > \frac{k}{2}$, the prior favours larger models and therefore the choice of m is crucial in this setting. If, however, p is random with a Beta-prior on p, prior odds are

$$\frac{P(M_1|y)}{P(M_2|y)} = \frac{\Gamma(1+k_1)}{\Gamma(1+k_2)} \frac{\Gamma(\frac{k-m}{m} + k - k_1)}{\Gamma(\frac{k-m}{m} + k - k_2)},$$

where k_1 and k_2 denote the number of parameters not set to zero in model M_1 and M_2 respectively. The authors state that, in this case models with k_2 around $\frac{k}{2}$ are always downweighted irrespectively of m and that the choice of m is less critical. They further discuss the influence of the simultaneous choice of g and m on the posterior odds. Suppose the posterior odds are equal to one, i.e.

$$1 = \frac{p(M_1|y)}{p(M_2|y)} = \frac{p(y|M_1)}{p(y|M_2)} \frac{p(M_1)}{p(M_2)}.$$

If the prior model probabilities are equal, the marginal likelihoods, which are influenced by g, would need to be similar. If the prior model probabilities are not the same, then the fit of one of the models has to be better to compensate for a smaller a priori probability.

The author's analysis of the interrelation between the choices of g and m shows that, for fixed p, there exists a clear trade-off between m and g. A larger m can be compensated by a small g, which increases the penalty for larger models. Ley and Steel advise against using $g = N$ because it is even more sensitive to m and state that the worst combination in terms of predictive performance is setting $p = 0.5$ and $g = N$. For a random p, though, the trade off between m and g almost disappears, such that the choice of m is not that relevant any more. Finally, Ley and Steel (2009) suggest to use a hierarchical structure for the prior on the model space and to set $g = k^2$.

Ley and Steel (2012) go one step further and use a Binomial-beta prior on the model size and combine this with a hierarchical prior structure for the parameters in order to analyse the influence of these prior choices on the resulting Bayes factors. For the latter, a g-prior is used and different choices of hyper-priors for g are compared in this setting[4]. The authors perform an extensive simulation study, but also compare the performance of the priors when applying the methodology to real data. If N is small in comparison to k, the Zellner-Siow and the Maruyama-George priors perform well. However, the authors recommend two priors in particular. First, the benchmark Beta prior, which is a Beta(b,c) prior on the shrinkage factor $\delta = \frac{g}{1+g}$ while setting c=0.01. Second, they recommend the hyper-g/N prior with a=3. Ley and Steel state that the use of hyper-priors helps to avoid the information paradox and makes the analysis more robust with respect to arbitrary prior assumptions.

Another interesting class of model space priors, that has not been studied as well so far, are dilution priors. The latter have been proposed by George (2010). In a comment on Hoeting et al. (1999), he states that one might need to compensate for redundancies between model classes and argues that dilution priors may be useful to avoid biasing away from good, but isolated models. George (2010) suggests three possible approaches. The first one is tessellation-defined dilution priors for a linear model variable selection setup. The main idea is to identify neighbourhoods of models with appropriate tesselations of the surface of a high-dimensional sphere and then assign assign uniform probabilities to neighbourhood models. The second approach is collinearity adjusted dilution priors where the model probability is also determined by a monotone function h of the determinant of the correlation matrix of the regressors in that model. If the columns of the covariate matrix are highly correlated and therefore redundant, the determinant of that matrix decreases to zero. Hence, the model probability is downweighted somewhat,

[4]The hyperpriors compared are, e.g., the Zellner-Siow prior, Beta-Shrinkage priors for different choices of parameters and the Hyper-g/n prior.

depending on the choice of h. One adaptation of this approach can be found in Durlauf et al. (2008), who assign priors to various combinations of proxies for different growth theories. The third possibility the authors suggest are model-distance-based dilution priors, which work for any class of models and are not restricted to linear models. A possible distance function could be the Hellinger distance of the marginal distributions of two models. The prior probability of a model is then a function of its distance to any other model. As the authors point out, variations and features of these priors and empirical applications need to be studied in future research.

Jointness

For the linear regression BMA framework, Ley and Steel (2007) suggest a measure called *jointness*, in order to be able to display some more informative aspects of the posterior beyond information contained in the marginals. In particular, *jointness* measures the dependence among explanatory variables. The concept relates to variables that are capturing different sources of relevant information (complements) while *disjointness* on the other hand relates to variables which perform similar roles, e.g., being highly collinear (substitutes). In that sense, the authors define *multivariate jointness* for a general set of regressors. This can be an additional useful tool to learn about relations between variables in complex empirical data sets.

4 Introducing Endogeneity

In the previous section we have seen, that for the linear regression model, many well-understood tools are available to the researcher. In practice, however, the linear regression model often does not suffice due to more complex relations between the variables of interest and violations of underlying assumptions. We therefore need to expand the methodology to a wider range of problems and resulting models.

One complication which occurs quite often is the so-called endogeneity problem. It arises when a regressor in a model is not predetermined, that is, if the regressor is not orthogonal to the error term (see Hayashi, 2000, Ch.3). This basically means that at least one of the explanatory variables in the model is correlated with the error term.

There are several possible causes for this issue. As described in Hayashi (2000), the first reason are measurement errors concerning the regressors of the model. Second, omitting variables that affect both the regressor variable(s) as well as the dependent variable(s) is a further potential cause. Finally, there might not be a one-sided causality relation between the dependent and the regressor variable(s).

An example for the latter is the relationship between police numbers and crime rates. On the one hand, the number of police officers has an influence on crime rates since the more police are patrolling the streets, the less crimes are committed. But, at the same time, crime rates might influence the number of police officers, e.g., due to policy decisions. If crime rates are high, it is more likely that more police officers are hired.

Suppose we are in the frequentist world and would like to estimate a linear regression model of the form

$$Y = X\gamma + \epsilon,$$

where at least one of the regressors in X is endogenous. If we ignore the endogeneity and estimate γ according to a standard frequentist estimation method like ordinary least-squares (OLS), our estimate $\hat{\gamma}^{OLS} = (X'X)^{-1}X'y$ will be biased, i.e.

$$E(\hat{\gamma}^{OLS}) = E((X'X)^{-1}X'y) = E((X'X)^{-1}X'(X\gamma + \epsilon)) = \gamma + E((X'X)^{-1}X'\epsilon).$$

The last part is not equal to zero if X and ϵ are correlated and will not vanish with the number of observations $N \to \infty$. Hence, the estimation procedure is inconsistent.

In this simple setting, one possible frequentist solution would be to estimate γ using a two-stage least-squares procedure (2SLS). In the first stage, the endogenous variable is regressed on instruments and on the remaining regressors of the main equation[5]. Then, one obtains predicted values of the endogenous variable, which are used in the main equation instead of the endogenous variable itself. Overall, this procedure leads to a consistent estimate of γ. The question is what happens if we estimate the above model in a BMA framework while ignoring the endogeneity problem.

4.1 The Instrumental Variables Regression Model

We define an instrumental variables regression model (IV model) using capital letters for matrices and lower case letters for vectors or scalars[6]. The IV model in structural form (**SF**) can be written as:

$$\begin{aligned} y_1 &= Y_2\beta + X\gamma + \epsilon_1, \\ Y_2 &= Z\Delta + X\Lambda + V_2, \end{aligned} \tag{4.1}$$

where y_1 is a $N \times 1$ dependent variable and Y_2 a $N \times m$ matrix of endogenous variables. Z is a $N \times k_Z$ matrix of instruments and assumed to be of full column rank. Δ is a $k_Z \times m$ matrix of parameters and β a $m \times 1$ vector of parameters. X is a $N \times k_X$ matrix of exogenous variables and γ is a $k_X \times 1$ vector, while Λ is a $k_x \times m$ matrix of parameters. Further, ϵ_1 is a $N \times 1$ vector of structural form errors and V_2 is a $N \times m$ matrix of reduced form errors. They are jointly normally distributed with

$$\mathrm{Var}\begin{pmatrix} \epsilon_{1j} \\ \mathrm{vec}(V_2)_j \end{pmatrix} = \Sigma = \begin{pmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where j denotes the observation index with $j = 1, ..., N$. Here, the vec(A) operator denotes the vectorisation of the $p \times q$ matrix A into a $pq \times 1$ column vector. If our system had, e.g., two endogenous variables ($m = 2$), then $V_2 = (v_{21} : v_{22})$ would be a matrix of dimension $N \times 2$ with the notation $(A : B)$ denoting the placement of matrix A and matrix B next to each other (concatenation)[7]. Then, $\mathrm{vec}(V_2) = (v'_{21}\, v'_{22})'$ is of dimension $2N \times 1$. Now, for every observation $j = 1, ..., N$, $\mathrm{vec}(V_2)_j = (v'_{21,j}\, v'_{22,j})'$ will be of dimension 2×1. Furthermore, $\mathrm{vec}(\epsilon_{1j} : V_{2j})$ and $\mathrm{vec}(\epsilon_{1i} : V_{2i})$ are assumed to be independent for $j \neq i$.

[5] Instruments are variables which are correlated with the endogenous variable, but not with the error term of the main model.

[6] For the sake of simplicity of notation, we do not distinguish between random variables and their realizations.

[7] The same notation is used for vectors and scalars, as well.

Therefore,

$$\begin{pmatrix} \epsilon_1 \\ \text{vec}(V_2) \end{pmatrix} \sim N(0, \Sigma \otimes I_N).$$

Substituting the reduced form equation in the structural equation gives the non-linearly restricted reduced form (**RRF**):

$$\begin{aligned} y_1 &= Z\Delta\beta + X(\gamma + \Lambda\beta) + v_1, \\ Y_2 &= Z\Delta + X\Lambda + V_2, \end{aligned} \tag{4.2}$$

where

$$v_1 = \epsilon_1 + V_2\beta.$$

This implies that

$$\begin{pmatrix} v_1 \\ V_2 \end{pmatrix} \sim N(0, \Omega \otimes I_N), \tag{4.3}$$

with

$$\Omega = \begin{pmatrix} \omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \beta & I_m \end{pmatrix}' \Sigma \begin{pmatrix} 1 & 0 \\ \beta & I_m \end{pmatrix}.$$

The resulting unrestricted reduced form (**URF**) is

$$\begin{aligned} y_1 &= Z\tau + X\xi + v_1, \\ Y_2 &= Z\Delta + X\Lambda + V_2, \end{aligned} \tag{4.4}$$

where τ is a $k \times 1$ vector and $\begin{pmatrix} v_1 \\ V_2 \end{pmatrix}$ is distributed as in (4.3). Using the structural form, the likelihood can be formulated as

$$\begin{aligned} p(y_1, Y_2|\beta, \Delta, \Lambda, \gamma, \Sigma) = (2\pi)^{-N}|\Sigma|^{-\frac{1}{2}N}\exp\Big[-\tfrac{1}{2}\text{tr}\{\Sigma((y_1 : Y_2) - (Y_2\beta : Z\Delta) - X(\gamma : \Lambda))' \\ ((y_1 : Y_2) - (Y_2\beta : Z\Delta) - X(\gamma : \Lambda))\}\Big]. \end{aligned}$$

Now, assuming for the moment that the endogenous regressors are known, we introduce the index i for model M_i with $i = 1, ..., I$, where the index indicates a distinct set of variables which are contained in the model. Then, the **SF** is equivalent to

$$\begin{aligned} y_1 &= Y_2\beta + X\gamma_i + \epsilon_1, \\ Y_2 &= Z\Delta_i + X\Lambda_i + V_2, \end{aligned} \tag{4.5}$$

with the **RRF** given by

$$\begin{aligned} y_1 &= Z\Delta_i\beta_i + X(\gamma_i + \Lambda_i\beta_i) + v_1, \\ Y_2 &= Z\Delta_i + X\Lambda_i + V_2, \end{aligned} \tag{4.6}$$

where

$$v_1 = \epsilon_1 + V_2\beta_i,$$

and the **URF** simply becomes

$$\begin{aligned} y_1 &= Z\tau_i + X\xi_i + v_1, \\ Y_2 &= Z\Delta_i + X\Lambda_i + V_2. \end{aligned} \tag{4.7}$$

Using (4.5), the likelihood can be formulated as

$$p(y_1, Y_2|\beta_i, \Delta_i, \Lambda_i, \gamma_i, \Sigma) = (2\pi)^{-N}|\Sigma|^{-\frac{1}{2}N}\exp\Big[-\tfrac{1}{2}\mathrm{tr}\{\Sigma((y_1 : Y_2) - (Y_2\beta_i : Z\Delta_i) - X(\gamma_i : \Lambda_i))' \\ ((y_1 : Y_2) - (Y_2\beta_i : Z\Delta_i) - X(\gamma_i : \Lambda_i))\}\Big].$$

4.2 Effects of Ignoring Endogeneity Within the Standard Linear Regression BMA-Framework

In order to analyse the effects of disregarded endogeneity in different settings, we will conduct a simulation study. For the general data generating process we follow Lenkoski et al. (2014), who base their simulation framework on Fernández et al. (2001a) and Raftery et al. (1997).

First, we construct a $N \times 15$ matrix of exogenous regressors $X = (X_1 : X_2)$, such that the first ten columns of X denoted by X_1 with $x_1, ..., x_{10}$ are i.i.d. draws from a $N(0,1)$ distribution. X_2 with $x_{11}, ..., x_{15}$ is then constructed as

$$(x_{11} \dots x_{15}) = (x_1 \dots x_5)(0.3\ \ 0.5\ \ 0.7\ \ 0.9\ \ 1.1)'(1 \dots 1) + E,$$

where E is a $N \times 5$ matrix of i.i.d. $N(0,3)$ draws. This structure induces some correlation between the first and the last five variables in X to reflect typical properties of real empirical data.

The instruments $Z = (Z_1 : Z_2)$ are constructed in the same fashion. Z_1 with $z_1, ..., z_5$ is a $N \times 5$ matrix of i.i.d. draws from a $N(0,1)$ distribution and Z_2 is a $N \times 5$ matrix with

$$(z_6 \dots z_{10}) = (z_1 \dots z_5)(0.3\ \ 0.5\ \ 0.7\ \ 0.9\ \ 1.1)'(1 \dots 1) + F,$$

where F is a $N \times 5$ matrix of i.i.d. $N(0,3)$ draws. The error terms ϵ_1 and ϵ_2 are drawn from a multivariate normal distribution with

$$\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 25 & 10.5 \\ 10.5 & 9 \end{pmatrix} \right).$$

The resulting correlation matrix is

$$\begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}.$$

For this simulation study, we use a simple setup and create three different versions with regards to how the second equation for y_2 is constructed. In the first version, y_2 is simply determined by one instrument, z_{10}, and the error term. In the second version, y_2 is determined by z_{10}, but also by x_{10}, which we assume to be an exogenous regressor that is not part of the main equation. This reflects the situation, in which we assume a variable to be a potential exogenous regressor in the main equation when in fact it is an instrument. In the third version, additionally to x_{10} and z_{10}, y_2 is also determined by x_{12}, which is also part of the main equation. Here, we want to understand how the consistency of the estimate of a coefficient, which is exogenous in the main equation, but also related to one of the endogenous variables as described above, is affected when endogeneity is ignored.

For each version, we estimate the system for three different sample sizes $N = 100$, $N = 1000$ and $N = 10000$, setting the seed to one each time. We run the chain with 100000 iterations, of which the first 10000 are disregarded as burn-in.

In each version, the true underlying data generating process is a two-equation system in which y_2 is an endogenous variable in the structural equation for y_1. However, y_2 will wrongly be assumed to be exogeneous. Hence, the second equation will be disregarded. Instead, we will simply estimate the structural equation for y_1 using the standard linear regression BMA approach of Section 3 using the BMS R package by Zeugner and Feldkircher (2015). We will use the benchmark prior for g suggested by Fernández et al. (2001a) and random prior inclusion probabilities for the covariates with regards to the prior of the model space. We repeat the procedure for each setup 100 times and present the averaged results.

Variable	$\text{PIP}_{N=100}$	$\text{PIP}_{N=10000}$	True Value	$\text{Mean}_{N=100}$	$\text{Mean}_{N=10000}$
$\mathbf{y_2}$	1.00	1.00	1	**1.47**	**1.5**
$\mathbf{x_1}$	0.83	1.00	-1.8	-1.59	-1.79
$\mathbf{x_2}$	0.68	1.00	1.5	1.14	1.51
x_3	0.08	0.01	0	-0.01	0.00
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{10}	0.08	0.01	0	0.02	0.00
$\mathbf{x_{11}}$	1.00	1.00	1	1.00	1.00
$\mathbf{x_{12}}$	1.00	1.00	-1.5	-1.47	-1.5
x_{13}	0.07	0.01	0	0.00	0.00
x_{14}	0.09	0.01	0	0.01	0.00
x_{15}	0.06	0.01	0	0.00	0.00

Table 1: Estimation results for sample sizes $N = 100$ and $N = 10000$ for version one (4.8). Variables that truly are part of the model are marked in bold blue. PIP denotes the posterior inclusion probability and Mean denotes the posterior mean. Values marked in red indicate inconsistencies of the methodology in this scenario.

First Version

$$\begin{aligned} y_1 &= y_2 - 1.8x_1 + 1.5x_2 + x_{11} - 1.5x_{12} + \epsilon_1, \\ y_2 &= z_{10} + \epsilon_2. \end{aligned} \tag{4.8}$$

The estimation results of (4.8) using a standard BMA approach for linear regression models are depicted in Table 1. We observe that for $x_4, ..., x_9$ the results are as expected. All of these variables are neither part of the structural equation nor the instrumental equation and the methodology correctly picks that up. The posterior inclusion probabilities as well as the posterior means for the respective coefficients go to zero with larger sample size. As these particular results are not of great importance for this analysis we do not explicitly display them here. The full table, however, is available in Appendix B.[8]

In this setup, the BMA methodology is still able to pick up the correct inclusion probabilities for all variables. They converge to one for variables which are truly part of the model and go to zero for those variables which are not truly part of the model with N becoming larger. Further, the estimation results for the coefficients of the exogenous variables in the model, x_1, x_2, x_{11} and x_{12}, seem to be consistent as the posterior means converge to their true value with growing sample size. With regards to the posterior mean of the parameter of the endogenous variable y_2, we can see however, that it does not converge to the true value with growing sample size. Consequently, the estimation

[8]The same holds for the following two versions of the simulation study.

Variable	$\text{PIP}_{N=100}$	$\text{PIP}_{N=10000}$	True Value	$\text{Mean}_{N=100}$	$\text{Mean}_{N=10000}$
$\mathbf{y_2}$	1.00	1.00	1	**1.44**	**1.50**
$\mathbf{x_1}$	0.75	1.00	-1.8	-1.39	-1.80
$\mathbf{x_2}$	0.67	1.00	1.5	1.10	1.50
x_3	0.11	0.01	0	-0.01	0.00
⋮	⋮	⋮	⋮	⋮	⋮
x_{10}	**0.52**	**1.00**	0	**0.99**	**1.66**
$\mathbf{x_{11}}$	1.00	1.00	1	0.99	1.00
$\mathbf{x_{12}}$	1.00	1.00	-1.5	-1.46	-1.50
x_{13}	0.09	0.03	0	0.01	0.00
x_{14}	0.10	0.02	0	0.00	0.00
x_{15}	0.12	0.02	0	-0.02	0.00

Table 2: Estimation results for sample sizes $N = 100$ and $N = 10000$ for version two (4.9). Variables that truly are part of the model are marked in bold blue. PIP denotes the posterior inclusion probability and Mean denotes the posterior mean. Values marked in red indicate inconsistencies of the methodology in this scenario.

procedure in this case leads to biased inference for the endogenous variables in the model.

Second Version

$$\begin{aligned} y_1 &= y_2 - 1.8x_1 + 1.5x_2 + x_{11} - 1.5x_{12} + \epsilon_1, \\ y_2 &= -3.3x_{10} + z_{10} + \epsilon_2. \end{aligned} \tag{4.9}$$

Now, in (4.9) we include x_{10} in the equation determining y_2. As x_{10} is not part of the structural equation, it is rather an instrument in the system. However, we assume x_{10} to be a potential exogenous regressor in the structural equation to evaluate the consequences of misspecifying an instrument in this context. The estimation results for (4.9) are displayed in Table 2.

We observe inconsistent estimates of the posterior mean of the coefficient of the endogenous variable y_2 as in version one. Additionally, we see that the inclusion probability of x_{10} wrongly converges to one despite not being part of the true underlying structural equation. Further, the estimate of the posterior mean of the coefficient of x_{10} is also inconsistent. Therefore, the effects of unaccounted endogeneity on the estimation results are even more profound if the relationship between the endogenous variable(s) and potential exogenous regressors is more complex as it often is the case in real empirical applications.

Variable	$PIP_{N=100}$	$PIP_{N=10000}$	True Value	$Mean_{N=100}$	$Mean_{N=10000}$
$\mathbf{y_2}$	1.00	1.00	1	**1.40**	**1.50**
$\mathbf{x_1}$	0.83	1.00	-1.8	-1.55	-1.79
$\mathbf{x_2}$	0.71	1.00	1.5	1.17	1.49
x_3	0.15	0.02	0	-0.12	0.00
⋮	⋮	⋮	⋮	⋮	⋮
x_{10}	**0.60**	**1.00**	0	**1.11**	**1.65**
$\mathbf{x_{11}}$	1.00	1.00	1	1.05	1.00
$\mathbf{x_{12}}$	1.00	1.00	-1.5	**−1.88**	**−2.00**
x_{13}	0.09	0.02	0	0.00	0.00
x_{14}	0.09	0.01	0	0.00	0.00
x_{15}	0.11	0.01	0	0.01	0.00

Table 3: Estimation results for sample sizes $N = 100$ and $N = 10000$ for version three (4.10). Variables that truly are part of the model are marked in bold blue. PIP denotes the posterior inclusion probability and Mean denotes the posterior mean. Values marked in red indicate inconsistencies of the methodology in this scenario.

Third Version

$$\begin{aligned} y_1 &= y_2 - 1.8x_1 + 1.5x_2 + x_{11} - 1.5x_{12} + \epsilon_1, \\ y_2 &= -3.3x_{10} + x_{12} + z_{10} + \epsilon_2. \end{aligned} \tag{4.10}$$

In contrast to the second version, we further add the exogenous regressor x_{12} to the equation determining y_2. x_{12} is therefore part of the structural equation and instrumental equation. The estimation results of (4.10) are displayed in Table 3. We observe that, additionally to the consequences we could already see in version two, the estimate for the posterior mean of the coefficient of x_{12} is inconsistent as well. Hence, estimates for an exogenous regressor which is part of the main structural equation while also being part of the instrumental equation for the endogenous variable, will be biased as well if the endogeneity is not taken into account properly.

We have now established that ignoring endogeneity has severe consequences in the BMA framework. Not only do we get biased estimates of the coefficient(s) of the endogenous variable(s), but, depending on the underlying structure, also potentially false posterior inclusion probabilities and biased posterior estimates of parameters of other variables in the model.

It seems to be essential to correctly classify exogenous variables and instruments which only appear in the instrumental equation (see, e.g., Lenkoski et al., 2014). Obviously, however, it is even more important to specify the system correctly and take into account all the endogenous variables in the model. We thus need a methodology which can deal with this fundamental problem. However, in real empirical application, we also have to deal with general model uncertainty. In the next section, we introduce some of the existing approaches which try to account for the above described problem while also incorporating model uncertainty.

4.3 Existing Approaches for Incorporating Model Uncertainty in the IV Model Case

Hybrid Approaches

In Durlauf et al. (2008), the focus is on understanding broader growth theories rather than the statistical methodology itself. The authors use a heuristic hybrid approach and apply the Bayesian averaging of classical estimators (BACE) methodology to a 2SLS framework. They consider frequentist estimators and combine them with Bayesian weights, while applying the BIC approximation for the Bayes factors[9]. One specific variable, the log of real GDP, is assumed to be endogenous. Hence, the authors do not allow for uncertainty about the possible endogeneity of other variables. For the model space, a dilution prior is used, where different combinations of growth theories constitute distinct models. Durlauf et al. suggest that it is possible to interpret their estimation method as a weighting of the limited information maximum likelihood (LIML) estimator with limited information BIC (LIBIC) weights. It has to be noted, however, that in the 2SLS framework, there exists no formal justification so far for using the BIC approximation for the weights as in the linear regression model with normal errors. This implies that one must be careful when interpreting the results. A similar approach as in Durlauf et al. (2008) is used by Durlauf et al. (2012) for an exactly identified system, where the number of endogenous regressors equals the number of instruments. Again, the focus is on the application, which is the investigation of the influence of religion on growth.

Although technically only allowing for weak exogeneity, Moral-Benito (2012) suggests a heuristic method for dealing with endogeneity and model uncertainty in a panel data setting. In contrast to strict exogeneity, current shocks are allowed to affect the future. The author suggests to use a Bayesian averaging of maximum likelihood estimates (BAMLE) approach. Again, the BIC approximation for the marginal likelihood is used to construct posterior model weights. A hierarchical Binomial-Beta prior is used for prior model probabilities. Moral-Benito first considers a dynamic panel with country-specific effects and strictly exogenous regressors and later derives a likelihood function for a

[9]The derivation of the former can be found in Raftery (1995).

dynamic panel data model that relaxes the assumption of strict exogeneity of the lagged dependent variable. This allows to eliminate the bias of the within-group (WG) estimator for dynamic panel data models.

(Semi-)Bayesian Framework

Tsangarides (2004) develops a new Limited Information Bayesian Model Averaging Estimator (LIBMA) for a dynamic panel data setting with endogenous variables. The main feature of this methodology is the derivation of the likelihood. In contrast to standard BMA, LIBMA is a limited information approach that relies on moment restrictions in the context of GMM rather than a complete stochastic specification. To construct a Bayesian limited information procedure based on a set of moments, the author starts with a Bayesian estimator, which minimizes the expected posterior loss[10]. The loss function is chosen in a way, such that it yields an estimator equivalent to the GMM estimator

$$l(\theta,\delta) = [g_N(\theta) - g_N(\delta)]' W_N [g_N(\theta) - g_N(\delta)], \tag{4.11}$$

where N is the sample size and W_N is a weighting matrix, or

$$l(\theta,\delta) = [\theta - \delta]' \tilde{W}_N [\theta - \delta], \tag{4.12}$$

with $\tilde{W}_N = \{\frac{\partial g(\tilde{\theta})}{\partial \theta}\}' W_N \{\frac{\partial g(\tilde{\theta})}{\partial \theta}\}$ and $\tilde{\theta} \in (\theta,\delta)$. The author follows Kim (2002) in constructing a semiparametric limited information likelihood based on the moment conditions to derive a limited information posterior. Instead of using the BIC approximation to obtain posterior model probabilities, Tsangarides (2004) uses the LIBIC approximation, which is a limited information version. For the derivation of posterior model probabilities with LIBIC, the errors are assumed to be multivariate normal.

Another (semi-) Bayesian approach is the instrumental variables two-stage BMA (2SBMA) approach by Lenkoski et al. (2014) . The authors extend the methodology of Durlauf et al. (2008) and develop a formal statistical foundation. They show that 2SBMA is a consistent procedure, which reduces the 'too many instruments bias' as well as the mean-squared error compared to the classical 2SLS estimator[11]. The data is processed in two stages and model uncertainty is addressed in both.

[10]The expected posterior loss is the difference between the true posterior density of some quantity θ and a choice δ made by the researcher.

[11]A key concern in frequentist 2SLS regression is that $\hat{\beta}^{2SBMA}$ is biased and that the extent of the bias increases with the number of terms that are added in the first stage with coefficients equal or close to zero. 2SLS regression is essentially a large sample procedure, which in practice often is problematic.

The analysis is based on the restricted reduced form of the IV-model as displayed in equation (4.2). Using the notation from the latter we additionally define $T = (\Delta : \Lambda)'$ and $\rho = (\beta : \gamma)'$. Then, slightly rewriting (4.2) leads to

$$\begin{aligned} y_1 &= ((Z : X)T : X)\rho + v_1, \\ Y_2 &= (Z : X)T + V_2, \end{aligned}$$

where

$$\text{Var}\begin{pmatrix} v_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix} = \Omega.$$

The authors assume $m = 1$ for their analysis and define $B = (Z : X)$, $\tilde{U}(T) = (BT : X)$, $\hat{U} = \tilde{U}(\hat{T})$ with $\hat{T} = (B'B)^{-1}B'Y_2$ and $\phi = \omega_{22}^{-1}\omega_{21}$. The estimation of the parameters is a stepwise procedure. The first stage set of models for Y_2 is denoted by $M = \{M_1, ..., M_K\}$ and the set of models for y_1 in the second stage is denoted by $L = \{L_1, ..., L_J\}$. The two-stage marginal and conditional likelihoods conditional on the model M_k and L_j are then given by

$$\begin{aligned} L(\theta | Y_2, B, \omega_{22,k}) &= \omega_{22,k}^{-\frac{N}{2}} \exp\left(-\frac{1}{2} \frac{(Y_2 - B_k T_k)'(Y_2 - B_k T_k)}{\omega_{22,k}} \right), \\ L(\rho | \phi_{k,j}, \omega_{11.2,k,j}, y_1, \hat{T}_k) &= \omega_{11.2,k,j}^{-\frac{N}{2}} \exp\left(-\frac{1}{2} \frac{(y_1 - \hat{U}_j\rho_j + \hat{V}_{2,k}\phi_{k,j})'(y_1 - \hat{U}_j\rho_j + \hat{V}_{2,k}\phi_{k,j})}{\omega_{11.2,k,j}} \right), \end{aligned}$$

where $\hat{V}_{2,k} = Y_2 - V_k \hat{T}_k$. In this procedure, first the marginal likelihood and then the conditional likelihood is maximised. In contrast to the classical 2SLS, in this setup, the authors allow for the covariates in X to be in model M_k, but not in model L_j and vice versa. However, the latter case would lead to substantial bias if X actually had large explanatory power with respect to Y_2. According to Lenkoski et al. (2014), this would be visible in $\hat{\Delta}_{k,j}$ since a large value of the latter would imply a failure in model selection at the first stage. Lenkoski et al. (2014) construct a two-stage unit information prior (UIP), which can be seen as a special version of Zellner's g-prior. A key assumption here is the normality of the error terms. The UIP is a normal prior with a mean at the MLE and variance equal to the inverse of the average information contained in one observation.

Therefore, for

$$p(\rho_j, T_k, \omega_{11.2,k,j}, \omega_{22,k}, \phi_{k,j}) = p(\rho_j, \omega_{11.2,k,j}, \phi_{k,j}|T_k, \omega_{22,k})p(T_k, \omega_{22,k}), \qquad (4.13)$$

a UIP is specified on $p(T_k, \omega_{22,k})$ first and then, conditional on T_k, a UIP is specified on ρ_j. The authors show that their estimator asymptotically resembles the classical 2SLS estimator. Hence, it is named 2SBMA with

$$\hat{\rho}^{2SBMA} = \sum_{k=1}^{K}\sum_{j=1}^{J} \pi_k q_{j|k} \hat{\rho}_{k,j}, \qquad (4.14)$$

where $q_{j|k} = p(L_j|\theta_k, M_k, Y)$ is the posterior model probability of model L_j given model M_k. π_k is the posterior model probability of model M_k, such that $\hat{T}^{BMA} = \sum_{k=1}^{K} \pi_k \hat{T}_k$. The authors further introduce a Bayesian version of the Sargan test for overidentification and the Cragg and Donald test for weak instruments. Finally, the method is applied to investigate development determinants. As this is a two-step procedure, it does not treat the system as a whole and therefore does not use all the information available in the data efficiently.

Fully Bayesian Framework

Koop et al. (2012) design a special reversible jump Markov chain Monte Carlo algorithm (RJMCMC) to deal with the IV-model in a BMA framework. In this paper, only over-identified and just-identified models are considered[12]. The flexible approach the authors employ allows for uncertainty about which instruments to include (i.e. uncertainty about the column dimension of Δ), uncertainty about the exogenous regressors X, implying uncertainty about γ, uncertainty about the restrictions on endogenous regressors (i.e. some coefficients in β might be restricted to zero) and uncertainty about the exogeneity assumptions, meaning uncertainty about the elements of σ_{21} in that some of the covariances between ϵ_1 and V_2 might be zero, as well. The authors use the algorithm of Holmes et al. (2002), but since this does not work well due to the high correlation between the model and the covariance matrix of the errors, Koop et al. extend it with the idea of simulated tempering (ST) put forward by Marinari and Parisi (1992). The authors expand the model space with 'cold models' which are based on an approximation of the posterior and are not of intrinsic interest. According to Koop et al. (2012), 'cold models' merely facilitate the movement between the models of interest, which they call

[12]That implies $k_z \geq m$ while Δ has full rank.

’hot’ models. The latter are defined by a likelihood function and an adapted Drèze (1976) prior, which is used for the parameters. For the calculation of posterior features, only draws of hot models are used. This algorithm can be extended if a different prior than the adapted Drèze prior is used for the parameters. For that purpose, ’super hot models’ need to be introduced. The posterior of these ’super hot model’ equals the posterior of ’hot models’ times an additional term. The algorithm explores the joint posterior distribution of parameters and models and is, due to its flexible setup, quite powerful.

However, there are several disadvantages. The algorithm potentially has problems with convergence due to slow mixing. It is not very efficient as only a fraction of the draws are actually used for inference. The bigger issues with this approach, however, is the complexity of the algorithm. It is very hard to comprehend, not to mention to get an understanding of how different prior choices actually influence the results. The algorithm is written in GAUSS, which is not used as frequently as , e.g., R by the research community. Taking all these things into account, the approach by Koop et al. (2012) is potentially not the most recommendable one to be used by applied researchers.

Karl and Lenkoski (2012) suggest a rather different and much simpler approach, namely instrumental variables BMA. It will be discussed in detail in the following section.

5 The Karl and Lenkoski (2012) Approach and Some Extensions

Using conditional Bayes factors (CBFs), Karl and Lenkoski (2012) describe a very intuitive algorithm to deal with model uncertainty within the IV model framework in a Bayesian fashion. They call this approach IVBMA. In contrast to the 2SBMA version of Lenkoski et al. (2014), IVBMA is a fully Bayesian approach.

The Karl and Lenkoski (2012) approach is based on a MC3-within-Gibbs algorithm since model moves are embedded in a Gibbs sampler. The authors extend the Gibbs sampler discussed in Rossi et al. (2006) and use CBFs to calculate posterior model probabilities. CBFs are based on comparing two models in a hierarchical system, conditionally on parameters that are not influenced by the models considered. In this framework, the CBFs are essentially ratios of the normalizing constants of multivariate normal densities.

The advantage of the above approach is that, due to a simple and conjugate prior structure, analytical expressions for the CBFs are available. Further, there already exists an R package on which the empirical analysis can be built.

Karl and Lenkoski (2012) implement their algorithm using standard normal priors for the coefficient parameters of the system. Further, they consider an inverted Wishart prior for the covariance of the system with an identity matrix as centring matrix and the degrees of freedom set to the size of the system plus three. Setting up the system regarding the prior structure is indeed user friendly in that no hyper-parameters have to be elicited. However, the latter is also tricky as it is not invariant with respect to scaling.

In the following section, we derive the full conditionals and the conditional Bayes factors for a more general setting, in which we introduce a more flexible prior structure and, in addition, assume a general number of endogenous variables. We then continue by comparing various different prior structures based on estimation results in a simulation setting. Finally, we develop a procedure for determining whether a variable is truly endogenous or not. This will help to enable the researcher to specify the correctly-sized system and get more efficient and accurate estimation results.

5.1 Derivation of the Full Conditionals and the Conditional Bayes Factors

In the following derivations, we follow Karl and Lenkoski (2012) closely. However, we assume that the number of possible endogenous variables m can be larger than one. Further, we choose a more general setup for the prior structures.

We consider model (4.1), i.e.

$$y_1 = Y_2\beta + X\gamma + \epsilon_1, \tag{5.1}$$

$$Y_2 = Z\Delta + X\Lambda + V_2, \tag{5.2}$$

where we number the structural equation (5.1) and the instrumental equation(s) (5.2) separately in order to make it easier for the reader to follow the derivations. Additionally, we now define $\rho = [\beta' \; \gamma']'$ being of dimension $(m + k_X) \times 1$, $T = \text{vec}[\Delta' \; \Lambda']'$ being of dimension $(k_Z + k_X)m \times 1$ and $D = (y_1, Y_2, X, Z)$.

We will assume the following prior structures for the parameters of the above system:

$$p(\rho) \sim N(0, A), \tag{5.3}$$

$$p(T) \sim N(0, B \otimes I_m), \tag{5.4}$$

$$p(\Sigma) \sim IW(\nu, I_{m+1}), \tag{5.5}$$

with A being a $(m + k_X) \times (m + k_X)$ and B being a $(k_Z + k_X) \times (k_Z + k_X)$ positive definite and symmetric matrix, respectively. These priors are very similar to those assumed by Karl and Lenkoski (2012), but allow for more general prior covariances within the normal priors for ρ and vec(T).

Derivation of Full Conditionals

In the following, we derive the full conditionals $p(\rho|T, \Sigma, D)$, $p(T|\rho, \Sigma, D)$ and $p(\Sigma|\rho, T, D)$.

$p(\rho|T, \Sigma, D)$:

1. Rewrite (5.1) with $V = [Y_2 \; X]$, such that $y_1 = V\rho + \epsilon_1$. Conditional on T, we observe V_2 in (5.2).

2. Conditioning on V_2, such that $y_1 = V\rho + \epsilon_1$, for $\epsilon_{1j}|V'_{2i} \sim N(0 + \Sigma_{12}\Sigma_{22}^{-1}(V'_{2j} - 0), \sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$, we have, $\epsilon_1|V_2 \sim N((\Sigma_{12}\Sigma_{22}^{-1}V'_2)', (\sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})I_N)$.

3. We subtract the mean and define $\bar{\epsilon}_1 = \epsilon_1 - (\Sigma_{12}\Sigma_{22}^{-1}V'_2)'$ and $\tilde{y}_1 = y_1 - (\Sigma_{12}\Sigma_{22}^{-1}V'_2)'$.

4. We have $\tilde{y}_1 = V\rho + \bar{\epsilon}_1$ with $\bar{\epsilon}_1 \sim N(0, (\sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})I_N)$ and

   define $\xi = \sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, such that $\bar{\epsilon}_1 \sim N(0, \xi I_N)$.

5. We can formulate the conditional marginal posterior as

$$
\begin{aligned}
p(\rho|\tilde{y}_1, V, T, \Sigma) &\propto p(\tilde{y}_1|\rho, V, T, \Sigma)p(\rho), \\
p(\rho|\tilde{y}_1, V, T, \Sigma) &\propto (2\pi)^{-\frac{N}{2}}|\xi I_N|^{-\frac{1}{2}}(2\pi)^{-\frac{m+k_X}{2}}|A|^{-\frac{1}{2}} \\
&\quad \times \exp[-\tfrac{1}{2}((\tilde{y}_1 - V\rho)'(\xi I_N)^{-1}(\tilde{y}_1 - V\rho) + \rho' A^{-1}\rho)].
\end{aligned}
$$

6. Define $\hat{\rho} = \xi^{-1}\tilde{y}_1' V \Xi^{-1}$ and $\Xi = [\xi^{-1}V'V + A^{-1}]$. Then,

$$
\begin{aligned}
p(\rho|\tilde{y}_1, V, T, \Sigma) &\propto (2\pi)^{-\frac{N}{2}}|\xi I_N|^{-\frac{1}{2}}(2\pi)^{-\frac{m+k_X}{2}}|A|^{-\frac{1}{2}} \\
&\quad \times \exp[-\tfrac{1}{2}(\xi^{-1}\tilde{y}_1'\tilde{y}_1 - 2\hat{\rho}'\Xi'\rho + \rho'\Xi\rho)], \\
&\propto (2\pi)^{-\frac{N}{2}}|\xi I_N|^{-\frac{1}{2}}(2\pi)^{-\frac{m+k_X}{2}}|A|^{-\frac{1}{2}} \\
&\quad \times \exp[-\tfrac{1}{2}(\xi^{-1}\tilde{y}_1'\tilde{y}_1 - \hat{\rho}'\Xi\hat{\rho})]\exp[-\tfrac{1}{2}(\rho - \hat{\rho})'\Xi(\rho - \hat{\rho})], \\
&\propto (2\pi)^{-\frac{N}{2}}|\xi I_N|^{-\frac{1}{2}}(2\pi)^{-\frac{m+k_X}{2}}|A|^{-\frac{1}{2}} \\
&\quad \times \exp[-\tfrac{1}{2}\xi^{-1}\tilde{y}_1'\tilde{y}_1]\exp[\tfrac{1}{2}\hat{\rho}'\Xi\hat{\rho}]\exp[-\tfrac{1}{2}(\rho - \hat{\rho})'\Xi(\rho - \hat{\rho})].
\end{aligned}
$$

7. Hence, $\rho|T, \Sigma, D \sim N(\hat{\rho}, \Xi^{-1})$.

$p(T|\rho, \Sigma, D)$:

Here, we can not simply follow the same steps as in Karl and Lenkoski (2012). In order to obtain an expression for the conditional in this more general case, we need to re-write the system as follows:

1. Insert (5.2) into (5.1), such that $y_1 = (Z\Delta + X\Lambda + V_2)\beta + X\gamma + \epsilon_1$.

2. Rewriting leads to $y_1 = (Z\ :\ X)\binom{\Delta}{\Lambda}\beta + X\gamma + v_1$ with $v_1 = \epsilon_1 + V_2\beta$.

3. We are now using the fact that $y_1 = \text{vec}(y_1)$ since y_1 is a vector and $\text{vec}(ABC) = (C' \otimes A)\text{vec}B$ for compatible matrices A, B and C. Hence,

$$y_1 = (\beta' \otimes (Z\ :\ X))\,\text{vec}\begin{pmatrix}\Delta \\ \Lambda\end{pmatrix} + X\gamma + v_1.$$

Now, define $y_1^* = y_1 - X\gamma$. Accordingly,

$$y_1^* = (\beta' \otimes (Z\ :\ X))\,\text{vec}\begin{pmatrix}\Delta \\ \Lambda\end{pmatrix} + v_1.$$

4. Rewriting the second equation (5.2) leads to $Y_2 = (Z \ : \ X)\binom{\Delta}{\Lambda} + V_2$. This is equivalent to

$$\operatorname{vec} Y_2 = (I_m \otimes (Z \ : \ X)) \operatorname{vec} \begin{pmatrix} \Delta \\ \Lambda \end{pmatrix} + \operatorname{vec} V_2.$$

5. Now, we can combine the equations from 3. and 4. and formulate the system as

$$\begin{pmatrix} y_1^* \\ \operatorname{vec} Y_2 \end{pmatrix} = \left[\begin{pmatrix} \beta' \\ I_m \end{pmatrix} \otimes (Z \ : \ X) \right] \operatorname{vec} \begin{pmatrix} \Delta \\ \Lambda \end{pmatrix} + \begin{pmatrix} v_1 \\ \operatorname{vec} V_2 \end{pmatrix}.$$

6. By defining $\tilde{y} = \binom{y_1^*}{\operatorname{vec} Y_2}$, $W = \left[\binom{\beta'}{I_m} \otimes (Z \ : \ X) \right]$, $T = \operatorname{vec}\binom{\Delta}{\Lambda}$ and $\tilde{\epsilon} = \binom{v_1}{\operatorname{vec} V_2}$, we can rewrite the model in 5.

$$\tilde{y} = W \ T + \tilde{\epsilon},$$

with $\tilde{\epsilon} \sim N(0, \Omega \otimes I_N)$, where $\Omega = \begin{pmatrix} \sigma_{11} + \beta' \Sigma_{22} \beta + 2\Sigma_{12}\beta & \Sigma_{12} + \beta' \Sigma_{22} \\ \Sigma_{21} + \Sigma_{22}\beta & \Sigma_{22} \end{pmatrix}$.

7. The likelihood is given by

$$p(\tilde{y}|W, T, \Omega) = (2\pi)^{-\frac{(1+m)N}{2}} |\Omega \otimes I_N|^{-\frac{1}{2}} \exp[-\tfrac{1}{2}(\tilde{y} - WT)'(\Omega \otimes I_N)^{-1}(\tilde{y} - WT)],$$

and the prior is equal to

$$p(T) = (2\pi)^{-\frac{(k_Z + k_X)m}{2}} |B \otimes I_m|^{-\frac{1}{2}} \exp[-\tfrac{1}{2} T'(B \otimes I_m)^{-1} T].$$

8. Similar to the steps above, we can define $\Psi = W'\Omega^{-1} \otimes I_N W + B^{-1} \otimes I_m$ and $\hat{T} = \tilde{y}'\Omega^{-1} \otimes I_N W \Psi^{-1}$. Then,

$$\begin{aligned} p(T|\rho, \Sigma, D) \propto & (2\pi)^{-\frac{(1+m)N}{2}} |\Omega \otimes I_N|^{-\frac{1}{2}} (2\pi)^{-\frac{(k_Z+k_X)m}{2}} |B \otimes I_m|^{-\frac{1}{2}} \\ & \times \exp[-\tfrac{1}{2}\tilde{y}'\Omega^{-1} \otimes I_N \tilde{y}] \exp[\tfrac{1}{2}\hat{T}'\Psi\hat{T}] \exp[-\tfrac{1}{2}(T - \hat{T})'\Psi(T - \hat{T})]. \end{aligned}$$

9. Hence, $T|\rho, \Sigma, D \sim N(\hat{T}, \Psi^{-1})$.

$p(\Sigma|\rho, T, D)$:

The last conditional posterior quickly follows as $\Sigma|\rho, T, D \sim IW(\hat{\Sigma}, \nu + N)$, where $\hat{\Sigma} = (I_{1+m} + [\epsilon_1 V_2]'[\epsilon_1 V_2])$.

Introducing Model Uncertainty and Resulting Conditional Bayes Factors

Following Karl and Lenkoski (2012), we define $\mathscr{L}$ as a collection of models in which each $L \in \mathscr{L}$ has different restrictions on ρ, such that $\rho_L \subseteq \mathbb{R}^{m+k_X}$. Similarly, we define $\mathscr{M}$ as a collection of models in which each $M \in \mathscr{M}$ has different restrictions on T, such that $T_M \subseteq \mathbb{R}^{(k_Z+k_X)m}$. The dimension of Σ is kept fixed.

The conditional joint distribution of $\tilde{y}_1$ and ρ_L under model uncertainty is

$$\begin{aligned} p(\tilde{y}_1|\rho_L, V, T_M, \Sigma)p(\rho_L) =&(2\pi)^{-\frac{N}{2}}|\xi I_N|^{-\frac{1}{2}}(2\pi)^{-\frac{L}{2}}|A_L|^{-\frac{1}{2}}\exp[-\tfrac{1}{2}\xi^{-1}\tilde{y}_1'\tilde{y}_1]\exp[\tfrac{1}{2}\hat{\rho}_L'\Xi_L\hat{\rho}_L] \\ &\times \exp[-\tfrac{1}{2}(\rho_L-\hat{\rho}_L)'\Xi_L(\rho_L-\hat{\rho}_L)], \end{aligned}$$

while the conditional joint distribution of $\tilde{y}$ and T_M is given by

$$\begin{aligned} p(\tilde{y}|T_M, V, \rho_L, \Sigma)p(T_M) =&(2\pi)^{-\frac{(1+m)N}{2}}|\Omega\otimes I_N|^{-\frac{1}{2}}(2\pi)^{-\frac{Mm}{2}}|B_M\otimes I_m|^{-\frac{1}{2}} \\ &\times \exp[-\tfrac{1}{2}\tilde{y}'\Omega^{-1}\otimes I_N\tilde{y}]\exp[\tfrac{1}{2}\hat{T}_M'\Psi_M\hat{T}_M] \\ &\times \exp[-\tfrac{1}{2}(T_M-\hat{T}_M)'\Psi_M(T_M-\hat{T}_M)]. \end{aligned}$$

The conditional posterior model probabilities are equal to the conditional Bayes factor since, for now, we assume that each model is equally likely. Hence, prior model probabilities cancel out while comparing models with each other. Therefore, we need to calculate the conditional marginal likelihoods for each conditional posterior

$$\begin{aligned} \int p(\tilde{y}_1, \rho_L|V, T_M, \Sigma)\,\mathrm{d}\rho_L =& \int p(\tilde{y}_1|\rho_L, V, T_M, \Sigma)p(\rho_L)\,\mathrm{d}\rho_L, \\ =&(2\pi)^{-\frac{N}{2}}|\xi I_N|^{-\frac{1}{2}}(2\pi)^{-\frac{L}{2}}|A_L|^{-\frac{1}{2}}\exp[-\tfrac{1}{2}\xi^{-1}\tilde{y}_1'\tilde{y}_1]\exp[\tfrac{1}{2}\hat{\rho}_L'\Xi_L\hat{\rho}_L] \\ &\times \int \exp[-\tfrac{1}{2}(\rho_L-\hat{\rho}_L)'\Xi_L(\rho_L-\hat{\rho}_L)\,\mathrm{d}\rho_L, \\ =&(2\pi)^{-\frac{N}{2}}|\xi I_N|^{-\frac{1}{2}}(2\pi)^{-\frac{L}{2}}|A_L|^{-\frac{1}{2}}\exp[-\tfrac{1}{2}\xi^{-1}\tilde{y}_1'\tilde{y}_1]\exp[\tfrac{1}{2}\hat{\rho}_L'\Xi_L\hat{\rho}_L] \\ &\times (2\pi)^{\frac{L}{2}}|\Xi_L|^{-\frac{1}{2}}, \\ =&(2\pi)^{-\frac{N}{2}}|\xi I_N|^{-\frac{1}{2}}\exp[-\tfrac{1}{2}\xi^{-1}\tilde{y}_1'\tilde{y}_1]\exp[\tfrac{1}{2}\hat{\rho}_L'\Xi_L\hat{\rho}_L]\,|\Xi_L|^{-\frac{1}{2}}\,|A_L|^{-\frac{1}{2}}. \end{aligned}$$

Hence, by comparing model L_1 and L_2, we obtain

$$
\begin{aligned}
BF_{12} =& \frac{(2\pi)^{-\frac{N}{2}} |\xi I_N|^{-\frac{1}{2}} \exp[-\frac{1}{2}\xi^{-1}\tilde{y}_1'\tilde{y}_1] \exp[\frac{1}{2}\hat{\rho}_{L_1}'\Xi_{L_1}\hat{\rho}_{L_1}] |\Xi_{L_1}|^{-\frac{1}{2}} |A_{L_1}|^{-\frac{1}{2}}}{(2\pi)^{-\frac{N}{2}} |\xi I_N|^{-\frac{1}{2}} \exp[-\frac{1}{2}\xi^{-1}\tilde{y}_1'\tilde{y}_1] \exp[\frac{1}{2}\hat{\rho}_{L_2}'\Xi_{L_2}\hat{\rho}_{L_2}] |\Xi_{L_2}|^{-\frac{1}{2}} |A_{L_2}|^{-\frac{1}{2}}}, \\
=& \frac{\exp[\frac{1}{2}\hat{\rho}_{L_1}'\Xi_{L_1}\hat{\rho}_{L_1}] |\Xi_{L_1}|^{-\frac{1}{2}} |A_{L_1}|^{-\frac{1}{2}}}{\exp[\frac{1}{2}\hat{\rho}_{L_2}'\Xi_{L_2}\hat{\rho}_{L_2}] |\Xi_{L_2}|^{-\frac{1}{2}} |A_{L_2}|^{-\frac{1}{2}}}.
\end{aligned}
$$

If A_L is not equal to the identity matrix with dimension L as in Karl and Lenkoski (2012), we have to consider the extra term $|A_L|^{-\frac{1}{2}}$ in the marginal.

Performing similar calculations for the second conditional posterior yields

$$
\begin{aligned}
\int p(\tilde{y}, T_M|\rho_L, \Sigma)\, \mathrm{d}\, T_M =& \int p(\tilde{y}|T_M, \rho_L, V, \Sigma) p(T_M)\, \mathrm{d}\, T_M \\
=& (2\pi)^{-\frac{(1+m)N}{2}} |\Omega \otimes I_N|^{-\frac{1}{2}} (2\pi)^{-\frac{Mm}{2}} |B_M \otimes I_m|^{-\frac{1}{2}} \exp[-\tfrac{1}{2}\tilde{y}'\Omega^{-1} \otimes I_N \tilde{y}] \\
& \times \exp[\tfrac{1}{2}\hat{T}_M'\Psi_M \hat{T}_M] \int \exp[-\tfrac{1}{2}(T_M - \hat{T}_M)'\Psi_M(T_M - \hat{T}_M)]\, \mathrm{d}\, T_M \\
=& (2\pi)^{-\frac{(1+m)N}{2}} |\Omega \otimes I_N|^{-\frac{1}{2}} (2\pi)^{-\frac{Mm}{2}} |B_M \otimes I_m|^{-\frac{1}{2}} \exp[-\tfrac{1}{2}\tilde{y}'\Omega^{-1} \otimes I_N \tilde{y}] \\
& \times \exp[\tfrac{1}{2}\hat{T}_M'\Psi_M \hat{T}_M] (2\pi)^{\frac{Mm}{2}} |\Psi_M|^{-\frac{1}{2}} \\
=& (2\pi)^{-\frac{(1+m)N}{2}} |\Omega \otimes I_N|^{-\frac{1}{2}} \exp[-\tfrac{1}{2}\tilde{y}'\Omega^{-1} \otimes I_N \tilde{y}] \exp[\tfrac{1}{2}\hat{T}_M'\Psi_M \hat{T}_M] \\
& \times |\Psi_M|^{-\frac{1}{2}} |B_M \otimes I_m|^{-\frac{1}{2}} \\
=& (2\pi)^{-\frac{(1+m)N}{2}} |\Omega \otimes I_N|^{-\frac{1}{2}} \exp[-\tfrac{1}{2}\tilde{y}'\Omega^{-1} \otimes I_N \tilde{y}] \exp[\tfrac{1}{2}\hat{T}_M'\Psi_M \hat{T}_M] \\
& \times |\Psi_M|^{-\frac{1}{2}} |B_M|^{-\frac{m}{2}},
\end{aligned}
$$

where $|B_M|^{-\frac{m}{2}}$ is an extra term compared to the results of Karl and Lenkoski (2012), which is required if B is not the identity matrix. Finally, comparing two models M_1 and M_2 leads to:

$$
BF_{1,2} = \frac{\exp[\frac{1}{2}\hat{T}_{M_1}'\Psi_{M_1}\hat{T}_{M_1}] |\Psi_{M_1}|^{-\frac{1}{2}} |B_{M_1}|^{-\frac{m}{2}}}{\exp[\frac{1}{2}\hat{T}_{M_2}'\Psi_{M_2}\hat{T}_{M_2}] |\Psi_{M_2}|^{-\frac{1}{2}} |B_{M_2}|^{-\frac{m}{2}}}.
$$

For the general implementation algorithm, see Karl and Lenkoski (2012).

5.2 Extensions to the Prior Structure

In the previous subsection (5.1), we derived the conditionals and conditional BFs for a more general case. We now investigate how different settings for A and B from the prior structures (5.3) and (5.4), respectively, will affect the results and whether improvements, e.g, in terms of accuracy of the posterior mean, can be achieved. We start by comparing four different prior settings:

a: $A_L = I_L$ and $B_M = I_M$:

The first setting is the one considered by Karl and Lenkoski (2012). This means that, in both stages, structural and instrumental, an identity matrix of size L and M, respectively, is used as a scaling matrix in the prior distributions for the parameters.

b: $A_L = gI_L$ and $B_M = gI_M$:

Here, we introduce a shrinkage factor g, which will typically increase the prior variance, thus making the prior flatter and more uninformative. The Bayes factor for comparing two different models for the structural equation is

$$BF_{12} = \frac{\exp[\frac{1}{2}\hat{\rho}'_{L_1}\Xi_{L_1}\hat{\rho}_{L_1}]\,|\Xi_{L_1}|^{-\frac{1}{2}}\,|A_{L_1}|^{-\frac{1}{2}}}{\exp[\frac{1}{2}\hat{\rho}'_{L_2}\Xi_{L_2}\hat{\rho}_{L_2}]\,|\Xi_{L_2}|^{-\frac{1}{2}}\,|A_{L_2}|^{-\frac{1}{2}}},$$

and therefore proportional to $\frac{|A_{L_1}|^{-\frac{1}{2}}}{|A_{L_2}|^{-\frac{1}{2}}}$. The latter equals $\frac{|gI_{L_1}|^{-\frac{1}{2}}}{|gI_{L_2}|^{-\frac{1}{2}}}$, which reduces to $g^{\frac{L_2-L_1}{2}}$. Hence, if $g = 100$, a smaller model with $\dim(L_1) = \dim(L_2) - 1$ will be preferred by a factor of ten. The larger g, the more pronounced the shrinkage.

A similar argument can be made for the instrumental part of the system. There, the Bayes factor is proportional to $\frac{|B_{M_1}|^{-\frac{m}{2}}}{|B_{M_2}|^{-\frac{m}{2}}}$. Hence, this expression in the given prior setting equals $\frac{|gI_{M_1}|^{-\frac{m}{2}}}{|gI_{M_2}|^{-\frac{m}{2}}}$, which reduces to $g^{\frac{m(M_2-M_1)}{2}}$. Again, the shrinkage intensity increases with the number of endogenous variables m. In the same setting, with $g = 100$, $\dim(M_1) = \dim(M_2) - 1$ and, e.g., two endogenous variables, such that $m = 2$, the smaller model will be preferred by a factor of 100. Although there are many options regarding the choice of g, we set $g = N$ in our simulations, making it dependent on the sample size.

c: $A_L = g\xi([Y_2X]'_L[Y_2X]_L)^{-1}$ and $B_M = I_M$:

In this setting, we additionally introduce some structure for the covariances between the regressors scaled by the factor g. We set $g = N$ as in **b**, implying that g is constant across models.

ξ changes in each step. It introduces some information from the covariance of the system which can be potentially exploited. For the remaining instrumental equations of the system, we use an identity matrix for the prior variance as Karl and Lenkoski (2012) suggest.

d: $A_L = g\xi([Y_2 X]'_L[Y_2 X]_L)^{-1}$ and $B_M = gI_M$:

This setting is similar to the one in **c**, except that we introduce a scaling factor g for the prior covariance of the instrumental equations. Here, again, g is the same for all models and is set to $g = N$.

Influence of ν on Estimation Results of the Posterior

In our simulation study, we consider a three-dimensional system and first set $\nu = 4$, which corresponds to $\dim(\Sigma) + 1$. Repeating all simulations while setting $\nu = 8$, we investigate whether the choice of ν has a sizeable influence on the estimation results of the posterior. We observe that, for a sample size of at least $N = 100$, the influence of ν is small if not negligible with regards to the posterior. Hence, in the following subsections we only present and discuss the results for $\nu = 4$. The simulation results where $\nu = 8$ can be found in Appendix C.

Simulation Set Up

For our simulation study, we stick to the setup discussed in section 4.2, but we augment the latter by an additional equation. We consider the following three-equation system

$$\begin{aligned} y_1 &= y_2 - y_3 - 1.8x_1 + 1.5x_2 + x_{11} - 1.5x_{12} + \epsilon_1, \\ y_2 &= 1.1x_1 - 0.5x_3 + 0.75x_{12} + 0.75z_1 - 2z_8 + \epsilon_2, \\ y_3 &= z_{10} + \epsilon_3. \end{aligned} \tag{5.6}$$

We investigate how the various prior structures perform within two different setups. In the first case, we estimate a system which is properly specified in the sense that y_2 and y_3 are actually endogenous. To achieve the latter, ϵ_1, ϵ_2 and ϵ_3 are drawn from a multivariate normal distribution with

$$\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \epsilon_{3i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 25 & 1.0 & 10.5 \\ 1.0 & 4 & 1.8 \\ 10.5 & 1.8 & 9 \end{pmatrix}\right).$$

The implied correlation matrix then is

$$\begin{pmatrix} 1 & 0.1 & 0.7 \\ 0.1 & 1 & 0.3 \\ 0.7 & 0.3 & 1 \end{pmatrix}.$$

In the second case, we estimate a system in which y_2 and y_3 are wrongly specified as endogenous. Here, the error terms ϵ_1, ϵ_2 and ϵ_3 are drawn from a multivariate normal distribution with

$$\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \epsilon_{3i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 25 & 0 & 0 \\ 0 & 4 & 1.8 \\ 0 & 1.8 & 9 \end{pmatrix}\right),$$

which implies the following true correlation matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.3 \\ 0 & 0.3 & 1 \end{pmatrix}.$$

We compare the performance of the different prior structures (**a**-**d**) based on the averaged results of 100 repetitions for both cases and for different sample sizes of $N = 100$ and $N = 500$. It is to be expected that the differences we observe vanish with larger sample size as the influence of the prior becomes smaller. Finally, we repeat the simulations while scaling the data. Eventually, we give some recommendations as to which prior structure works best in a given setting.

As before, we reduce the display of our results in the following subsection as to not overwhelm the reader with too many tables. We show the interesting parts of the results for a sample size of $N = 100$. The complete tables as well as the results for larger sample sizes can be found in Appendix C.

Simulation Results for Non-Standardised Data

First Case - Correctly Specified System

In this paragraph, we look at the estimation results using the prior structures **a-d** discussed above while the system is correctly specified. y_2 is closer to being exogenous and y_3 is highly endogenous with both being indeed correctly categorized as such.

	PIP_a	PIP_b	PIP_c	PIP_d	True Value	$Mean_a$	$Mean_b$	$Mean_c$	$Mean_d$
$\mathbf{y_2}$	1.00	1.00	1.00	1.00	1	0.99	1.00	1.00	1.00
$\mathbf{y_3}$	1.00	1.00	1.00	1.00	-1	-0.97	-1.01	-0.95	-0.97
$\mathbf{x_1}$	0.96	0.94	0.94	0.97	-1.8	-1.45	-1.70	-1.70	-1.72
$\mathbf{x_2}$	0.91	0.87	0.85	0.91	1.5	1.15	1.34	1.30	1.37
x_3	0.38	0.08	0.18	0.17	0	-0.01	-0.02	-0.01	-0.03
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{10}	0.35	0.06	0.16	0.14	0	0.01	0.00	0.01	0.01
$\mathbf{x_{11}}$	1.00	1.00	1.00	1.00	1	1.01	1.02	1.02	1.02
$\mathbf{x_{12}}$	1.00	1.00	1.00	1.00	-1.5	-1.49	-1.52	-1.51	-1.51
x_{13}	0.19	0.03	0.19	0.19	0	0.00	0.00	0.00	0.00
x_{14}	0.17	0.02	0.16	0.15	0	0.00	0.00	0.00	0.00
x_{15}	0.19	0.02	0.17	0.16	0	-0.01	0.00	0.00	0.00

Table 4: Estimated posterior inclusion probabilities (PIP) and posterior mean (Mean) for variables in structural equation (5.6) for correctly specified system for different prior settings (**a**-**d**) based on an average of 100 samples of size $N = 100$. $\nu = 4$ in prior for Σ. Variables marked in bold blue are truly part of the model.

Table 4 depicts the estimated posterior inclusion probabilities (PIP) and the estimated posterior mean for each prior structure. The PIPs for the variables that are truly part of the model are all either one or close to one and all priors work equally well in that respect. However, with regards to variables that are not part of the model, in particular x_3 to x_{10}, the original prior by Karl and Lenkoski (2012) (setting **a**) does not perform as well as priors **b-d**. It seems that the prior does not impose enough complexity penalty in the overall setting and therefore tends to lead to larger PIPs for variables that are actually not part of the model. In that regard, prior structure **b** performs best. Given that we only used a relatively small sample size of $N = 100$, the estimates of the posterior mean (Mean) are quite close to the true values. The posterior means for the coefficients of variables that are not part of the model are consistently close to zero or zero for all prior structures. Overall, the estimates are very similar for priors **b-d**. Also, the estimates of the posterior mean for, e.g., x_1 and x_2 are better than the estimates under prior **a** as they are much closer to the true values.

Table 5 shows the estimated lower 0.05 and upper 0.95 quantiles of the posterior distributions of the parameters. For the original prior **a**, the quantiles are the widest. Therefore, the overall posterior estimation results are the least accurate for prior **a**. Prior **b** and **d** perform best, while prior **b** leads to the most accurate posterior results. Probably since, here, with $g = 100$, a higher complexity penalty is induced than, e.g., in prior **a**, where $g = 1$.

In Table 6, we see the estimated covariances for the entire system of equations. They are all fairly similar with the one under prior **b** being closest to the true one.

True Value	$q_{0.05,\mathbf{a}}$	$q_{0.95,\mathbf{a}}$	$q_{0.05,\mathbf{b}}$	$q_{0.95,\mathbf{b}}$	$q_{0.05,\mathbf{c}}$	$q_{0.95,\mathbf{c}}$	$q_{0.05,\mathbf{d}}$	$q_{0.95,\mathbf{d}}$
1	0.88	1.09	0.91	1.09	0.88	1.09	0.90	1.08
-1	-1.22	-0.75	-1.27	-0.77	-1.19	-0.73	-1.21	-0.75
-1.8	-2.14	-0.68	-2.37	-0.98	-2.48	-0.85	-2.37	-1.04
1.5	0.42	1.85	0.65	2.02	0.49	2.10	0.72	2.03
0	-0.46	0.45	-0.15	0.06	-0.26	0.27	-0.29	0.17
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	-0.38	0.43	-0.05	0.08	-0.20	0.25	-0.14	0.20
1	0.80	1.22	0.84	1.21	0.81	1.23	0.83	1.20
-1.5	-1.72	-1.26	-1.72	-1.32	-1.74	-1.28	-1.71	-1.30
0	-0.09	0.07	-0.01	0.01	-0.08	0.07	-0.08	0.06
0	-0.07	0.07	0.00	0.01	-0.07	0.07	-0.05	0.06
0	-0.08	0.07	-0.01	0.00	-0.08	0.07	-0.07	0.06

Table 5: Estimated lower 0.05 quantile ($q_{0.05}$) and upper 0.95 quantile ($q_{0.95}$) for coefficients of variables in structural equation (5.6) for correctly specified system for different prior settings (**a**-**d**) based on an average of 100 samples of size $N = 100$. $\nu = 4$ in prior for Σ.

$\hat{\Sigma}_{\mathbf{a}}$	$\hat{\Sigma}_{\mathbf{b}}$	$\hat{\Sigma}_{\mathbf{c}}$	$\hat{\Sigma}_{\mathbf{d}}$
$\begin{pmatrix} 24.09 & 0.84 & 9.82 \\ 0.84 & 3.84 & 1.60 \\ 9.82 & 1.60 & 8.34 \end{pmatrix}$	$\begin{pmatrix} 25.27 & 0.85 & 10.33 \\ 0.85 & 4.22 & 1.68 \\ 10.33 & 1.68 & 8.66 \end{pmatrix}$	$\begin{pmatrix} 23.71 & 0.83 & 9.67 \\ 0.83 & 3.83 & 1.60 \\ 9.67 & 1.60 & 8.34 \end{pmatrix}$	$\begin{pmatrix} 24.18 & 0.84 & 9.98 \\ 0.84 & 4.22 & 1.69 \\ 9.98 & 1.69 & 8.65 \end{pmatrix}$
$\hat{\Sigma}^{\text{corr}}_{\mathbf{a}}$	$\hat{\Sigma}^{\text{corr}}_{\mathbf{b}}$	$\hat{\Sigma}^{\text{corr}}_{\mathbf{c}}$	$\hat{\Sigma}^{\text{corr}}_{\mathbf{d}}$
$\begin{pmatrix} 1 & 0.09 & 0.69 \\ 0.09 & 1 & 0.28 \\ 0.69 & 0.28 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.07 & 0.67 \\ 0.07 & 1 & 0.26 \\ 0.67 & 0.26 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.14 & 0.65 \\ 0.14 & 1 & 0.29 \\ 0.65 & 0.29 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.15 & 0.65 \\ 0.15 & 1 & 0.29 \\ 0.65 & 0.29 & 1 \end{pmatrix}$

Table 6: Estimated posterior means of covariance matrices and corresponding correlation matrices in structural equation (5.6) for correctly specified system for different prior settings (**a**-**d**) based on an average of 100 samples of size $N = 100$. $\nu = 4$ in prior for Σ.

	PIP_a	PIP_b	PIP_c	PIP_d	True Value	$Mean_a$	$Mean_b$	$Mean_c$	$Mean_d$
$\mathbf{y_2}$	1.00	1.00	1.00	1.00	1	0.98	0.99	0.98	0.98
$\mathbf{y_3}$	0.99	0.99	0.99	0.99	-1	-0.97	-0.98	-0.97	-0.97
$\mathbf{x_1}$	0.90	0.77	0.83	0.83	-1.8	-1.28	-1.49	-1.55	-1.55
$\mathbf{x_2}$	0.83	0.64	0.71	0.71	1.5	1.01	1.09	1.16	1.17
x_3	0.42	0.11	0.17	0.18	0	0.01	0.00	0.01	0.00
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_{10}	0.39	0.08	0.15	0.15	0	0.02	0.01	0.01	0.01
$\mathbf{x_{11}}$	0.99	0.99	0.98	0.99	1	1.01	1.04	1.03	1.03
$\mathbf{x_{12}}$	1.00	1.00	1.00	1.00	-1.5	-1.48	-1.53	-1.52	-1.52
x_{13}	0.24	0.04	0.18	0.18	0	0.00	0.00	0.00	0.00
x_{14}	0.22	0.03	0.15	0.15	0	0.00	0.00	0.00	0.00
x_{15}	0.23	0.04	0.17	0.17	0	-0.01	0.00	0.01	-0.01

Table 7: Estimated posterior inclusion probabilities (PIP) and posterior mean (Mean) for variables in structural equation (5.6) for different prior setting based on an average of 100 samples of size $N = 100$ and $\nu = 4$ in prior for Σ. For incorrectly specified system. Variables marked in bold blue are truly part of the model.

Second Case - Wrongly Assumed Endogeneity

We investigate how the different priors perform under an unnecessarily complex system. Both y_2 and y_3 are not truly endogenous, but are assumed to be. We therefore expect to see some kind of efficiency loss of the estimation procedure, i.e., in terms of accuracy of the estimates of the posterior mean or the PIPs.

Table 7 displays the estimated posterior inclusion probabilities and the posterior means for the different prior structures. We see that, in comparison to a correctly specified system, the results seem to converge somewhat slower to the true values. The PIP for, e.g., x_1 and x_2 reduces for all priors. For variables that are not part of the model, the PIP increases slightly, which occurs mostly for priors **a** and **b**.

The posterior means are also not as close to the true value as in the above case of a correctly specified system. Here, prior **c** and **d** perform best. Overall, the latter seem to be more robust in that the efficiency loss seems to be somewhat smaller. It has to be noted that this problem becomes less pronounced with larger sample sizes. As mentioned before, results for $N = 500$ can be found in Appendix C. Since in practice, however, the researcher often deals with smaller sample sizes, it is useful to keep the aforementioned effect in mind while specifying a model.

In Table 8, we see the lower 0.05 and upper 0.95 quantiles. The differences between the various prior structures is not big, except for prior **a**, which implies wider quantile ranges, for instance, for x_3 to x_{10}. Overall, the posterior distributions are less peaked

True Value	$q_{0.05,\mathbf{a}}$	$q_{0.95,\mathbf{a}}$	$q_{0.05,\mathbf{b}}$	$q_{0.95,\mathbf{b}}$	$q_{0.05,\mathbf{c}}$	$q_{0.95,\mathbf{c}}$	$q_{0.05,\mathbf{d}}$	$q_{0.95,\mathbf{d}}$
1	0.85	1.10	0.86	1.12	0.85	1.11	0.85	1.11
-1	-1.22	-0.72	-1.25	-0.72	-1.23	-0.71	-1.23	-0.71
-1.8	-2.12	-0.45	-2.53	-0.56	-2.56	-0.63	-2.55	-0.62
1.5	0.22	1.85	0.27	2.08	0.31	2.16	0.32	2.15
0	-0.53	0.60	-0.17	0.23	-0.27	0.36	-0.28	0.38
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	-0.48	0.55	-0.11	0.15	-0.23	0.28	-0.23	0.29
1	0.74	1.28	0.77	1.31	0.76	1.30	0.76	1.30
-1.5	-1.76	-1.19	-1.81	-1.24	-1.80	-1.23	-1.80	-1.23
0	-0.14	0.12	-0.02	0.03	-0.10	0.09	-0.10	0.09
0	-0.12	0.12	-0.01	0.01	-0.09	0.09	-0.08	0.09
0	-0.14	0.11	-0.03	0.02	-0.11	0.08	-0.10	0.08

Table 8: Estimated lower 0.05 quantile ($q_{0.05}$) and upper 0.95 quantile ($q_{0.95}$) for coefficients of variables in structural equation (5.6) for different prior setting based on an average of 100 samples of size $N = 100$ and $\nu = 4$ in prior for Σ. For incorrectly specified system.

$\hat{\Sigma}_{\mathbf{a}}$	$\hat{\Sigma}_{\mathbf{b}}$	$\hat{\Sigma}_{\mathbf{c}}$	$\hat{\Sigma}_{\mathbf{d}}$
$\begin{pmatrix} 24.87 & -0.02 & -0.14 \\ -0.02 & 3.85 & 1.68 \\ -0.14 & 1.68 & 8.19 \end{pmatrix}$	$\begin{pmatrix} 25.42 & -0.05 & -0.02 \\ -0.05 & 4.02 & 1.73 \\ -0.02 & 1.73 & 8.42 \end{pmatrix}$	$\begin{pmatrix} 25.01 & -0.03 & -0.16 \\ -0.03 & 3.85 & 1.68 \\ -0.16 & 1.68 & 8.19 \end{pmatrix}$	$\begin{pmatrix} 25.03 & -0.04 & -0.13 \\ -0.04 & 4.23 & 1.74 \\ -0.13 & 1.74 & 8.56 \end{pmatrix}$
$\hat{\Sigma}_{\mathbf{a}}^{\text{corr}}$	$\hat{\Sigma}_{\mathbf{b}}^{\text{corr}}$	$\hat{\Sigma}_{\mathbf{c}}^{\text{corr}}$	$\hat{\Sigma}_{\mathbf{d}}^{\text{corr}}$
$\begin{pmatrix} 1 & 0.00 & -0.01 \\ 0.00 & 1 & 0.30 \\ -0.01 & 0.30 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.00 & 0.00 \\ 0.00 & 1 & 0.30 \\ 0.00 & 0.30 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.00 & -0.01 \\ 0.00 & 1 & 0.30 \\ -0.01 & 0.30 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.00 & -0.01 \\ 0.00 & 1 & 0.29 \\ -0.01 & 0.29 & 1 \end{pmatrix}$

Table 9: Estimated posterior means of the covariance matrices and corresponding correlation matrices for structural equation (5.6) under different simulation settings based on an average of 100 samples of size $N = 100$ and $\nu = 4$ in prior for Σ. For incorrectly specified system.

and therefore the estimates of the posterior less accurate in comparison to the correctly specified system.

Table 9 displays the estimated covariances. For prior **a**,**c** and **d**, the results are very similar. $\hat{\Sigma}_b$ is closest to the true covariance structure and therefore prior **b** is slightly preferable to the others in that regard.

Simulation Results for Standardised Data

How these priors perform will, particularly for prior structure **a** and **b**, depend on the scaling of the data. We now investigate how the results compare if we standardize the data before the estimation procedure. We de-mean the data and divide each variable by its standard deviation.

In order to be able to compare the estimation results for the different prior settings for scaled data, we would need to re-scale the posterior of the coefficients since we do not know the true value of the parameters in the model for the scaled data. However, as it is not completely straightforward to re-scale in this case, we compare the estimation

	PIP_a	PIP_b	PIP_c	PIP_d	$Mean_{N=5000}$	$Mean_a$	$Mean_b$	$Mean_c$	$Mean_d$
$\mathbf{y_2}$	1.00	1.00	1.00	1.00	0.76	0.76	0.76	0.75	0.75
$\mathbf{y_3}$	1.00	1.00	1.00	1.00	-0.45	-0.43	-0.43	-0.42	-0.42
$\mathbf{x_1}$	0.91	0.75	0.95	0.94	-0.18	-0.17	-0.14	-0.17	-0.16
$\mathbf{x_2}$	0.79	0.65	0.87	0.90	0.15	0.12	0.11	0.13	0.13
x_3	0.08	0.01	0.16	0.16	0.00	0.00	0.00	0.00	0.00
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_{10}	0.06	0.01	0.14	0.13	0.00	0.00	0.00	0.00	0.00
$\mathbf{x_{11}}$	1.00	1.00	1.00	1.00	0.34	0.35	0.35	0.34	0.34
$\mathbf{x_{12}}$	1.00	1.00	1.00	1.00	-0.51	-0.51	-0.51	-0.51	-0.50
x_{13}	0.09	0.01	0.18	0.17	0.00	0.00	0.00	0.00	0.00
x_{14}	0.07	0.01	0.15	0.14	0.00	0.00	0.00	0.00	0.00
x_{15}	0.08	0.01	0.16	0.15	0.00	0.00	0.00	0.00	0.00

Table 10: Estimated posterior inclusion probabilities (PIP) and posterior mean (Mean) for variables in structural equation (5.6) for correctly specified system and **scaled** data for different prior settings based on an average of 100 samples of size $N = 100$. $\nu = 4$ in prior for Σ. Variables marked in bold blue are truly part of the model.

results for a smaller sample size of $N = 100$ with the estimation results for $N = 5000$[13] since we can assume posterior consistency.

First Case - Correctly Specified System

The results for the posterior inclusion probabilities and the posterior mean estimates for a correctly specified system and a sample size of $N = 100$ can be found in Table 10. We can see that for standardized data the different priors lead to quite similar results in terms of posterior mean estimates. The differences are marginal. The posterior inclusion probabilities are similar, as well. Prior **d** is slightly better in including the true variables in the model than the others. Prior **b**, however, is best in reflecting which variables are not part of the model.

Table 11 shows the estimated quantiles under the different prior settings. The results are again very similar and no tangible differences can be detected between the priors with regards to how pronounced the quantile range of the posterior distribution is.

[13]We use ten samples of this size and average over the results. It is worth mentioning that, for sample sizes this large, the prior has no influence anymore and the results are indistinguishable for different prior settings.

$\text{Mean}_{N=5000}$	$q_{0.05,\mathbf{a}}$	$q_{0.95,\mathbf{a}}$	$q_{0.05,\mathbf{b}}$	$q_{0.95,\mathbf{b}}$	$q_{0.05,\mathbf{c}}$	$q_{0.95,\mathbf{c}}$	$q_{0.05,\mathbf{d}}$	$q_{0.95,\mathbf{d}}$
0.76	0.68	0.83	0.68	0.83	0.67	0.83	0.68	0.82
-0.45	-0.54	-0.33	-0.55	-0.33	-0.52	-0.32	-0.53	-0.32
-0.18	-0.24	-0.08	-0.22	-0.07	-0.24	-0.09	-0.23	-0.09
0.15	0.05	0.20	0.04	0.18	0.05	0.20	0.07	0.20
0.00	-0.01	0.01	0.00	0.00	-0.03	0.02	-0.03	0.02
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.00	-0.01	0.01	0.00	0.00	-0.02	0.02	-0.01	0.02
0.34	0.28	0.42	0.28	0.41	0.27	0.41	0.28	0.41
-0.51	-0.59	-0.43	-0.58	-0.44	-0.58	-0.43	-0.58	-0.43
0.00	-0.01	0.01	0.00	0.00	-0.03	0.02	-0.03	0.02
0.00	-0.01	0.01	0.00	0.00	-0.02	0.02	-0.02	0.02
0.00	-0.01	0.01	0.00	0.00	-0.02	0.02	-0.02	0.02

Table 11: Estimated lower 0.05 quantile ($q_{0.05}$) and upper 0.95 quantile ($q_{0.95}$) for variables in structural equation (5.6) for correctly specified system and **scaled** data for different prior settings based on an average of 100 samples of size $N = 100$. $\nu = 4$ in prior for Σ.

$\hat{\Sigma}_{\mathbf{a}}$	$\hat{\Sigma}_{\mathbf{b}}$	$\hat{\Sigma}_{\mathbf{c}}$	$\hat{\Sigma}_{\mathbf{d}}$
$\begin{pmatrix} 0.24 & 0.01 & 0.21 \\ 0.01 & 0.08 & 0.05 \\ 0.21 & 0.05 & 0.41 \end{pmatrix}$	$\begin{pmatrix} 0.25 & 0.01 & 0.22 \\ 0.01 & 0.09 & 0.05 \\ 0.22 & 0.05 & 0.42 \end{pmatrix}$	$\begin{pmatrix} 0.24 & 0.01 & 0.20 \\ 0.01 & 0.08 & 0.05 \\ 0.20 & 0.05 & 0.41 \end{pmatrix}$	$\begin{pmatrix} 0.24 & 0.01 & 0.21 \\ 0.01 & 0.09 & 0.05 \\ 0.21 & 0.05 & 0.42 \end{pmatrix}$
$\hat{\Sigma}^{\text{corr}}_{\mathbf{a}}$	$\hat{\Sigma}^{\text{corr}}_{\mathbf{b}}$	$\hat{\Sigma}^{\text{corr}}_{\mathbf{c}}$	$\hat{\Sigma}^{\text{corr}}_{\mathbf{d}}$
$\begin{pmatrix} 1 & 0.08 & 0.66 \\ 0.08 & 1 & 0.26 \\ 0.66 & 0.26 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.06 & 0.66 \\ 0.06 & 1 & 0.25 \\ 0.66 & 0.25 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.08 & 0.66 \\ 0.08 & 1 & 0.26 \\ 0.66 & 0.26 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.07 & 0.66 \\ 0.07 & 1 & 0.25 \\ 0.66 & 0.25 & 1 \end{pmatrix}$

Table 12: Estimated posterior means of the covariance matrices and corresponding correlation matrices for structural equation (5.6) for correctly specified system and **scaled** data for different prior settings based on an average of 100 samples of size $N = 100$. $\nu = 4$ in prior for Σ.

The estimated posterior mean of the covariance matrix using a large sample of $N = 5000$ equals

$$\hat{\Sigma}_{N=5000} = \begin{pmatrix} 0.25 & 0.01 & 0.23 \\ 0.01 & 0.07 & 0.05 \\ 0.23 & 0.05 & 0.43 \end{pmatrix}.$$

In Table 12, we see the estimated covariance matrices for the different prior structures for a sample size of $N = 100$. All of them are fairly similar with $\hat{\Sigma}_{\mathbf{b}}$ being the closest to $\hat{\Sigma}_{N=5000}$. Prior **b** therefore leads to somewhat more accurate results in that respect.

Second Case - Wrongly Assumed Endogeneity

As we did for the non-standardised data, we want to investigate the effect of estimating an unnecessarily big system of equations by assuming y_2 and y_3 to be endogenous even though they are not.

	$PIP_\mathbf{a}$	$PIP_\mathbf{b}$	$PIP_\mathbf{c}$	$PIP_\mathbf{d}$	$Mean_{N=5000}$	$Mean_\mathbf{a}$	$Mean_\mathbf{b}$	$Mean_\mathbf{c}$	$Mean_\mathbf{d}$
$\mathbf{y_2}$	1.00	1.00	1.00	1.00	0.70	0.69	0.69	0.68	0.68
$\mathbf{y_3}$	1.00	0.98	1.00	1.00	-0.42	-0.41	-0.40	-0.40	-0.40
$\mathbf{x_1}$	0.75	0.48	0.82	0.81	-0.16	-0.13	-0.09	-0.14	-0.14
$\mathbf{x_2}$	0.61	0.33	0.69	0.70	0.13	0.09	0.06	0.10	0.10
x_3	0.10	0.01	0.17	0.18	0.00	0.00	0.00	0.00	0.00
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{10}	0.07	0.01	0.14	0.15	0.00	0.00	0.00	0.00	0.00
$\mathbf{x_{11}}$	1.00	0.99	1.00	1.00	0.31	0.32	0.32	0.32	0.32
$\mathbf{x_{12}}$	1.00	1.00	1.00	1.00	-0.47	-0.47	-0.47	-0.47	-0.47
x_{13}	0.11	0.01	0.18	0.18	0.00	0.00	0.00	0.00	0.00
x_{14}	0.09	0.01	0.15	0.15	0.00	0.00	0.00	0.00	0.00
x_{15}	0.10	0.02	0.16	0.16	0.00	0.00	0.00	0.00	0.00

Table 13: Estimated posterior inclusion probabilities (PIP) and posterior mean (Mean) for variables in structural equation (5.6) for different prior settings based on an average of 100 samples of size $N = 100$ and $\nu = 4$ in prior for Σ and scaled data and incorrectly specified system. Variables marked in bold blue are truly part of the model.

We observe similar effects as we have with non-standardised data. Table 13 displays the PIPs and the posterior mean for the different prior structures. Overall, the accuracy decreases for all priors. The posterior inclusion probabilities for the variables that are truly part of the model, e.g. x_1 and x_2, are smaller for all priors, but particularly for prior **b**. Further, the posterior means are not as close to their true value as it is the case for a correctly specified system. Here, prior **b** seems to perform worst, as well.

Table 14 displays the upper and lower quantile of the posterior. The spread of the distribution is as wide as it is for a correctly specified system, but less symmetric around the true value.

If we estimate the posterior mean of the covariance using a large sample with $N = 5000$, we have the following result

$$\hat{\Sigma}_{N=5000} = \begin{pmatrix} 0.21 & 0.00 & 0.00 \\ 0.00 & 0.07 & 0.05 \\ 0.00 & 0.05 & 0.43 \end{pmatrix}.$$

As we assume posterior consistency, we conclude that the closer the estimated covariance is to the aforementioned one for a small sample size, the better the corresponding prior performs. Table 15 shows the estimated covariances for $N = 100$. They are very similar for the different priors, with the one under prior **d** perhaps being closest to the covariance given above.

$\text{Mean}_{N=5000}$	$q_{0.05,\mathbf{a}}$	$q_{0.95,\mathbf{a}}$	$q_{0.05,\mathbf{b}}$	$q_{0.95,\mathbf{b}}$	$q_{0.05,\mathbf{c}}$	$q_{0.95,\mathbf{c}}$	$q_{0.05,\mathbf{d}}$	$q_{0.95,\mathbf{d}}$
0.76	0.60	0.78	0.59	0.78	0.59	0.77	0.59	0.77
-0.45	-0.51	-0.30	-0.52	-0.29	-0.51	-0.29	-0.51	-0.29
-0.18	-0.23	-0.05	-0.19	-0.03	-0.23	-0.05	-0.23	-0.05
0.15	0.02	0.19	0.01	0.14	0.03	0.19	0.03	0.19
0.00	-0.01	0.02	0.00	0.00	-0.02	0.03	-0.03	0.03
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.00	-0.01	0.01	0.00	0.00	-0.02	0.03	-0.02	0.03
0.34	0.24	0.41	0.24	0.41	0.23	0.40	0.23	0.40
-0.51	-0.56	-0.38	-0.56	-0.38	-0.56	-0.38	-0.56	-0.38
0.00	-0.02	0.02	0.00	0.00	-0.03	0.03	-0.03	0.03
0.00	-0.01	0.02	0.00	0.00	-0.03	0.03	-0.03	0.03
0.00	-0.02	0.01	0.00	0.00	-0.03	0.02	-0.03	0.02

Table 14: Estimated lower 0.05 quantile ($q_{0.05}$) and upper 0.95 quantile ($q_{0.95}$) for coefficients of variables in structural equation (5.6) for different prior settings based on an average of 100 samples of size $N = 100$ and $\nu = 4$ in prior for Σ and scaled data, without endogeneity.

$\hat{\Sigma}_{\mathbf{a}}$	$\hat{\Sigma}_{\mathbf{b}}$	$\hat{\Sigma}_{\mathbf{c}}$	$\hat{\Sigma}_{\mathbf{d}}$
$\begin{pmatrix} 0.22 & 0.00 & 0.00 \\ 0.00 & 0.08 & 0.05 \\ 0.00 & 0.05 & 0.41 \end{pmatrix}$	$\begin{pmatrix} 0.23 & 0.00 & 0.00 \\ 0.00 & 0.09 & 0.05 \\ 0.00 & 0.05 & 0.42 \end{pmatrix}$	$\begin{pmatrix} 0.21 & 0.00 & 0.00 \\ 0.00 & 0.08 & 0.05 \\ 0.00 & 0.05 & 0.41 \end{pmatrix}$	$\begin{pmatrix} 0.21 & 0.00 & 0.00 \\ 0.00 & 0.09 & 0.05 \\ 0.00 & 0.05 & 0.42 \end{pmatrix}$
$\hat{\Sigma}^{\text{corr}}_{\mathbf{a}}$	$\hat{\Sigma}^{\text{corr}}_{\mathbf{b}}$	$\hat{\Sigma}^{\text{corr}}_{\mathbf{c}}$	$\hat{\Sigma}^{\text{corr}}_{\mathbf{d}}$
$\begin{pmatrix} 1 & 0.00 & 0.00 \\ 0.00 & 1 & 0.27 \\ 0.00 & 0.27 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & -0.01 & 0.00 \\ -0.01 & 1 & 0.26 \\ 0.00 & 0.26 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.00 & 0.00 \\ 0.00 & 1 & 0.27 \\ 0.00 & 0.27 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.00 & 0.00 \\ 0.00 & 1 & 0.26 \\ 0.00 & 0.26 & 1 \end{pmatrix}$

Table 15: Estimated posterior means of the covariance matrices and corresponding correlation matrices for structural equation (5.6) under different simulation settings based on an average of 100 samples of size $N = 100$ and $\nu = 4$ in prior for Σ and scaled data and incorrectly specified system.

Main Results of This Simulation Study

The previous results suggest that, if the data is not standardised, the adapted priors **b**, **c** and **d** clearly improve the estimation results in terms of more accurate PIPs and slightly better point estimates of the posterior mean. When it comes to correctly estimating the posterior inclusion probabilities, prior **b** performed best in this study. With regards to accurate posterior mean estimates and credible intervals, prior **d** yielded the best results.

If we standardise the data, the differences between the priors are not that noticeable. However, since in this case the results are not as interpretable as when we use non-standardised variables, it might not be very practical to perform standardisation in empirical applications. Therefore, the researcher has the option to use prior **d**, for instance, along with the original data.

Another thing that became clear is that there is something to gain by trying to specify the system correctly. Wrongly assuming exogeneity leads to inconsistent results as we have seen previously in section 4.2. However, wrongly assuming endogeneity also comes with a cost and that is less accurate results due to efficiency loss. It would therefore be very useful to have some kind of tool available that helps the researcher to determine whether a certain variable should be classified as endogenous or not.

5.3 Savage-Dickey Density Ratio for Endogeneity

An applied researcher's reflex might be to assume endogeneity and to estimate the 'bigger system' to avoid inconsistencies if there is any doubt about the exogeneity assumption of certain variables in the model. However, imposing more structure and complexity that is not needed leads to the inefficient use of data and less accurate estimation results and can significantly increase the computational effort required.

We therefore suggest to use a Savage-Dickey density ratio (SD-ratio) to determine whether a variable is endogenous or not based on the work of Dickey (1971). Verdinelli and Wassermann (1995) and Wagenmakers et al. (2010) excellently portray the intuition behind the SD-ratio. The basically is the ratio of the prior over the posterior at a certain point. The idea can be described more technically in the following way. Assume that θ represents the entire set of parameters in our model and ω and ϕ are subsets of θ, such that $\phi = \theta \setminus \omega$. We want to test the model $M_0 : \omega = \omega_0$ against the Model $M_1 : \omega \neq \omega_0$, where ω_0 constitutes a specific value. We now define two priors:

- p is the prior of $\theta = (\omega, \phi)$ under M_1 with prior density $p(\omega, \phi)$.

- p_0 is the prior of ϕ under M_0 with the condition $p_0(\phi) = p(\phi|\omega = \omega_0)$. When $\omega = \omega_0$, the prior of ϕ under M_1 should be equal to the prior of ϕ under M_0. Then, the Savage-Dickey density ratio for comparing M_0 and M_1 is basically equal to (5.7) if ϕ fulfils exactly the same role under M_0 and M_1.

If $L(\cdot,\cdot)$ is the corresponding likelihood function and D denotes the data, we have

$$\text{SD-ratio} = BF_{\omega_0,\omega} = \frac{p(D|H_0)}{p(D|H_1)} = \frac{\int L(\omega_0,\phi)p_0(\phi)\,d\phi}{\int\int L(\omega,\phi)p(\omega,\phi)\,d\omega\,d\phi} = \frac{p(\omega=\omega_0|D,H_1)}{p(\omega=\omega_0|H_1)}. \tag{5.7}$$

Then, the SD-ratio (5.7) implies that the BF for $\omega = \omega_0$ is computed by dividing the density value of the marginal posterior of ω by the density value of the prior of ω at the point of interest ω_0 (see Wagenmakers et al., 2010).

In our example, testing for endogeneity means testing whether specific elements of Σ are zero. We have that Σ is a $(1+m)\times(1+m)$ matrix and we are interested in testing whether $\sigma_{1+m,1} = 0$. If that were the case, we could conclude that y_{1+m} is not endogenous.

Hence, we need to evaluate the marginal prior and posterior of $\sigma_{1+m,1}$ at zero and then calculate the ratio between the density value of the posterior and the density value of the prior. If the resulting BF is big enough, such that we can argue that there is substantial evidence for H_0, we can reduce the system of equations by one and assign y_{1+m} to the set of exogenous regressors. Kass and Raftery (1995) provide some guidelines regarding the interpretation of the magnitude of Bayes-factors.

Prior to computing the density ratio, we have that $\Sigma \sim IW(I_{1+m},\nu)$. If we then reduce the system by one equation, we have that Σ_{red} now is a matrix of dimension $m\times m$. As Σ_{red} is a submatrix of Σ, we need to ensure that the prior of Σ_{red} is the same as the marginal original prior for that submatrix. Therefore, the induced prior for Σ_{red} is obtained by conditioning the prior of Σ on $\sigma_{1,1+m} = 0$, such that $\Sigma_{\text{red}} \sim IW(I_m, \nu-1)$.

For a general number of potentially endogenous variables, we suggest a sequential procedure, starting with the smallest estimated correlation between the structural variable and the assumed endogenous variable. In addition, we can see that the prior choice for the parameter of interest is crucial and influential in this kind of test. The degrees of freedom, e.g., have a huge influence on how flat the prior is and therefore on the BF. The subsequent simulation studies and results are helpful in terms of giving more general suggestions regarding reasonable choices for the prior.

Simulation Study for the Savage-Dickey Density Ratio

We start by assuming a $k-$equation system. That is, one main equation of interest and $k-1=m$ possibly endogenous variables. We want to examine whether the smallest covariance of any of the potentially endogenous variables with the main dependent variable is actually zero, implying exogeneity rather than endogeneity.

We have $\Sigma \sim IW_k(Q,\nu)$, where ν are the degrees of freedom and Q is a $k \times k$ matrix. We assume that ν is an integer to facilitate the implementation and that $\nu \geq k$, such that we have a proper distribution.

Σ, with $k = k_1 + k_2$, can be partitioned, such that

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

with Σ_{11} being a $k_1 \times k_1$ matrix, Σ_{12} being a $k_1 \times k_2$ matrix ($\Sigma_{21} = \Sigma'_{12}$) and Σ_{22} being a $k_2 \times k_2$ matrix. For our setting, Σ_{11} will be a scalar, hence $\Sigma_{11} = \sigma_{11}$. Consequently, Σ_{21} will be a $(k-1) \times 1$ vector and Σ_{22} a $(k-1) \times (k-1)$ matrix. We are interested in the marginal density of $\Sigma_{21} = (\sigma_{21}\sigma_{31}...\sigma_{k1})'$, and in particular the element with the smallest corresponding correlation in absolute terms, which we will assume to be σ_{k1}. As the order of the equations is irrelevant and they can be rearranged, this assumption is just for notational convenience. From that point on, we can follow a 'testing down' approach.

Non-Parametric Approach

Since we do not know the marginal density of σ_{k1}, we need to find a way to estimate $P(\sigma_{k1} = 0)$. The first approach that comes to mind is to generate a sample of Σ, extract σ_{k1} each time and use the generated sample to estimate the density and the density value at zero via non-parametric density estimation. In this context, it is well known that the choice of the bandwidth can have a considerable influence on the estimation results. In order to obtain an idea of the accuracy of the estimate, the computational cost and the influence of the number of degrees of freedom, we start by drawing samples from the prior for σ_{12} based on a two-equation system. More precisely we draw samples from an $IW(I_2,\nu)$ distribution with varying ν.

Table 16 reports the results for the estimated density value at zero for 2 to 10 degrees of freedom using a rule of thumb bandwidth which relies upon a normality assumption of the underlying density. The estimates do change somewhat with varying sample sizes

	ν								
draws	2	3	4	5	6	7	8	9	10
10000	0.32	1.09	2.04	3.14	4.43	5.79	7.37	8.91	10.68
50000	0.38	1.15	2.11	3.22	4.51	5.90	7.47	9.03	10.76
100000	0.38	1.17	2.15	3.27	4.56	5.99	7.48	9.07	10.78

Table 16: Density value estimates at zero for the marginal prior density of σ_{21} from a two-equation system with 10000, 50000 and 100000 draws based on non-parametric density estimation using a rule-of -thumb bandwidth (normal reference) and varying degrees of freedom (2-10).

	ν								
	2	3	4	5	6	7	8	9	10
$P(\sigma_{21} = 0)$	~ 0	0.25	0.97	1.51	2.86	4.74	6.32	7.38	9.94
$P(\sigma_{31} = 0)$	-	~ 0	0.27	0.89	1.98	2.74	4.59	6.75	7.96

Table 17: Density value estimates at zero for the marginal prior density of σ_{21} from a two-equation system and σ_{31} from a three-equation system based on non-parametric density estimation where the bandwidth is chosen via cross-validation criterion and using **10000 draws**.

(number of draws) and the results indicate that we need a rather big sample size to obtain more reliable results. This, in turn, is computationally quite expensive.

In the next step, we estimate $P(\sigma_{21} = 0)$ based on a two-equation system for varying degrees of freedom as well as $P(\sigma_{31} = 0)$ based on a three-equation system and varying degrees of freedom ($\Sigma \sim IW(I_3, \nu)$) using a cross-validation criterion to determine the bandwidth. We want to see how much influence the bandwidth choice actually has on the results. For this estimation, only a sample of 10000 draws was used as the procedure is very expensive computationally. The results can be found in Table 17.

If we compare the results for the two-equation system ($P(\sigma_{21}) = 0$) and a sample size of 10000 draws, we observe that they differ quite a bit. Further, it is not immediately clear which results are more accurate in being closer to the true value within this setting. Hence, we have no means to decide which bandwidth to choose.

It is possible that the number of draws is not sufficient to accurately estimate the density. This is true especially for lower degrees of freedom since, in that case, the probability mass is more spread and less concentrated, such that we do not have enough observations for the region of interest. For the posterior, especially in the case of endogeneity, the density value at zero can be quite small. In the latter case, non-parametric density estimation is likely to perform poorly since it is usually more precise in areas with higher probability mass where most of the observations lie. In this situation, it seems that we need many more observations, such that the estimates eventually converge to the true value. However, given that this procedure is quite time-consuming, just increasing the number of draws does not seem to be a feasible solution to the problem.

Hence, it appears that kernel density estimation is not very well-suited to deliver the best results in the given setting. Therefore, it would be beneficial to find another way to obtain estimations for the marginal density.

Rao-Blackwell Type Approach

Another option is to consider a type of estimation which is often referred to as 'Rao-Blackwell' estimation (see, e.g., Robert, 2007). Here, we can exploit the properties of conditional densities.

We have that $\Sigma \sim IW_k(Q, \nu)$. By partitioning Q, where Q is a $k \times k$ positive-definite symmetric matrix, in the same fashion as Σ, we can use the following implementation strategy[14]:

1. $\Sigma_{11} \sim IW_{k_1}(Q_{11}, \nu - k_2)$
2. Define $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ with $\Sigma_{22.1} \sim IW_{k_2}(Q_{22} - Q_{21}Q_{11}^{-1}Q_{12}, \nu)$,
3. $\Sigma_{11}^{-1}\Sigma_{12}|\Sigma_{22.1} \sim MN_{k_1,k_2}(Q_{11}^{-1}Q_{12}, \Sigma_{22.1} \otimes Q_{11}^{-1})$, and thus
   $\Sigma_{12}|\Sigma_{11}, \Sigma_{22.1} \sim MN_{k_1,k_2}(\Sigma_{11}Q_{11}^{-1}Q_{12}, \Sigma_{22.1} \otimes \Sigma_{11}Q_{11}^{-1}\Sigma_{11}')$.

Now, we can calculate the density value at zero for the element of interest of Σ_{12} by using conditional normality as well as by repeatedly drawing values for Σ_{11} and $\Sigma_{22.1}$, plugging the latter in and averaging over the results. Since we have a known form for the distribution of interest, the results for this approach should be rather exact.

In our case the expressions simplify somewhat. σ_{11} is a scalar and follows an Inverse-Gamma distribution, which is a special case of the Inverse-Wishart distribution[15].

1. $\sigma_{11} \sim IG(\frac{\nu-(k-1)}{2}, \frac{q_{11}}{2})$.
2. Define $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\sigma_{11}^{-1}\Sigma_{12}$ with $\Sigma_{22.1} \sim IW_{k-1}(Q_{22} - Q_{21}q_{11}^{-1}Q_{12}, \nu)$.
3. $\sigma_{11}^{-1}\Sigma_{12}|\Sigma_{22.1} \sim N_{k-1}(q_{11}^{-1}Q_{12}, \Sigma_{22.1}q_{11}^{-1})$, such that
   $\Sigma_{12}|\sigma_{11}, \Sigma_{22.1} \sim N_{k-1}(\sigma_{11}q_{11}^{-1}Q_{12}, \Sigma_{22.1}\sigma_{11}q_{11}^{-1}\sigma_{11}')$.

In order to obtain marginal **posterior** estimates, we can apply the same algorithm, starting from an IW distribution for Σ with hyper-parameters updated by the sample. See subsection 5.1. From the estimation results, we can then extract $\hat{\Sigma}$ and proceed in the following way:

[14]This follows standard results which can be found, e.g., in Drèze and Richard (1983).

[15]More information on the Inverse-Gamma distribution can be found in Appendix A.

draws	2	3	4	5	6	7	8	9	10
10000	0.49	1.28	2.24	3.43	4.69	6.15	7.71	9.28	11.13
50000	0.50	1.27	2.25	3.40	4.70	6.11	7.65	9.31	11.09
100000	0.50	1.27	2.26	3.39	4.68	6.10	7.66	9.30	11.08

Table 18: Density value estimates at zero for the marginal prior density of σ_{21} from a two-equation system with 10000, 50000 and 100000 draws obtained by algorithm described above and for varying degrees of freedom (2-10).

$\Sigma \sim IW_k(\hat{\Sigma}, \nu + N)$ and $\hat{\Sigma} = \begin{pmatrix} \hat{\sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix}$, implying that

1. $\sigma_{11} \sim IG(\frac{\nu+N-(k-1)}{2}, \frac{\hat{\sigma}_{11}}{2})$.

2. $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\sigma_{11}^{-1}\Sigma_{12}$ with $\Sigma_{22.1} \sim IW_{k-1}(\hat{\Sigma}_{22} - \hat{\Sigma}_{21}\hat{\sigma}_{11}^{-1}\hat{\Sigma}_{12}, \nu + N)$.

3. $\sigma_{11}^{-1}\Sigma_{12}|\Sigma_{22.1} \sim N_{k-1}(\hat{\sigma}_{11}^{-1}\hat{\Sigma}_{21}, \Sigma_{22.1}\hat{\sigma}_{11}^{-1})$, and thus

   $\Sigma_{12}|\sigma_{11}, \Sigma_{22.1} \sim N_{k-1}(\sigma_{11}\hat{\sigma}_{11}^{-1}\hat{\Sigma}_{21}, \Sigma_{22.1}\sigma_{11}^{2}\hat{\sigma}_{11}^{-1})$.

5.3.1 Marginal Prior Simulations

In the beginning, we need to establish how accurate the simulation results are depending on the sample size. We first implement the algorithm for a two-equation system and for varying degrees of freedom, where $\Sigma \sim IW_2(I_2, \nu)$. Table 18 shows the estimation results for $p(\sigma_{21} = 0)$ for three different sample sizes (number of draws). Even for a smaller sample size of 10000, the estimate already seems to be quite exact as the results are nearly the same for a large sample size of 100000.

We can now proceed by analysing the prior behaviour in more detail using a sample size of 50000. Table 19 shows the marginal prior density estimation results for σ_{k1} at the point zero based on a k-equation system for varying degrees of freedom, where $\Sigma \sim IW_k(I_k, \nu)$. The first distinctive feature is the diagonal structure, which reflects a sort of similar behaviour or shape independent from the size of the system. The latter is a very useful property as we can use this characteristic to generate other density values for even bigger systems without having to simulate the data each time.

The second important aspect is that we can observe that, the higher the number of the degrees of freedom, the more peaked the prior distribution becomes around zero, which implies stronger prior beliefs about y_k being exogenous. For the Bayes factor to favour the hypothesis that y_k is indeed exogenous, the marginal posterior density value of σ_{k1} at zero needs to be large. Since we will look at the ratio between marginal **posterior** density value at 0 (perhaps less affected by ν) and marginal **prior** density

					ν				
k	2	3	4	5	6	7	8	9	10
2	0.49	1.27	2.25	3.39	4.70	6.14	7.65	9.32	11.09
3	-	0.50	1.27	2.26	3.38	4.69	6.16	7.65	9.30
4	-	-	0.49	1.27	2.25	3.40	4.69	6.12	7.65
5	-	-	-	0.50	1.27	2.24	3.39	4.69	6.11
6	-	-	-	-	0.50	1.27	2.24	3.41	4.69
7	-	-	-	-	-	0.51	1.28	2.23	3.40
8	-	-	-	-	-	-	0.49	1.27	2.26
9	-	-	-	-	-	-	-	0.49	1.26
10	-	-	-	-	-	-	-	-	0.49

Table 19: Density value estimates at zero for the marginal prior density of σ_{k1} of a k-equation system and for varying degrees of freedom (2-10). $\Sigma \sim IW_k(I_k, \nu)$.

value at zero (directly affected by ν), the choice of the number of degrees of freedom will be important.

The central question is whether the above properties stem from the fact that we use the identity matrix as a centring matrix. We therefore need to continue to analyse results from slightly different configurations.

In order to determine how much the shape and therefore the estimation results for a particular density value of the marginal is driven by the choice of the centring matrix in the prior, we investigate four additional cases. In case a), we set the off-diagonal elements to 0.3 instead of zero. This reflects a prior belief of mild endogeneity of the variables we categorize as such and mild correlation among themselves. In case b), we adjust the values such that a prior belief of mild endogeneity and moderate correlation among the potentially endogenous variables is represented. In c), we set all off-diagonal elements to 0.5, which in turn, will reflect a prior belief about moderate endogeneity and moderate correlation among the variables.

Table 20 depicts simulation results for density estimates of $p(\sigma_{k1} = 0)$ when $\Sigma \sim IW_k(Q_{k,a}, \nu)$, with

$$Q_{k,a} = \begin{pmatrix} 1 & 0.3 & \cdots & 0.3 \\ 0.3 & 1 & \cdots & 0.3 \\ \vdots & \vdots & \ddots & \vdots \\ 0.3 & 0.3 & \cdots & 1 \end{pmatrix}.$$

				ν					
k	2	3	4	5	6	7	8	9	10
2	0.46	1.11	1.86	2.69	3.52	4.38	5.25	6.09	6.91
3	-	0.45	1.11	1.86	2.68	3.53	4.39	5.24	6.09
4	-	-	0.45	1.10	1.86	2.68	3.54	4.40	5.24
5	-	-	-	0.45	1.11	1.87	2.69	3.53	4.40

Table 20: Density value estimates at zero for the marginal prior density of σ_{k1} of a k-equation system and varying degrees of freedom (2-5). $\Sigma \sim IW_k(Q_{k,a}, \nu)$.

				ν					
k	2	3	4	5	6	7	8	9	10
2	0.46	1.11	1.86	2.69	3.52	4.38	5.25	6.09	6.91
3	-	0.45	1.10	1.87	2.68	3.53	4.39	5.25	6.09
4	-	-	0.45	1.11	1.87	2.69	3.53	4.37	5.26
5	-	-	-	0.45	1.11	1.86	2.69	3.54	4.40

Table 21: Density value estimates at zero for the marginal prior density of σ_{k1} of a k-equation system and varying degrees of freedom (2-5) . $\Sigma \sim IW_k(Q_{k,b}, \nu)$.

Table 21 shows simulation results for density estimates of $p(\sigma_{k1} = 0)$ when $\Sigma \sim IW_k(Q_{k,b}, \nu)$, with

$$Q_{k,b} = \begin{pmatrix} 1 & 0.3 & \cdots & 0.3 \\ 0.3 & 1 & \cdots & 0.5 \\ \vdots & \vdots & \ddots & \vdots \\ 0.3 & 0.5 & \cdots & 1 \end{pmatrix}.$$

Table 22 displays simulation results for density estimates of $p(\sigma_{k1} = 0)$ when $\Sigma \sim IW_k(Q_{k,c}, \nu)$, with

$$Q_{k,c} = \begin{pmatrix} 1 & 0.5 & \cdots & 0.5 \\ 0.5 & 1 & \cdots & 0.5 \\ \vdots & \vdots & \ddots & \vdots \\ 0.5 & 0.5 & \cdots & 1 \end{pmatrix}.$$

We can see that, in all three cases, the diagonal structure is preserved, which is why only results up to a system size of $k = 5$ are depicted in the above tables. For larger systems, we can extrapolate and use the fact that, for a $k-$dimensional system, the density value at zero for, e.g., k degrees of freedom is (almost) the same for all k[16]. Further, the higher the assumed degree of endogeneity, that is, the higher the values

[16]Small differences in the results are most likely due to simulation inaccuracies.

	ν								
k	2	3	4	5	6	7	8	9	10
2	0.38	0.83	1.27	1.66	1.98	2.23	2.42	2.56	2.62
3	-	0.38	0.83	1.26	1.65	1.97	2.23	2.42	2.55
4	-	-	0.37	0.82	1.27	1.66	1.98	2.23	2.43
5	-	-	-	0.38	0.82	1.27	1.65	1.98	2.23

Table 22: Density value estimates at zero for the marginal prior density of σ_{k1} of a k-equation system and varying degrees of freedom (2-5). $\Sigma \sim IW_k(Q_{k,c}, \nu)$.

in the first column and row, the flatter the marginal prior seems to be. The degrees of freedom have less influence on the shape than in the case where the centring matrix is the identity. It is interesting to note that the assumed correlation between the endogenous variables does not affect this behaviour as we see no changes in the estimated density values between the results in Table 20 and Table 21.

In the next step, we will analyse how the marginal posterior and therefore the density value is influenced by the choice of prior for Σ. We will compare the results using the prior suggested by Karl and Lenkoski (2012) as a benchmark with $\Sigma \sim IW_k(Q_{k,c}, \nu)$. The latter has the advantage that our prior choice with regards to the number of degrees of freedom will not deeply impact the results of the BF as the marginal prior density value does not change considerably with varying degrees of freedom. However, the researcher needs to decide a priori about the direction of the endogeneity as she needs to determine whether to center the prior over positive, zero or negative correlation between the endogenous and structural variable.

5.3.2 Marginal Posterior Simulations

The three-equation simulation is based on the simulation setup in (4.2). We use the following set of equations:

$$y_1 = y_2 - y_3 - 1.8x_1 + 1.5x_2 + x_{11} - 1.5x_{12} + \epsilon_1,$$
$$y_2 = 1.1x_1 - 0.5x_3 + 0.75x_{12} + 0.75z_1 - 2z_8 + \epsilon_2,$$
$$y_3 = z_{10} + \epsilon_3.$$

The error terms ϵ_1, ϵ_2 and ϵ_3 are drawn from a multivariate normal distribution with zero mean and, depending on the simulation setting, varying covariance matrices. The implied correlation matrix of the error terms has the following structure

$$\begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & 0.3 \\ \rho_{31} & 0.3 & 1 \end{pmatrix}.$$

As there will be some variability with respect to the sample and, consequently, regarding the results, we repeat the simulation procedure for each setting 100 times in order to make our inference more robust. Consequently, the results shown in the following tables are the average estimates over all the runs.

In practice, a sequential procedure would probably be the most obvious approach to test for endogeneity and in order to be able to reduce the system of equations. The question is how to determine which variable should be tested first and how to proceed afterwards. In reality, there is of course no special ordering of the equations and solely looking at the covariances might be misleading because we need to somehow take into account the scale of the variances, as well. Therefore, we will not just inspect the smallest covariance, but we will consider $\hat{\Sigma}$ and calculate the estimated correlation matrix. Then, we will choose the variable with the smallest correlation with y_1 in absolute terms. For the SD ratio itself, the marginal posterior density of the covariance of that variable will be used. The analysis for a two-equation system can be found in Appendix D. Below, we discuss the results for a three-equation system as it is potentially more interesting. The main conclusions, however, are the same.

Tables 23 - 27 display the density value at zero of the marginal posterior density of σ_{13} and the corresponding SD ratio value for different degrees of freedom, under different priors for Σ and different priors for the model parameters.

We begin with a setting where the endogeneity is quite strong for y_2 as well as for y_3. Therefore, the smallest true correlation is $\rho_{13} = 0.6$. The marginal density values and the corresponding SD ratios are displayed in Table 23. We can see that, even for a small sample size of $N = 100$ and all the different prior settings, the endogeneity is clearly picked up as the density values at zero are all virtually zero.

Table 24 depicts the results for the case in which y_2 is highly endogenous with $\rho_{12} = 0.7$ and y_3 is mildly endogenous with $\rho_{13} = 0.3$. As the true correlation is smaller than in the previous setting, the marginal density value of the covariance of σ_{13} is not quite zero for a smaller sample size of $N = 100$. However, for a larger sample size, this weak endogeneity is clearly picked up anyway as we can observe in Table 25. Here, we

Setting with $\rho_{12} = 0.7$ and $\rho_{13} = 0.6$							
		marginal posterior at zero			SD-ratio		
prior for Σ	ν	I-I	g-g	gprvar-g	I-I	g-g	gprvar-g
$\Sigma \sim IW(I_3, \nu)$	3	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	6	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	10	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
$\Sigma \sim IW(Q_{3,c}, \nu)$	3	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	6	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	10	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0

Table 23: Density value of the marginal posterior of σ_{13} at zero and **SD-ratio** from a three-equation system for a sample size of $N = 100$.

Setting with $\rho_{12} = 0.7$ and $\rho_{13} = 0.3$							
		marginal posterior at zero			SD-ratio		
prior for Σ	ν	I-I	g-g	gprvar-g	I-I	g-g	gprvar-g
$\Sigma \sim IW(I_3, \nu)$	3	2.82	2.18	2.57	5.64	4.36	5.14
	6	2.87	2.28	2.70	0.85	0.67	0.80
	10	3.04	2.41	2.87	0.33	0.26	0.31
$\Sigma \sim IW(Q_{3,c}, \nu)$	3	2.81	2.17	2.57	7.39	5.71	6.78
	6	2.86	2.27	2.69	1.73	1.38	1.63
	10	3.04	2.40	2.86	1.19	0.94	1.12

Table 24: Density value of the marginal posterior of σ_{13} at zero and **SD-ratio** from a three-equation system for a sample size of $N = 100$.

see the results for the same setting, but for a sample size of $N = 500$. The estimate of the marginal density seems to have converged as the density, in this case, is much more peaked and concentrated around the true value.

Another interesting feature is the similarity between the results for the marginal density estimates between the different priors for Σ. Interestingly, this is also the case when $\rho_{12} = -0.7$ and $\rho_{13} = -0.3$. The results for the latter scenario can be found in appendix D in Table A.48. Even if the direction of the endogeneity is specified wrongly in the prior for Σ, the influence on the marginal posterior density is mild as a prior with a centring matrix like $Q_{3,c}$ is flatter than the prior using an identity.

It is therefore recommendable to use a prior similar to $\Sigma \sim IW(Q_{3,c}, \nu)$ since it has the advantage that the influence of the choice of ν on the SD ratio is much smaller. At the same time, similarly good estimation results for the posterior of Σ are implied as compared to the approach of using the identity as centring matrix in the prior.

In Table 26 and Table 27, we see the marginal density estimates of σ_{13} at zero for $N = 100$ and $N = 500$, respectively, when no endogeneity is present. We can observe that, for $N = 100$, the density value of the marginal posterior of σ_{13} at zero is not close to zero anymore as was to be expected. The larger the sample, the more concentrated

Setting with $\rho_{12} = 0.7$ and $\rho_{13} = 0.3$							
		marginal posterior at zero			SD-ratio		
prior for Σ	ν	I-I	g-g	gprvar-g	I-I	g-g	gprvar-g
$\Sigma \sim IW(I_3, \nu)$	3	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	6	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	10	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
$\Sigma \sim IW(Q_{3,c}, \nu)$	3	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	6	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	10	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0

Table 25: Density value of the marginal posterior of σ_{13} at zero and **SD-ratio** from a three-equation system for a sample size of $N = 500$.

Setting with $\rho_{12} = 0.4$ and $\rho_{13} = 0$							
		marginal posterior at zero			SD-ratio		
prior for Σ	ν	I-I	g-g	gprvar-g	I-I	g-g	gprvar-g
$\Sigma \sim IW(I_3, \nu)$	3	17.46	17.01	17.09	34.92	34.02	34.18
	6	18.68	18.18	18.25	5.53	5.38	5.40
	10	20.37	19.77	19.84	2.19	2.13	2.13
$\Sigma \sim IW(Q_{3,c}, \nu)$	3	17.46	17.00	17.09	45.95	44.74	44.97
	6	18.69	18.16	18.25	11.33	11.01	11.06
	10	18.90	19.75	19.84	7.41	7.75	7.78

Table 26: Density value of the marginal posterior of σ_{13} at zero and **SD-ratio** from a three-equation system for a sample size of $N = 100$.

the density mass becomes around the true value of zero as depicted for $N = 500$. Again, there are no big differences in terms of scale between the results for the different priors of the model parameters I-I, g-g and gprvar-g.

One particular challenge is the choice of ν in the prior. We see that, with larger ν, the posterior becomes more peaked, but overall, the influence on the posterior is not crucial. This is different with regards to the prior, however, and thus the choice of ν will have a noticeable influence on the SD-ratio. If we choose ν too large, the resulting Bayes factor becomes easily unfavourable towards the hypothesis that the variable we are testing is actually exogenous. Therefore, the SD-ratio would reflect a more conservative and risk-averse approach of the researcher as, in this case, the marginal posterior density value needs to be of very large magnitude, implying a very strong signal within the data. For smaller sample sizes and more complex systems, this effect tends to become more rare.

On the other hand, if ν is chosen to be too small, the hypothesis of exogeneity is favoured disproportionally as the prior marginal density value at zero is smaller than zero if, e.g, $\nu =$**dim**(Σ). Altogether, the above problem is less pronounced for a prior in which the centring matrix is set to $Q_{k,c}$ instead of I_k, but it still exists.

Setting with $\rho_{12} = 0.4$ and $\rho_{13} = 0$							
		marginal posterior at zero			SD-ratio		
prior for Σ	ν	I-I	g-g	gprvar-g	I-I	g-g	gprvar-g
$\Sigma \sim IW(I_3, \nu)$	3	191.71	191.94	193.63	383.42	383.88	387.26
	6	194.18	194.49	196.19	57.45	57.54	58.04
	10	197.54	197.92	199.67	21.24	21.28	21.47
$\Sigma \sim IW(Q_{3,c}, \nu)$	3	191.59	191.90	193.59	504.18	505.00	509.45
	6	194.10	194.45	196.16	117.64	117.85	118.88
	10	197.48	197.88	199.64	77.44	77.60	78.29

Table 27: Density value of the marginal posterior of σ_{13} at zero and SD-ratio from a three-equation system for a sample size of $N = 500$.

True Correlation	Estimated Correlation	True Covariance	Estimated Covariance
$\begin{pmatrix} 1.00 & 0.4 & 0.0 \\ 0.4 & 1 & 0.3 \\ 0.0 & 0.3 & 1.00 \end{pmatrix}$	$\begin{pmatrix} 1.00 & 0.39 & -0.01 \\ 0.39 & 1 & 0.3 \\ -0.01 & 0.3 & 1.00 \end{pmatrix}$	$\begin{pmatrix} 25.00 & 4.00 & 0.00 \\ 4.00 & 4.00 & 1.80 \\ 0.00 & 1.80 & 9.00 \end{pmatrix}$	$\begin{pmatrix} 25.44 & 4.08 & -0.11 \\ 4.08 & 4.24 & 1.80 \\ -0.11 & 1.80 & 8.66 \end{pmatrix}$

Table 28: True and estimated posterior mean of the covariance matrices from first run of testing-down setup.

Eventually, we would recommend to set $\nu \geq \mathbf{dim}(\Sigma) + 1$ when $\Sigma \sim IW(I_k, \nu)$ and $\nu \geq \mathbf{dim}(\Sigma) + 2$ when $\Sigma \sim IW(Q_{k,c}, \nu)$. The actual threshold value for which the null can not be rejected has to be chosen by the researcher. The complexity of the data of the application will certainly play a role. In the end, the researcher must decide how much evidence she needs to favour the assumption of exogeneity.

Testing Down - Example

In the following example, we want to demonstrate the procedure as it is supposed to be implemented. We use the same set of equations as in the previous subsection. For the model parameters, we consider a gprvar-g prior structure and assume $\Sigma \sim IW(Q_{3,c}, 5)$.

Table 28 reports the true covariance setting of this example as well as the estimated covariance. In the first run, we still assume that y_2 and y_3 are endogenous variables. Therefore, we have a system of three equations overall. In Table 29, the estimation results of the posterior inclusion probabilities, posterior means as well as posterior quantiles are displayed. The estimates do not indicate anything unusual. They seem to converge to the true values and do not suggest that the system is wrongly specified. However, the covariance $\hat{\sigma}_{13} = -0.11$ is quite small. We therefore calculate the marginal posterior density at zero of $\hat{\sigma}_{13}$ and compute the SD-ratio in order to determine whether we can potentially reduce the system by one equation. The estimated density value is **17.85**. The resulting SD-ratio in this case is **14.17**.

	PIP	True Value	Mean	Lower	Upper
$\mathbf{y_2}$	1.00	1.00	0.98	0.85	1.10
$\mathbf{y_3}$	1.00	-1.00	-0.97	-1.20	-0.73
$\mathbf{x_1}$	0.84	-1.80	-1.59	-2.62	-0.64
$\mathbf{x_2}$	0.77	1.50	1.22	0.37	2.11
x_3	0.26	0.00	0.19	-0.13	0.81
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{10}	0.15	0.00	0.01	-0.24	0.26
$\mathbf{x_{11}}$	1.00	1.00	1.04	0.79	1.28
$\mathbf{x_{12}}$	1.00	-1.50	-1.52	-1.81	-1.23
x_{13}	0.18	0.00	0.00	-0.10	0.09
x_{14}	0.15	0.00	0.01	-0.06	0.08
x_{15}	0.17	0.00	-0.01	-0.10	0.08

Table 29: Estimated posterior inclusion probabilities (PIP), posterior mean (Mean) as well as 0.05 (Lower) and 0.95 (Upper) quantile for variables in structural equation (5.6) based on an average of 100 samples with $N = 100$. The centring matrix is set to $Q_{3,c}$ and $\nu = 5$. Red variables are truly part of the model and assumed to be endogenous while blue variables are truly part of the model and assumed to be exogenous.

True Correlation	Estimated Correlation	True Covariance	Estimated Covariance
$\begin{pmatrix} 1.00 & 0.4 \\ 0.4 & 1.00 \end{pmatrix}$	$\begin{pmatrix} 1.00 & 0.42 \\ 0.42 & 1.00 \end{pmatrix}$	$\begin{pmatrix} 25.00 & 4.00 \\ 4.00 & 4.00 \end{pmatrix}$	$\begin{pmatrix} 25.16 & 4.15 \\ 4.15 & 3.92 \end{pmatrix}$

Table 30: True and estimated covariance matrices from second run of testing-down setup.

We now need to decide whether there is sufficient support for the hypothesis of exogeneity of y_3 as compared to y_3 being endogenous. The hypothesis of exogeneity, implying that $\sigma_{13} = 0$, is about 14 times more likely than the hypothesis of endogeneity, meaning $\sigma_{13} \neq 0$.

If we assume that y_3 is exogenous based on the above result, we re-estimate the system by only assuming y_2 to be endogenous. The prior for Σ is now an $IW_2(Q_{2,c}, 4)$ distribution. As the system is reduced by one equation, the degrees of freedom have to be adjusted accordingly. Table 30 displays the corresponding covariance matrices for this run and Table 31 shows the estimation results for the other parameters of the model. In this particular example, the estimation results do not improve dramatically, but are very similar in both cases. If we calculate the SD-ratio for the marginal posterior of $\hat{\sigma}_{12}$, we find that the posterior density value at zero is **0.22** and, consequently, the SD-ratio value is **0.17**. We would therefore correctly conclude that y_2 is indeed endogenous and stop with the 'testing down' - procedure at this point.

	PIP	True Value	Mean	Lower	Upper
$\mathbf{y_2}$	1.00	1.00	0.97	0.84	1.10
$\mathbf{y_3}$	1.00	-1.00	-1.04	-1.23	-0.84
$\mathbf{x_1}$	0.84	-1.80	-1.57	-2.59	-0.62
$\mathbf{x_2}$	0.76	1.50	1.21	0.35	2.12
x_3	0.22	0.00	0.10	-0.16	0.66
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{10}	0.15	0.00	0.01	-0.23	0.25
$\mathbf{x_{11}}$	1.00	1.00	1.04	0.79	1.29
$\mathbf{x_{12}}$	1.00	-1.50	-1.52	-1.80	-1.23
x_{13}	0.15	0.00	0.00	-0.09	0.08
x_{14}	0.15	0.00	0.01	-0.06	0.09
x_{15}	0.17	0.00	-0.01	-0.09	0.08

Table 31: Estimated posterior inclusion probabilities (PIP), posterior mean (Mean) as well as 0.05 (Lower) and 0.95 (Upper) quantile for variables in structural equation (5.6) based on an average of 100 samples with $N = 100$. The centring matrix is set to $Q_{2,c}$ and $\nu = 4$. Red variables are truly part of the model and assumed to be endogenous while blue variables are truly part of the model and assumed to be exogenous.

5.4 Empirical Application - Drivers of Corruption

> "*Corruption is an insidious plague that has a wide range of corrosive effects on societies. It undermines democracy and the rule of law, leads to violations of human rights, distorts markets, erodes the quality of life, and allows organized crime, terrorism and other threats to human security to flourish.*[17]"
>
> *Kofi Annan*

Corruption in all its forms has been an important issue on the international politics agenda for several years. A general definition is given by Transparency International as "the abuse of entrusted power for private gain"[18]. A more precise definition, particularly of political corruption, which is in the center of the following analysis, is given by Mark Philip in Heywood (2015):

> "Corruption in politics occurs where a public official violates the rules and/or norms of office, to detriment of the interests of the public (or some subsection thereof) who is the designated beneficiary of that office, to benefit themselves and a third party who rewards or otherwise incentivises the public official to gain access to goods or services they would not otherwise obtain."

[17] Statement On The Adoption By The General Assembly Of The United Nations Convention Against Corruption, New York, 31 October 2003.

[18] https://www.transparency.org/what-is-corruption#costs-of-corruption

Some reasons as to why this topic is of such importance include the fact that the costs of corruption have been estimated to amount to five per cent of global GDP (see Heywood, 2015). Along those lines, Rose-Ackerman (1999) states that "high levels of corruption are associated with lower levels of investment and growth". That, in turn, affects the well-being of societies at large. Transparency International states that "corruption is a major obstacle to democracy and the rule of law" and "corrodes the social fabric of society" [19].

The United Nations (UN) adopted the *United Nations Convention On Corruption* on 9th December 2003, which is in effect since 14th December 2005. It has 140 signatories and 186 parties (as of 26th June 2018) [20]. Among other things, chapter 1 article 1a states that its purpose is "to promote and strengthen measures to prevent and combat corruption more efficiently and effectively". To achieve this goal, it is of great concern to understand the main drivers of corruption in order to be able to take effective measures.

Most research aiming to determine the drivers of corruption has been of empirical nature as there are only some theoretical foundations underlying this subject. Jetter and Parmeter (2018) provide a summary of these empirical studies regarding both the types of methods used with most being OLS regressions and the respective significant findings. However, none of the studies incorporate model uncertainty, which, particularly in the given context with many potential drivers, is highly problematic. Further, only some of them account for potential reversed causality problems, i.e. endogeneity.

Therefore, Jetter and Parmeter (2018) perform a comprehensive analysis of a set of 36 potential corruption determinants. Their sample consists of 123 countries ($N = 123$), where each observation is the averaged value of annual data from 2001 to 2010. Their data set is comprised of 11 different sources, which are described in detail in the appendix of their paper. Instruments for potentially endogenous variables are constructed from lagged values (averages from 1991 to 2000).

Corruption as the dependent variable is measured by the Corruption Perception Index (CPI), which is a measure created by Transparency International. It is based on surveys of experts and business people and ranges from zero to ten, where higher numbers represent less corruption[21].

[19] See https://www.transparency.org/what-is-corruption#costs-of-corruption

[20] A list of countries who signed and/or ratified this treaty can be found here: https://www.unodc.org/unodc/en/corruption/ratification-status.html

[21] Note that, more recently, Transparency International has started using a scale of zero to 100 for the CPI, but the overall conceptual framework is similar.

In their empirical application, Jetter and Parmeter distinguish institutional, economic, cultural and geographic factors. From these 36 potential drivers, 12 are assumed to be endogenous. The authors explain in great detail the categorisation of each variable and what they expect the direction of each effect should be. To make it easier for the reader, one can find a table in Appendix E which gives an overview of the variable names and what they represent based on the information in Table A.1. from Jetter and Parmeter (2018).

Using a chain of three million draws and disregarding the first 200.000 draws as burn-in, the authors first estimate the influence of all potential drivers of corruption within a standard linear regression framework using the BMS package. Then, assuming 12 of the 36 variables to be endogenous[22], they estimate the system using the ivbma package, hence using the original prior set-up by Karl and Lenkoski (2012)[23]. Both approaches lead to different posterior means as well as posterior inclusion probabilities as is to be expected.

Jetter and Parmeter find that mostly institutional characteristics matter, less so cultural characteristics. The *rule of law* variable, for example, has a PIP of one in both frameworks. In the standard BMA framework, *property rights* has a PIP of 0.9 and the *absence of political rights* has a PIP of 0.85. The *participation of women* with a PIP of 0.97 and the *urbanisation rate* with a PIP of 0.92 seem to be important, as well.

Within the IVBMA framework, *government effectiveness* becomes more important with a PIP of 0.88 and the PIP of *property rights* decreases drastically to 0.01. The *absence of political rights* variable now only has a PIP of 0.68. Here, the *participation of women* seems to be somewhat less important with a PIP of 0.80 and the same holds for *urbanisation rate* with a PIP of 0.85. However, *duration of primary education* seems to be more relevant in this framework with a PIP of 0.75.

We know from our simulation studies that, on the one hand, ignoring endogeneity leads to false inclusion probabilities and false posterior means, but on the other hand, specifying a system that is too large causes efficiency losses up to the point that the PIPs, for example, can be far off the true values. Hence, this is a good opportunity to try out the testing-down approach from above, starting with the full 13-equation system and

---

[22]The variables assumed to be endogenous are *property rigths*, *rule of law*, *absence of press freedom*, *trade freedom*, *government size*, *level of democracy*, *absence of political rights*, *government effectiveness*, *ln(GDP/cap)*, *trade openness*, *imports as share of GDP* and *FDI*.

[23]Jetter and Parmeter (2018) perform the same ivbma analysis for developing countries only, as well. The most important findings also hold for the analysis of this sub-sample of $N = 95$. Further, the authors state that, in pre-analyses, variables like *foreign aid flows*, *Gini coefficients* and *inflation rates* were included. However, none were of importance in any setting, which is why they were omitted for the main analysis.

reducing it step-by-step if there is sufficient evidence that the variable we are testing is actually exogenous. Further, we have seen that the choice of the prior, particularly for the parameters of the structural equation, has noticeable influence on the results. We therefore want to investigate how the results change depending on the prior choice. The latter will certainly be useful information for other applied researchers, making them aware that their choices will affect the study outcome to a certain degree.

For the above reasons, we revisit this particular problem and re-estimate the system using the various priors we investigated earlier and apply the testing-down procedure. Additionally, we want to check whether we can obtain a model which has a better out-of-sample fit than an approach based on simply estimating the entire system.

Empirical Results

First, we look at the results when setting the prior for Σ to $IW(I_{m+1}, m+2)$. Table 32 displays some of the posterior estimation results for the different model parameter prior structures. It contains the posterior inclusion probabilities PIP_f for the full-system-sized models and PIP_r for the down-tested-sized models. Further, the table reports the posterior means, again for the full systems, Mean_f, and for the respective down-tested systems, Mean_r. The setting I-I within the full system (first two columns of the results in Table 32) basically replicates the results of Jetter and Parmeter (2018) with very small differences being present due to setting $\nu = 14$ instead of $\nu = 16$ in the prior for Σ[24]. As before, the prior on the model space is uniform across all models.

We see that the most important drivers of corruption (highlighted in light grey) seem to be *rule of law*, *level of democracy*, *absence of political rights*, *seats held by women*, *duration of primary education* and *urbanization rate*. They have a PIP of at least 0.70 in most of the models and settings. Hence, they are the same main drivers that Jetter and Parmeter (2018) found, except for *government-effectiveness*, which, using other priors and estimating the model based on a smaller system, does not seem to be that important.

In order to reduce the full system to the necessary size only, we follow the testing-down procedure described in the previous subsection and use a (conservative) threshold value for the Bayes-factor of 50. The reduced model results displayed in Table 32 represent the smallest possible system of equations according to the procedure.

[24]The length of the Markov chain is also set to 3,000,000 with a burn-in of 200,000 for each estimation in the testing-down procedure. For the cross-validation study, we reduced the number of draws to 1,000,000 with a burn-in of 100,000.

With regards to the down-tested models, we notice that they are slightly different for each prior structure. In the original setting (I-I), we end up with six endogenous variables instead of twelve. Using the g-g prior structure leads us to only having four endogenous variables and the gprvar-g prior implies the biggest reduction in system size with only three endogenous variables at the end.

If we compare the posterior means of the full systems, we notice that there sometimes are tangible differences, e.g., the Mean_f of *rule of law* is 1.37 under the original prior structure, 1.69 under the g-g prior and 1.55 under the gprvar-g prior structure. In the simulation study in the previous subsection, we saw that different prior structures perform differently, particularly when the system is unnecessarily big, and that the prior choice for the model parameters matters. These simulation findings seem to be underpinned by the empirical results, as well.

We also observed that the posterior estimates changed somewhat when estimating a reduced version of the system. We therefore expect the Mean_r estimates to be different from those of the full-sized system. Further, we would also expect them to be more similar between the priors when comparing the tested-down versions.

The Mean_r estimates do indeed change in comparison to the Mean_f estimates of the respective full model with the Mean_r estimates also becoming somewhat more similar between the different priors. The variable *rule of law*, for instance, now has a posterior mean of 1.81 using the original prior structure, a posterior mean of 1.86 for the g-g prior and a posterior mean of 1.89 for the gprvar-g prior in the respective reduced system model.

Further, Table 33 shows the results of the leave -one -out cross validation study we additionally conducted in order to have a measure regarding the performance of the priors in this empirical setting. The ivbma-package already contains a function measuring the squared error (SE), the absolute error (AE), the predictive variance (VAR) and the continuous rank predictive score (CRPS) (see, e.g., Gneiting and Raftery, 2007). We further implemented the log-predictive score (LPS), which was introduced by Good (1952) as it is a quite common measure to evaluate predictive performance[25]. In order to make the implementation more feasible for future empirical applications using larger systems, we parallelized the cross-validation function across observations. We observe that, for the full system, the predictive accuracy measure results as depicted in Table 33 are actually the same regardless of the prior choice. However, the fit is slightly improved

[25]One application using the LPS can be found , e.g., in Fernández et al. (2001a).

	I-I				g-g				gprvar-g			
Variable	PIP_f	$Mean_f$	PIP_r	$Mean_r$	PIP_f	$Mean_f$	PIP_r	$Mean_r$	PIP_f	$Mean_f$	PIP_r	$Mean_r$
property rights	0.01	0	0.16	0	0.05	0	0	0	0.33	0	0.15	0
rule of law	1	1.37	1	1.81	1	1.69	1	1.86	1	1.55	1	1.89
absence of press freedom	0.04	0	0.29	0	0	0	0	0	0.15	0	0.17	0
trade freedom	0.49	0.01	0.44	0	0.39	0.01	0.02	0	0.9	0.02	0.44	0.01
government size	0.03	0	0.18	0	0	0	0	0	0.16	0	0.17	0
level of democracy	0.33	0.03	0.7	0.05	0.64	0.06	0.8	0.07	0.69	0.04	0.91	0.08
absence of political rights	0.79	0.2	0.87	0.28	0.88	0.3	0.88	0.31	0.98	0.3	0.96	0.35
government effectiveness	0.86	0.57	0.44	0.19	0.32	0.2	0.13	0.07	0.63	0.31	0.31	0.12
ln(GDP/cap)	0.33	0.06	0.36	0.06	0.06	0.01	0	0	0.22	0.02	0.33	0.05
trade openness	0.01	0	0.42	0	0	0	0.58	0.16	0.31	0	0.31	0
imports as share of GDP	0.01	0	0.31	0	0	0	0	0	0.23	0	0.22	0
FDI	0.04	0	0.52	-0.01	0	0	0	0	0.44	0.02	0.17	0
years democratic	0.04	0	0.41	0	0	0	0	0	0.32	0	0.28	0
federal	0.15	-0.01	0.11	-0.01	0.02	0	0.01	0	0.12	-0.01	0.1	-0.01
British	0.21	-0.03	0.16	-0.02	0.04	-0.01	0.01	0	0.15	-0.02	0.13	-0.02
Spanish	0.19	0.02	0.1	0.01	0.03	0	0.02	0	0.16	0.03	0.12	0.02
Portuguese	0.31	0.07	0.23	0.06	0.07	0.02	0.06	0.02	0.32	0.1	0.19	0.05
French	0.14	0	0.1	-0.01	0.01	0	0.01	0	0.1	0	0.11	-0.01
Dutch	0.53	-0.19	0.23	-0.07	0.18	-0.08	0.09	-0.03	0.38	-0.13	0.27	-0.09
Europe	0.23	-0.04	0.11	-0.01	0.02	0	0.02	0	0.16	-0.03	0.12	-0.01
Africa	0.21	-0.03	0.22	-0.05	0.02	0	0.02	0	0.14	-0.01	0.18	-0.03
Asia	0.43	-0.13	0.13	-0.02	0.09	-0.03	0.03	0	0.33	-0.08	0.19	-0.04
South America	0.22	0.04	0.14	0.02	0.03	0.01	0.04	0.01	0.14	0.02	0.13	0.02
North America	0.19	-0.01	0.09	0.01	0.02	0	0.02	0	0.1	0.01	0.09	0
seats held by women	0.85	0.02	1	0.02	0.9	0.02	0.97	0.03	1	0.02	1	0.02
duration secondary education	0.23	0.03	0.39	0.04	0.03	0	0.11	0.02	0.15	0	0.24	0.02
duration primary education	0.85	0.18	0.83	0.15	0.89	0.19	0.45	0.09	0.89	0.16	0.81	0.13
sec. education enrollment rate	0	0	0.11	0	0	0	0	0	0.12	0	0.11	0
common law	0.14	-0.01	0.12	-0.01	0.02	0	0.01	0	0.11	-0.01	0.1	-0.01
ln(population)	0.05	0	0.22	0.01	0	0	0.01	0	0.31	-0.02	0.15	0
ethnic fractionalisation	0.28	0.06	0.12	0.02	0.04	0.01	0.08	0.03	0.11	0.01	0.14	0.03
language fractionalisation	0.37	0.13	0.3	0.11	0.04	0.01	0.27	0.13	0.17	0.04	0.22	0.07
religious fractionalisation	0.6	0.24	0.46	0.21	0.42	0.22	0.11	0.04	0.66	0.3	0.48	0.24
natural resources rents	0	0	0.25	0	0	0	0	0	0.11	0	0.1	0
share of Protestants	0.45	0.21	0.21	0.11	0.2	0.15	0.13	0.08	0.19	0.09	0.25	0.15
urbanization rate	0.88	0.77	0.81	0.81	0.87	1	0.59	0.73	0.87	0.79	0.86	0.9

Table 32: Posterior estimation results for structural equation using different priors while setting the prior of Σ to $\Sigma \sim IW(I_{m+1}, m+2)$. $PIP_{f,r}$ represents the posterior inclusion probability for the full system and the down-tested system, respectively. $Mean_{f,r}$ represents the posterior mean for the full system and the down-tested system, respectively. Variables marked in bold blue are assumed to be endogenous. Markings in light grey emphasise the most important drivers of corruption.

for the down-tested systems. The model obtained in the testing down procedure using the gprvar-g prior gives the best results in this context.

We now want to examine how our results depend on the choice of the prior for Σ. We repeat the entire estimation procedure above while setting the prior for Σ to $IW(Q_{m+1,c}, m+3)$, where $Q_{m+1,c}$ is the centring matrix with the diagonal elements set to one and the off-diagonal elements set to 0.5. Table 34 displays the posterior estimation results (PIPs and Means) of the full-system-sized model and the down-tested version, again, for three different prior structures for the model parameters and using the adapted prior structure for Σ. The first interesting finding to notice is that, in this setting, the variables that seem to be most important are the same as above, although the results regarding the posterior inclusion probabilities are more mixed between the priors. Interestingly, changing the prior on Σ has a tangible effect on the overall results. For instance, the tested-down version of the system of equations for each prior is quite

	Full			Down-tested		
Measure	I-I	g-g	gprvar-g	I-I	g-g	gprvar-g
SE	0.29	0.29	0.29	0.30	0.29	0.29
AE	0.44	0.44	0.44	0.44	0.44	0.43
VAR	0.35	0.35	0.35	0.33	0.31	0.30
CRPS	0.35	0.35	0.35	0.34	0.34	0.33
LPS	0.83	0.83	0.83	0.83	0.81	0.80

Table 33: Cross-validation predictive accuracy for structural equation for different prior settings based on the full set of equations on the left (Full) and down-tested set of equations on the right (Down-tested) using the identity as centring matrix in prior for Σ.

different from the one above. Under the original prior I-I, we have two endogenous variables at the end of the testing-down procedure in contrast to six. Under the g-g prior, we only end up with one endogenous variable instead of four and, under the gprvar-g prior, we have a bigger system eventually with seven endogenous variables instead of three.

Another effect to notice is that, if we compare the tested-down versions between the two different settings for the prior of Σ for each prior structure, we observe that the remaining endogenous variables in the respective smaller system in either prior setting are not necessarily a subset of the endogenous variables in the larger reduced system. For instance, when using the original prior I-I while setting $\Sigma \sim IW(I_{m+1}, m+2)$, the testing down procedure results in classifying *property rights*, *absence of press-freedom*, *level of democracy*, *ln(GDP/cap)*, *trade-openness* and *FDI* as endogenous variables. Under the same prior, but while setting $\Sigma \sim IW(Q_{m+1,c}, m+3)$, the testing down procedure results in classifying only *property rights* and *absence of political rights* as endogenous variables. The latter, however, has been classified as exogenous in the first set-up. We therefore conclude that the prior choice for Σ has profound influence on the estimation results and needs to be considered carefully.

In comparison to the setting above, the PIPs are similar on average for the variables *rule of law*, *absence of political rights* and *seats held by women*. For *duration of primary education* and *urbanization rate*, they are slightly higher on average and for *level of democracy*, the PIPs are slightly lower in this setting. The posterior means are also somewhat different in comparison to the results in Table 32. The latter is not so much the case with regards to the full system, but certainly with regards to the tested-down models.

This finding may not be too surprising as the models for the reduced systems differ across the two settings. The estimated effects on corruption here are therefore somewhat smaller on average for the *rule of law*, the *level of democracy*, the *absence of political rights* and the *seats held by women*. The Mean_r estimates of the *duration of primary education* and the *urbanization rate* are somewhat larger on average.

Furthermore, the variation of the Mean_r estimate values in Table 34 between the different priors seems to be more pronounced than it is the case in Table 32. For instance, the Mean_r of the variable *rule of law* varies between 1.17, 2.06 and 1.42. Hence, the phenomenon we observed in the simulation study in the previous subsection, implying that the results across the different priors converge if the system is specified more accurately, does not appear to be applicable here. One possible reason could be that the adapted prior for Σ in this empirical example is less in line with the true underlying covariance structure than the original one and that, consequently, the testing-down procedure does not work as well.

Table 35 displays the cross-validation results for the full and the reduced system for different model parameter priors while using the adapted prior for Σ with $Q_{m+1,c}$ as centring matrix. For the full system, we do not see a difference when comparing the priors I-I, g-g and gprvar-g as well as the original prior structure for Σ. For the tested down systems, however, we observe that, under the gprvar-g prior, we find an improved predictive performance in contrast to the full-system-sized model. However, the improvements are slightly smaller than when using the original prior structure for Σ.

Discussion

From an application perspective, the analysis itself could potentially be improved if a different unit of assessment than the nation state was used since corruption can vary dramatically in different regions within one country. However, it could be difficult to actually collect data at this level. Additionally, there are transnational networks and relationships as well as interactions that should be incorporated in the analysis, too. It seems very likely that there are geographical proximity effects, implying that countries which are close to countries that are less corrupt are also less likely to be corrupt themselves and vice versa. And, lastly, it would be interesting to analyse the private sector in this context as well. The latter is relevant as there are many public services which are delivered by the private sector.

With regards to the methodology, as we obtain the same cross-validation-results for the full-sized system for all settings, we observe that the choice of the prior is not overly important when it comes to the predictive ability given a certain model specification.

	I-I				g-g				gprvar-g			
Variable	PIP_f	$Mean_f$	PIP_r	$Mean_r$	PIP_f	$Mean_f$	PIP_r	$Mean_r$	PIP_f	$Mean_f$	PIP_r	$Mean_r$
property rights	**0.01**	**0**	**0.01**	**0**	**0.04**	**0**	**0**	**0**	**0.38**	**0**	**0.72**	**0.01**
rule of law	**1**	**1.39**	1	1.17	**1**	**1.64**	1	2.06	**1**	**1.58**	**1**	**1.42**
absence of press freedom	**0.08**	**0**	0.01	0	**0.01**	**0**	0	0	**0.17**	**0**	**0.14**	**0**
trade freedom	**0.42**	**0.01**	0.02	0	**0.57**	**0.01**	0	0	**0.97**	**0.02**	0.47	0.01
government size	**0.02**	**0**	0.08	0	**0**	**0**	0	0	**0.19**	**0**	**0.24**	**0**
level of democracy	**0.11**	**0**	0.02	0	**0.33**	**0.03**	0.97	0.1	**0.63**	**0.03**	0.77	0.04
absence of political rights	**0.51**	**0.08**	**0.09**	**0.01**	**0.85**	**0.18**	0.98	0.4	**0.95**	**0.24**	0.98	0.28
government effectiveness	**0.7**	**0.39**	0.9	0.57	**0.29**	**0.16**	0.09	0.04	**0.39**	**0.15**	0.33	0.11
ln(GDP/cap)	**0.45**	**0.1**	0.33	0.07	**0.12**	**0.03**	0.01	0	**0.16**	**0**	**0.41**	**0.07**
trade openness	**0.01**	**0**	0	0	**0**	**0**	0	0	**0.29**	**0**	0.52	0.02
imports as share of GDP	**0.01**	**0**	0.01	0	**0**	**0**	0	0	**0.35**	**0**	**0.5**	**-0.04**
FDI	**0.05**	**0**	0.01	0	**0.01**	**0**	0	0	**0.5**	**-0.01**	**0.55**	**-0.01**
years democratic	0.03	0	0.01	0	0	0	0	0	0.22	0	0.36	0
federal	0.13	-0.01	0.13	0	0.01	0	0.02	0	0.09	0	0.11	-0.01
British	0.19	-0.03	0.26	-0.05	0.03	0	0.04	-0.01	0.13	-0.01	0.14	-0.02
Spanish	0.17	0	0.17	-0.01	0.02	0	0.02	0	0.15	0.02	0.14	0.02
Portuguese	0.25	0.05	0.19	0.02	0.04	0.01	0.03	0.01	0.34	0.1	0.18	0.04
French	0.14	0	0.16	-0.02	0.01	0	0.01	0	0.09	0	0.11	-0.01
Dutch	0.63	-0.25	0.46	-0.16	0.15	-0.06	0.11	-0.04	0.4	-0.14	0.25	-0.07
Europe	0.2	-0.03	0.15	0	0.02	0	0.02	0	0.15	-0.03	0.16	-0.03
Africa	0.19	-0.03	0.2	-0.03	0.02	0	0.02	0	0.13	-0.01	0.19	-0.05
Asia	0.33	-0.09	0.18	-0.03	0.06	-0.01	0.06	-0.01	0.25	-0.05	0.22	-0.05
South America	0.25	0.05	0.2	0.03	0.03	0.01	0.03	0	0.15	0.03	0.18	0.04
North America	0.22	-0.03	0.21	-0.02	0.02	0	0.02	0	0.09	0	0.1	0
seats held by women	0.92	0.02	0.64	0.01	0.94	0.02	0.94	0.02	1	0.02	0.99	0.02
duration secondary education	0.35	0.05	0.83	0.15	0.03	0	0.03	0.01	0.12	0	0.24	0.02
duration primary education	0.93	0.21	0.99	0.3	0.97	0.23	0.98	0.18	0.94	0.17	0.79	0.14
sec. education enrollment rate	0	0	0	0	0	0	0	0	0.18	0	0.1	0
common law	0.13	0	0.17	-0.01	0.02	0	0.02	0	0.1	0	0.12	-0.01
ln(population)	0.04	0	0.04	0	0	0	0	0	0.2	-0.01	0.23	-0.01
ethnic fractionalisation	0.24	0.04	0.23	0.03	0.03	0.01	0.05	0.02	0.12	0.02	0.13	0.02
language fractionalisation	0.46	0.18	0.33	0.1	0.09	0.04	0.04	0.01	0.24	0.07	0.37	0.16
religious fractionalisation	0.62	0.24	0.48	0.17	0.45	0.22	0.73	0.46	0.78	0.38	0.37	0.14
natural resources rents	0	0	0	0	0	0	0	0	0.12	0	0.4	-0.01
share of Protestants	0.49	0.24	0.58	0.36	0.18	0.12	0.19	0.15	0.17	0.07	0.23	0.13
urbanization rate	0.86	0.76	0.94	1.05	0.91	1.05	0.99	1.36	0.86	0.74	0.82	0.85

Table 34: Posterior estimation results for structural equation using different priors while setting the prior of Σ to $\Sigma \sim IW(Q_{m+1,c}, m+3)$. $PIP_{f,r}$ represents the posterior inclusion probability for the full system and the down-tested system, respectively. $Mean_{f,r}$ represents the posterior mean for the full system and the down-tested system, respectively. Variables marked in bold blue are assumed to be endogenous. Markings in light grey emphasise the most important drivers of corruption.

	Full			Down-tested		
Measure	I-I	g-g	gprvar-g	I-I	g-g	gprvar-g
SE	0.29	0.30	0.29	0.33	0.29	0.29
AE	0.44	0.44	0.44	0.45	0.43	0.43
VAR	0.35	0.35	0.35	0.35	0.31	0.31
CRPS	0.35	0.35	0.35	0.35	0.34	0.34
LPS	0.83	0.83	0.83	0.88	0.82	0.81

Table 35: Cross-validation predictive accuracy for structural equation for different prior settings based on the full set of equations on the left (Full) and the down-tested set of equations on the right (Down-tested) using $Q_{m+1,c}$ as centring matrix in prior for Σ.

However, the choice of the prior does have an influence on the testing down procedure. That is, both the choice of the prior on the model parameters and the choice of the prior for Σ.

It seems to be the case that the adapted prior structure for Σ we used here leads to slightly worse model selection when applying the testing-down procedure. Possibly, the assumed direction of the endogeneity has more influence on the results for the empirical sample than in the smaller simulation example. If that was the case, it would be advantageous to be more careful in specifying the centring matrix in the prior for Σ in empirical applications. If the researcher could put more insight into the design of the centring matrix with regards to the direction of the endogeneity based on the causal relations of the underlying research subject, it would be worthwhile to use an adapted prior for Σ. This is, however, an issue that needs to be investigated further.

Using the testing-down procedure in order to reduce the system to an appropriate size is recommendable anyway since we can achieve a better predictive performance and it will reduce the computational effort significantly[26]. This will be particularly useful when, e.g., the system needs to be re-estimated with an updated sample.

Overall, based on the results of this empirical study under both prior settings for Σ, it is recommendable to use the gprvar-g prior structure as the latter performed best.

[26]The time needed to compute the cv-results for the whole system on a cluster using 41 cores amounts to roughly 12 hours, whereas the time needed to compute the cv-results for a tested-down system with, say, only two endogenous variables, only amounts to roughly three hours on the same cluster.

6 An Alternative Modelling Approach

Until now we only examined methodology that already incorporates model uncertainty in an IV-model framework and, with the Karl and Lenkoski (2012) approach, we found a sound solution to build on. It seems to work well in empirical applications and is easy to understand due to its relatively simple structure and assumptions. For a more complete overview, however, it could be interesting to also study more analytical approaches which potentially give a bit more detailed insight into specific behaviour of the posterior distribution.

In order to do that, we first take one step back and investigate Bayesian frameworks for the IV-model that could be adapted to account for model uncertainty.

6.1 Bayesian Methods for IV-models

Many of the approaches examined in the following assume that no exogenous regressors are part of the model. In this context, the **SF** in (4.1) reduces to:

$$\begin{aligned} y_1 &= Y_2\beta + \epsilon_1, \\ Y_2 &= Z\Delta + V_2. \end{aligned}$$

The **RRF** from (4.2) then collapses to

$$\begin{aligned} y_1 &= Z\Delta\beta + v_1, \\ Y_2 &= Z\Delta + V_2, \end{aligned}$$

with $v_1 = \epsilon_1 + V_2\beta$ and the **URF** from (4.4) then equates to

$$\begin{aligned} y_1 &= Z\tau + v_1, \\ Y_2 &= Z\Delta + V_2. \end{aligned}$$

Distributional assumptions about the error terms are the same as in subsection 4.1.

The **Drèze prior** was originally proposed by Drèze (1976) for the analysis of a single equation from a simultaneous equation system in a limited information setting[27]. A diffuse prior is used for the parameters of the equation of interest and no prior information other than the list of exogenous and endogenous variables is introduced for the parameters

[27] The IV-model is a special case of a simultaneous equation system.

of the remaining equations in the system. The latter are integrated out analytically. The kernel of the posterior density for the regression coefficients is a ratio of t-kernels, which the author describes as the class of poly-t densities. Drèze (1977) provides a comprehensive analysis of the properties of poly-t densities, e.g., showing that these densities are typically asymmetric and multimodal. Drèze (1976) derives the joint posterior probability density for a vector of coefficients of the structural equation.

Tsurumi (1985) extends the analysis and derives analytical expressions for the marginal posteriors of individual coefficients of the structural equation using the cumulant generating function for a multivariate normal distribution. Yet, Maddala (1976) states that we need to be careful using non-informative priors since these can lead to sharp posterior distributions even in underidentified models. He shows through an example that, even though the likelihood is flat because one of the equations in the model is not identified, the Drèze approach still leads to posterior distributions that are sharp. Thus, he concludes that, in this setting, the Bayesian analysis with diffuse priors may be meaningless when the true parameter values violate the rank conditions for identification.

Kleibergen and Zivot (2003) analyse the Drèze prior in the context of the IV model. The flat diffuse prior for the parameters of the **SF** above is given by:

$$p_{SF}(\beta, \Delta, \Sigma) \propto |\Sigma|^{-\frac{1}{2}(k+m+2)}. \tag{6.1}$$

The implied prior on the **RRF** parameters is then

$$p_{RRF}(\beta, \Delta, \Omega) \propto |\Omega|^{-\frac{1}{2}(k+m+2)}. \tag{6.2}$$

One of the advantages of the Drèze approach is the invariance property between the flat prior on the SF coefficients and the RRF coefficients. The resulting marginal posterior for β is a 1-1 poly-t density and is proportional to the concentrated likelihood function of β used for limited information maximum likelihood (LIML). The resulting marginal posterior for Δ is a ratio of poly-t densities and is proportional to the kernel of a Student-t density centred at the ordinary least squares (OLS) regression of y_1 on Y_2. The marginal posterior of β has its mode close to $\hat{\beta}_{LIML}$ if the model is slightly overidentified, but for highly overidentified models, the marginal posterior mode can be very different from the latter. However, in case of an increasing number of good instruments, the posterior behaves more and more like the sampling density of the 2SLS estimate. Kleibergen and

Zivot also point out some undesirable features of this prior. The marginal posteriors are not invariant with respect to the ordering of y_1 and Y_2 in overidentified models. If the instruments are weak, the posterior of β is bimodal, although this diminishes rapidly as the number of instruments increases. The marginal posterior of Δ has a non-integrable asymptote at $\Delta = 0$, which is problematic since the posterior favours values of Δ near 0 regardless of the information in the data. Further, Kleibergen and Zivot stress the fact that the marginal posterior of β is sensitive to superfluous instruments. As Maddala (1976) points out, in the just-identified case, the posterior is not proper, but can be made so by simply adding superfluous variables to Z.

The **Bayesian two-stage (B2S)** prior structure is an improved version of the Drèze prior and is also proposed by Kleibergen and Zivot (2003). It is named that way since the resulting posterior has properties similar to the frequentist 2SLS estimator. Kleibergen and Zivot argue that one of the main 'flaws' of the Drèze prior is the assumption of independence between β and Δ, although it is a priori known that the model is uninformative about β when $\Delta = 0$. The authors incorporate this knowledge in the prior structure in a 2SLS fashion with β being estimated conditionally on an estimate of Δ by adding a term where β is not identified if Δ has reduced rank. The authors reparametrize the RRF slightly such that:

$$
\begin{aligned}
y_1 &= Z\Delta\beta + e_1 + V_2\phi, \\
y_2 &= Z\Delta + V_2,
\end{aligned}
$$

where, due to the normality assumption of the error process of the underlying structural equations $\phi = \Omega_{22}^{-1}\Omega_{21}$, $\mathrm{Var}(e_1) = \omega_{11.2} - \omega_{12}\Omega_{22}^{-1}\omega_{21}$ and $e_1 \perp V_2$. They continue to suggest the following prior for the parameters of the above system:

$$
p_{RRF}^{B2S}(\beta, \phi, \omega_{11.2}, \Omega_{22}) \propto |\omega_{11.2}|^{-(m+1)}|\Omega|^{-\frac{1}{2}(m+k)}|\Delta' Z' Z\Delta|^{\frac{1}{2}}, \tag{6.3}
$$

where the last term is the main difference compared to the Drèze prior. The conditional and marginal posteriors of the parameters derived by the authors show the similarities with the frequentist 2SLS estimation procedure. All in all, the posteriors behave similarly as in the Drèze approach, but somewhat more like the classical 2SLS. The mean of the conditional posterior of β evaluated at $\hat{\phi}$ and $\hat{\Delta}$ is equal to $\hat{\beta}_{2SLS}$. In contrast to the Drèze prior, the asymptote of Δ at $\Delta = 0$ does not exist due to $|\Delta' Z' Z\Delta|^{\frac{1}{2}}$. However, these priors are still quite similar. Therefore some of the drawbacks of the Drèze prior still remain. The posteriors in the B2S approach are not invariant to the ordering of the endogenous variables either and the marginal posterior of β is still sensitive to the addition of superfluous instruments.

Lopes and Polson (2014) focus their analysis on a two dimensional system and propose a **Cholesky-based prior** for the covariance matrix Σ in the SF. The structural equation is used since β and Σ are intertwined in the reduced form, such that independent priors for both parameters would be counter-intuitive. The authors state that the Cholesky-based prior is a much more flexible alternative to the inverted Wishart prior and is also more realistic than the uniformative Jeffreys prior. The inverted Wishart prior has only one tightness parameter and thus can not be informative for some elements and less informative for others. For that reason, Lopes and Polson model the components of the recursive conditional regression that arises from the Cholesky decomposition of Σ.

Let $\Sigma = AHA'$, where A is a 2×2 upper triangular matrix with ones on the main diagonal and upper triangular component $a_{12} = \frac{\sigma_{12}}{\sigma_{22}}$, while $H = \text{diag}(\sigma_{1|2}, \sigma_{22})$. The reverse transformation is given by $\sigma_{12} = a_{21}\sigma_{22}$ and $\sigma_{11} = \sigma_{1|2} + \frac{\sigma_{12}^2}{\sigma_{22}}$. It follows that $A^{-1}V_j \sim N(0, H)$ with $V_j = (\epsilon'_{1j}\, V'_{2j})'$. The error process can be rewritten as $V_{2j} \sim N(0, \sigma_{22})$ and $\epsilon_{1j}|V_{2j} \sim N(a_{12}V_{2j}, \sigma_{1|2})$ where a_{12} measures the strength of the correlation between ϵ_{1j} and V_{2j}, while $\sigma_{1|2}$ is the conditional residual variance. Thus, one can specify independent prior distributions for σ_{22}, a_{12} and $\sigma_{1|2}$. The advantage of this prior is the relative freedom to independently quantify the uncertainty for the individual components of Σ. The authors suggest to assign Inverted-Gamma priors to σ_{22} and $\sigma_{1|2}$ as well as a(n) (inverse) Gaussian prior for a_{12}. A normal prior is put on all the remaining parameters of the system.

Conley et al. (2008) suggest a **Bayesian semi-parametric approach** for the estimation of the IV model. In contrast to most of the methods found in the literature , the authors here want to build a more flexible framework that does not rely on the normality assumption of the error term and therefore the likelihood. For that reason, Conley et al. (2008) assume linear structural and reduced-form equations but model the error distribution in a non-parametric way. Overall, this estimation procedure is very flexible and is able to handle very general forms of heterogeneity in the error distribution. Further, if the errors are non-normal, this procedure is more efficient than standard Bayesian or classical methods. However, it is also not as intuitive and requires many parameters to be estimated and many hyper-parameters to be chosen.

One interesting approach, which can be extended, is the natural conjugate prior by Hoogerheide et al. (2007). The authors call it natural conjugate since the marginal prior as well as the marginal posterior of the structural parameter β, which, in most applications, will be of most interest, have the same functional form[28]. Hence, we can obtain some understanding of how the data updates our prior information about β. Furthermore, this prior treats all parameters of the system at once and does not lead to a step-wise procedure. In addition, for a typical setting of the hyperparameters, this prior structure is proper on the model specific parameters and hence, in principle, suitable for a model uncertainty framework. Another advantage is that, at least for the case where $m = 1$, analytical expressions for, e.g, the marginal prior and posterior of β and Ω are available. Further, there already exists an algorithm to calculate the marginal prior and posterior of the structural parameter β. One disadvantage might be that many hyper-parameters need to be elicited in advance, but, if we understand their influence, this fact can also provide a tool for making meaningful prior assumptions, leading to robust results. Hence, in the next sub-section, we will have a closer look at the natural conjugate approach by Hoogerheide et al. (2007).

6.2 The Hoogerheide et al. (2007) Approach and some Extensions

Hoogerheide et al. (2007) suggest a very elegant way to put a prior directly on the parameters of the RRF. We write down the system of equations, again, to make it easier for the reader to keep track of the notation. Distributional assumptions for the error terms are the same as in the previous parts of the thesis.

The analysis concentrates on the case where $m = 1$. Since y_2 in that case is a vector, we will use δ instead of Δ and v_2 instead of V_2. Hence, the **RRF** is parametrised as

$$y_1 = Z\delta\beta + v_1,$$
$$y_2 = Z\delta + v_2,$$

and the **URF** is given by

$$y_1 = Z\tau + v_1,$$
$$y_2 = Z\delta + v_2,$$

[28]This holds for the marginal prior and posterior of Ω as well.

where y_1, y_2, v_1 and v_2 are vectors of dimension $N \times 1$, Z is a matrix of dimension $N \times k_Z$, δ and τ are $k_Z \times 1$ vectors and β,i.e. the structural parameter, is a scalar. The corresponding covariance matrix Ω of $(v_{1,j}, v_{2,j})$ for $j = 1, ..., N$, is of dimension 2×2. Using the above notation, the likelihood can be expressed as

$$\begin{aligned} p(y_1, y_2|\delta, \beta, \Omega) \propto |\Omega|^{-\frac{N}{2}} \exp\Big\{ -\tfrac{1}{2}\mathrm{tr}[\Omega^{-1}((y_1 : y_2) \\ - Z(\delta\beta : \delta))'((y_1 : y_2) - Z(\delta\beta : \delta))]\Big\}. \end{aligned} \tag{6.4}$$

Derivation of the Natural Conjugate Prior

Below we describe the scheme Hoogerheide et al. (2007) use to obtain a prior on the parameters of the RRF by specifying a prior on the parameters of the URF. Their motivation is to use the property that, in the IV model, the RRF results from a reduced rank restriction on the parameter matrix of the URF. Hence, they specify a natural conjugate prior on the parameters of the URF and then impose rank reduction on its parameter matrix to obtain a prior on the parameters of the RRF. This procedure basically consists of four steps:

1. Hoogerheide et al. (2007) start with a prior specification on the parameters of the URF with $p^{nc}_{URF}(\tau, \delta|\Omega)$ being a matrix normal prior on (τ, δ) given Ω and $p^{nc}_{URF}(\Omega)$ being an inverted-Wishart prior on Ω, such that

$$\begin{aligned} p_{URF}(\tau, \delta|\Omega) &\propto |A_0|^{\frac{1}{2}} |\Omega|^{-\frac{k}{2}} \\ &\quad \times \exp[-\tfrac{1}{2}\mathrm{tr}[\Omega^{-1}((\tau : \delta) - \delta_0(\beta_0 : 1))' A_0((\tau : \delta) - \delta_0(\beta_0 : 1))]], \\ p_{URF}(\Omega) &\propto |\Omega|^{-\frac{1}{2}(\mu_0+3)} \exp[-\tfrac{1}{2}\mathrm{tr}(\Omega^{-1}\Omega_0)], \end{aligned}$$

   where A_0 is a $k \times k$ scale matrix. β_0 is a scalar location hyper-parameter that reflects prior information on the structural parameter β, δ_0 is a $k \times 1$ vector with prior information on the reduced form parameter vector δ, Ω_0 is a 2×2 scale matrix reflecting the prior information on the covariance matrix Ω and μ_0 are the prior degrees of freedom in the distribution of Ω. By choosing A_0 appropriately, this prior can be seen as a multivariate version of the g-prior in the linear regression case, as is explained in the next subsection.

2. The parameter matrix $(\tau : \delta)$ is normalized, such that

$$\Theta = Q(\tau : \delta)W, \tag{6.5}$$

where Q is a $k \times k$ and W a 2×2 scale matrix. These matrices can be chosen, e.g., such that Θ corresponds to the matrix of t-values of (τ, δ). This step is performed to obtain a certain invariance property to avoid statistical paradoxes[29]. Here, for the natural conjugate prior, they specify $Q = (A_0 + Z'Z)^{\frac{1}{2}}$ and $W = \Omega^{-\frac{1}{2}}$. The prior of step 1 can be formulated as:

$$p_{URF}(\tau, \delta, \Omega) = p_{URF}(\tau, \delta|\Omega)p_{URF}(\Omega),$$

such that

$$p_{URF}(\Theta, \Omega) = p_{URF}(\Theta|\Omega)p_{URF}(\Omega),$$

with $p_{URF}(\Theta|\Omega) = p_{URF}(\Theta(\tau, \delta)|\Omega)|J((\tau, \delta), \Theta)|$ where $J(a, b)$ is the Jacobian of the transformation from $a \rightarrow b$.

3. Referring to work of Kleibergen (1997) and Kleibergen and van Dijk (1998), Hoogerheide et al. use the fact that there exists a one-to-one mapping between Θ and β, δ, ζ in such a way that the singular value decomposition (SVD) of Θ can be as well expressed in terms of β, δ and ζ with ζ being a $k-1$ dimensional vector equal to the smallest singular value pre-multiplied by a matrix[30]:

$$\Theta = Q[\delta(\beta : 1) + \delta_{\perp}\zeta(\beta : 1)_{\perp}]W, \tag{6.7}$$

---

[29]The authors refer here to the Borel-Kolmogorov paradox.

[30]The SVD of Θ can be written as

$$\begin{aligned} \Theta = USV' &= \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & U_{22} \end{pmatrix} \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix} \begin{pmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \end{pmatrix}, \\ &= \begin{pmatrix} u_{11} \\ u_{21} \end{pmatrix} s_1 \begin{pmatrix} v_{11} & v_{21} \end{pmatrix} + \begin{pmatrix} u_{12} \\ U_{22} \end{pmatrix} s_2 \begin{pmatrix} v_{12} & v_{22} \end{pmatrix}, \end{aligned} \tag{6.6}$$

where for $m = 1$, U is an orthogonal $k \times k$ matrix, V is an orthonormal 2×2 matrix, while the $k \times 2$ matrix S contains the non-negative singular values in decreasing order on the main diagonal, such that $S = \begin{pmatrix} s_1 & \vdots & 0 \\ 0 & \vdots & s_2 \end{pmatrix}$ with s_1 being a scalar and s_2 being a $(k-1) \times 1$ vector. Hence, the first element of s_2 is the smallest singular value and the other elements are zero. Further, u_{11}, v_{11}, v_{12}, v_{21} and v_{22} are scalars, u_{21} is a $(k-1) \times 1$ vector, u_{12} is a $1 \times (k-1)$ vector, while U_{22} is a $(k-1) \times (k-1)$ matrix. The decomposition (6.6) is subsequently expressed in terms of the right-hand side variables in (6.7). E.g., $\zeta = (U_{22}U_{22}')^{-\frac{1}{2}} U_{22}s_2$. See Hoogerheide et al. (2007) for details.

where $(\beta:1)_\perp$ is of dimension 1×2 and $\delta_\perp$ is of dimension $k\times(k-1)$, such that:

$$(\beta:1)_\perp WW'(\beta:1)' \equiv 0, \text{ and } (\beta:1)_\perp WW'(\beta:1)'_\perp \equiv 1,$$
$$\delta'_\perp QQ'\delta \equiv 0, \text{ and } \delta'_\perp QQ'\delta_\perp \equiv I_{k-1}.$$

Then, $p_{URF}(\beta,\delta,\zeta,\Omega) = p_{URF}(\Theta(\delta,\beta,\zeta)|\Omega)|J(\Theta,(\delta,\beta,\zeta))|p_{URF}(\Omega)$.

4. The prior of (β,δ,Ω) in the RRF then equals the conditional prior of (β,δ,Ω) given $\zeta=0$ of the URF. Here, the main underlying idea is that the URF constitutes a multivariate linear regression model, while the RRF results from imposing a non-linear parameter restriction. The latter is equivalent to setting the smallest singular value of the parameter matrix Θ to zero. Hence, the prior on the parameters of the RRF is proportional to the prior on the parameters of the URF given that the above restriction holds (see Kleibergen and van Dijk, 1998).

Performing all these steps leads to the resulting natural conjugate prior on the RRF:

$$\begin{aligned} p^{nc}_{RRF}(\beta,\delta,\Omega) \propto\ & |\delta'(A_0+Z'Z)\delta|^{\frac{1}{2}}|(\beta:1)\Omega^{-1}(\beta:1)'|^{\frac{1}{2}(k-1)}|A_0|^{\frac{1}{2}} \\ & \times|\Omega|^{-\frac{1}{2}(\mu_0+4)}\exp\Big[-\tfrac{1}{2}\mathrm{tr}(\Omega^{-1}\Omega_0) \\ & -\tfrac{1}{2}\mathrm{tr}\{\Omega^{-1}(\delta(\beta:1)-\delta_0(\beta_0:1))'A_0(\delta(\beta:1)-\delta_0(\beta_0:1))\}\Big]. \end{aligned} \tag{6.8}$$

For the case of the Jeffrey's prior, which is a special case of the natural conjugate prior for $A_0=0$, $\Omega_0=0$ and $\mu_0=0$, the authors prove that the above procedure is valid. They derive the Jeffrey's prior on the RRF as it is defined, more precisely as the square root of the determinant of the information matrix. Then, Hoogerheide et al. (2007) put a Jeffrey's prior on the parameters of the URF and apply the above procedure. The obtained prior on the parameters of the RRF is the same as the directly derived Jeffrey's prior. This shows that the rank-reducing approach is indeed a sensible one. Most of the time, however, the direct derivation of prior structures on the RRF parameters is a challenging task, whereas developing a prior on the URF is more straightforward. The latter fact makes the above approach particularly useful.

As to the Jeffrey's prior, the latter is not suitable if we want to incorporate model uncertainty since it is improper on the model-specific parameters. For that reason, we need to choose the hyper-parameters differently. As mentioned above, another special case of the natural conjugate prior is the so called G-prior (see Hoogerheide et al., 2007).

For $A_0 = gZ'Z$ with g being a positive scalar, the resulting G-prior will be a proper prior on model specific parameters.

The G-Prior

In this section, we want to have a closer look at the behaviour of the prior structure and the resulting posterior when imposing the the G-prior. Hoogerheide et al. (2007) use it in their application, but do not discuss in detail, how the choice of the individual hyper-parameters affects the results. Instead, they use stylized facts from the application context to assign values to the hyper-parameters. This is not always possible in all applications. Hence, it will be important to obtain a better understanding of how each parameter influences prior and posterior behaviour.

Simulation Study of Prior Behaviour

The prior can be expressed as

$$\begin{aligned} p^g_{RRF}(\beta,\delta,\Omega) \propto\; & |\delta'Z'Z\delta|^{\frac{1}{2}}|(\beta:1)\Omega^{-1}(\beta:1)'|^{\frac{1}{2}(k-1)}|g(1+g)|^{\frac{1}{2}} \\ & \times |\Omega|^{-\frac{1}{2}(\mu_0+4)} \times \exp\Big[-\tfrac{1}{2}\mathrm{tr}(\Omega^{-1}\Omega_0) \\ & -\tfrac{1}{2}\mathrm{tr}\{g\,\Omega^{-1}(\delta(\beta:1)-\delta_0(\beta_0:1))'Z'Z(\delta(\beta:1)-\delta_0(\beta_0:1))\}\Big], \end{aligned} \tag{6.9}$$

where g is the weight that is attached to the prior as compared to the likelihood in the posterior. All other parameters are defined as above. Hoogerheide et al. (2007) derive the marginal prior for β,Ω as

$$p(\beta,\Omega)^g_{RRF} = p(\beta|\Omega)_{RRF}\,q(\Omega)_{RRF}, \tag{6.10}$$

with

$$\begin{aligned} p(\beta|\Omega)^g_{RRF} \propto\; & |(\beta:1)\Omega^{-1}(\beta:1)'|^{-1} \\ & \times \sum_{h=0}^{\infty}\left(\frac{1}{2}\frac{g((\beta:1)\Omega^{-1}(\beta_0:1)')^2\delta_0'Z'Z\delta_0}{(\beta:1)\Omega^{-1}(\beta:1)'}\right)^h 2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2h+1))}{\Gamma(\frac{1}{2}(k+2h))h!},\; k\geq 1, \\ q(\Omega)^g_{RRF} \propto\; & |\Omega|^{-\frac{1}{2}(\mu_0+4)}\exp[-\tfrac{1}{2}\mathrm{tr}\{\Omega^{-1}(\Omega_0+g(\beta_0:1)'\delta_0'Z'Z\delta_0(\beta_0:1))\}]. \end{aligned}$$

We can see that, by integrating out δ, the term $|(\beta:1)\Omega^{-1}(\beta:1)'|^{-1}$, which influences the tail behaviour of the marginal prior of β, no longer depends on the number of instruments. One should note that the marginal prior of $\beta|\Omega$ has Cauchy-tails, which leads to the fact that no finite moments larger or equal to one exist. Usually, we will be mostly interested in the structural parameter β. Hence, it is useful to focus on studying the marginal prior of β and how it is influenced by the choice of hyper-parameters. Hoogerheide et al. (2007) suggested an algorithm to simulate the latter:

1. Specify a grid of values of β, with $\beta^1, ..., \beta^L$.
2. Generate T values of Ω^t, $t = 1, ..., T$, from an inverted-Wishart distribution with scale matrix $\Omega_0 + g(\beta_0 : 1)'\delta_0' Z' Z \delta_0(\beta_0 : 1)$ and $\mu_0 + 2$ degrees of freedom.
3. Compute: $p^g_{RRF}(\beta^l|\Omega) = \frac{1}{T}\sum_{t=1}^T p^g_{RRF}(\beta^l|\Omega^t)$ for $l = 1, ..., L$.
4. Compute the marginal prior of β : $p^g_{RRF}(\beta^l) = \frac{1}{c} p^g_{RRF}(\beta^l|\Omega)$, $l = 1, ..., L$, with $c = \sum_{i=1}^L (\beta^{i+1} - \beta^i) p^g_{RRF}(\beta^i|\Omega)$.

We now can use the above algorithm in a simulation example to study the behaviour of the marginal prior of β more closely. The structural equation of the true model looks as follows:

$$
\begin{aligned}
y_1 &= y_2 + \epsilon_1, \\
y_2 &= 0.75 z_1 - 2 z_8 + v_2,
\end{aligned}
$$

where y_1, y_2, z_1, z_8, v_2 and ϵ_1 are all $N \times 1$ random vectors. We start by drawing v_2 as a standard normal random vector. Then, ϵ_1 is created as $\epsilon_1 = a \times v_2 + \xi$, where a is a scalar determining the strength of the endogeneity and ξ is a standard normal random vector of dimension $N \times 1$. Afterwards, we construct a matrix of instruments $Z = (Z^{(1)} : Z^{(2)})$, which is of dimension $N \times 10$. $Z^{(1)}$ is a standard normal random matrix of dimension $N \times 5$. $Z^{(2)} = Z^{(1)}K + F$, where $K = (0.3, 0.5, 0.7, 0.9, 1.1)'(1, 1, 1, 1, 1)$ is a matrix inducing correlation and F is a standard normal random matrix of the same dimension as $Z^{(1)}$. The index i in z_i indicates which instrument and, hence, which column of Z has been used as instrument in the true model[31]. In this model the true value of the structural parameter is $\beta = 1$. Accordingly, the RRF of the true model is the following:

$$
\begin{aligned}
y_1 &= 0.75 z_1 - 2 z_8 + v_1, \\
y_2 &= 0.75 z_1 - 2 z_8 + v_2,
\end{aligned}
$$

[31] For our example, we only need z_1 and z_8. But as we want to use this example later on to evaluate how the prior structure can be used in the case of model uncertainty, we already introduce a slightly bigger setup here

with $v_1 = \epsilon_1 + v_2$.

For the following simulation, we set $N = 100$ (at first), $T = 500$ and we repeat steps 1 to 4 of the above algorithm $R = 10$ times, such that we obtain ten different samples, and average over the results. As $p^g_{RRF}(\beta^l|\Omega), l = 1, ..., L$, contains an infinite sum, we need to determine a fixed value for h and cut the sum at that point (see equation (6.10)). Choosing $h = 25$ seems an appropriate choice since simulating with, e.g., $h = 40$ made no difference with regards to the results. The simulation was repeated with three different settings for β_0, namely $\beta_0 = 0$, $\beta_0 = 1$ and $\beta_0 = -1$. In the following graphs, the choice of β_0 is represented by the vertical line in each one of them. The remaining hyperparameters were set to $g = \frac{1}{N} = 0.01$, $\mu_0 = 10$, $\delta_0 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$ and $\Omega_0 = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$. To analyse the influence of the latter, one at a time is changed while the others stay fixed.

Influence of the Hyperparameters on the Marginal Prior of β

β_0: This parameter influences the location of the marginal prior of β. Further, for $\beta_0 = 0$, the marginal prior of β seems to be fairly symmetric around β_0, whereas for $\beta_0 = 1$ and $\beta_0 = -1$, the distribution gets more skewed (see Figures 1-8). The influence on the skewness is symmetric with regards to positive and negative values, implying that the distribution is positively skewed for positive values of β_0 and negatively skewed for negative values of β.

g: Figure 1 shows that the marginal prior of β gets more peaked the larger g is. This is in line with the interpretation of g in the linear regression setup, where it indicates the relative information in the prior.

μ_0: This parameter influences the marginal prior of β only indirectly through the inverted-Wishart distribution we draw Ω from. In Figure 2, one can see that the marginal prior of β gets more peaked the larger μ_0 becomes. μ_0 is the precision parameter of the inverted-Wishart distribution and it also influences the mean of the distribution of Ω, which is $\frac{\Omega_0+c}{\mu_0-1}$, where $c = g(\beta_0 : 1)'\delta_0' Z' Z \delta_0(\beta_0 : 1)$. Fixing Ω_0, if μ_0 gets large, the mean of the distribution we draw from gets smaller in the sense that the entries of the mean matrix become smaller[32]. This reflects the fact, that a priori we assume smaller variances and covariances for the error terms of our

[32]For the mean to be defined, the degrees of freedom need to be larger than 3 in this case. Note that, in the above algorithm, we draw from an inverted-Wishart distribution with $\mu_0 + 2$ degrees of freedom. This means that if $\mu_0 \geq 2$, then $\mu_0 + 2$ will always be larger or equal to 4 and therefore the mean of the inverted-Wishart distribution is well-defined.

model equations. The smaller the variance of the error term, the more precisely the parameter can be identified.

δ_0: Figure 3 shows the influence of $\delta_0 = \binom{\delta_{11,0}}{\delta_{21,0}}$ on the marginal prior of β when both parameters change simultaneously. The marginal prior of β becomes less informative as both components in δ_0 are shrunk to zero. Intuitively, this makes sense. If $\delta_0 \to 0$, we have the case that we assume that the instruments do not help. This implies that β is difficult to identify and hence, the marginal prior of β becomes centred at zero with large spread [33].

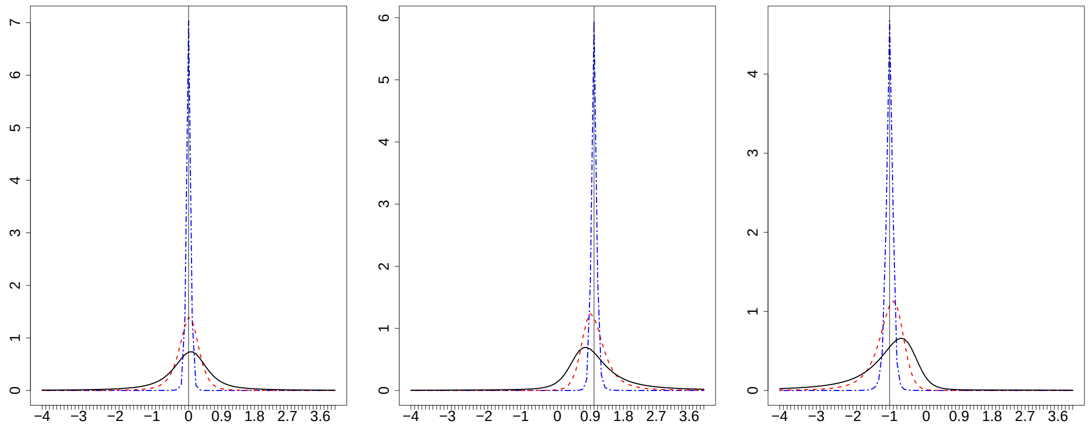
Figure 4 depicts the influence of $\delta_0 = \binom{\delta_{11,0}}{0.5}$ on the marginal prior of β letting only the first component of δ_0 go to zero while fixing the other at 0.5. Figure 5 shows the same for the second component of δ_0, such that $\delta_0 = \binom{0.5}{\delta_{21,0}}$, letting only the second component in δ_0 go to zero while fixing the other at 0.5. In both cases, a reduction in scale towards zero of the component leads to a flatter marginal prior of β_0, similar to the case where both components get shrunk towards zero. The finding that the marginal prior seems to be influenced more by the shrinkage of the second component could possibly be explained by the fact, that in this example, the instruments have not been standardized and that the second instrument, z_8, is larger in scale, such that the factor $\delta_0' Z' Z \delta_0$[34] has more influence on the marginal prior of β.

Ω_0: All components of Ω_0 influence the marginal prior of β only indirectly. Changes in Ω_0 imply changes in the scale matrix of the inverted-Wishart distribution we draw from and ultimately the mean of the latter. For both components of Ω_0, $\omega_{11,0}$ and $\omega_{22,0}$, we see that the larger they are, the more uninformative the marginal prior of β becomes. However, the magnitude of the effects on the marginal prior of β is different for $\omega_{11,0}$ and $\omega_{22,0}$, see Figure 7 and Figure 8, respectively. That possibly occurs due to the fact that ω_{11} interacts differently with β than ω_{22}. Consider the term $(\beta : 1)\Omega^{-1}(\beta : 1)'$, which is equal to $\frac{1}{\omega_{11}\omega_{22}-\omega_{12}^2}(\beta^2\omega_{22} - 2\beta\omega_{12} + \omega_{11})$ as well as the term $(\beta : 1)\Omega^{-1}(\beta_0 : 1)'$, which is equal to $\frac{1}{\omega_{11}\omega_{22}-\omega_{12}^2}(\beta\beta_0\omega_{22} - \beta\omega_{12} - \beta_0\omega_{12} + \omega_{11})$[35]. With respect to $\omega_{22,0}$ we can say that it reflects our prior

[33]Note that the software R defines $0^0 = 1$. Hence, by setting $\delta_0 = \binom{0}{0}$ in the simulation, $p(\beta|\Omega)^g_{RRF}$ will not be zero as the first term of the infinite sum, $0^0 2^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(k+2h+1))}{\Gamma(\frac{1}{2}(k+2h))h!}$, will be positive.

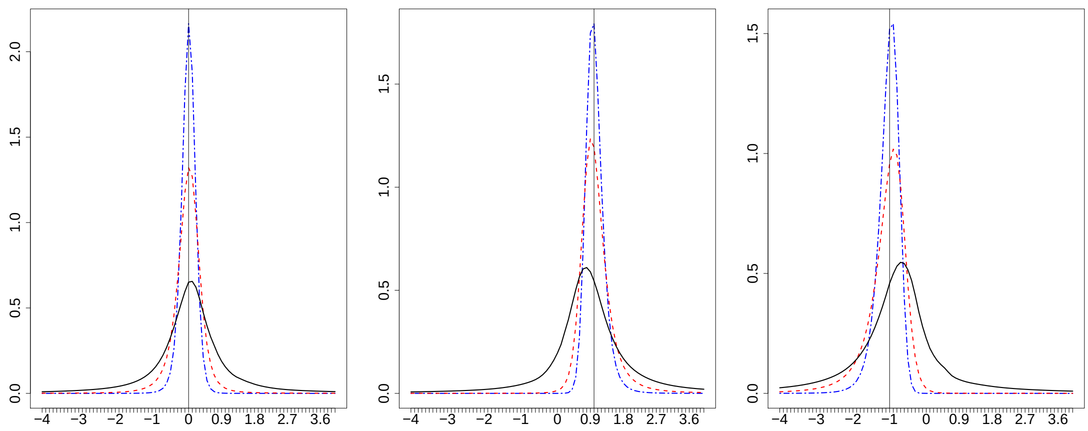
[34]Here in the simulation example, $\delta_0' Z' Z \delta_0$ collapses to $\delta_{11,0}^2 \sum_{j=1}^N z_{1,j}^2 + 2\delta_{11,0}\delta_{12,0} \sum_{j=1}^N z_{1,j} z_{8,j} + \delta_{12,0}^2 \sum_{j=1}^N z_{8,j}^2$, where j is the observation index.

[35]In both cases, we can rewrite the terms in that way since $m = 1$.



1.1: $\beta_0 = 0$ 1.2: $\beta_0 = 1$ 1.3: $\beta_0 = -1$

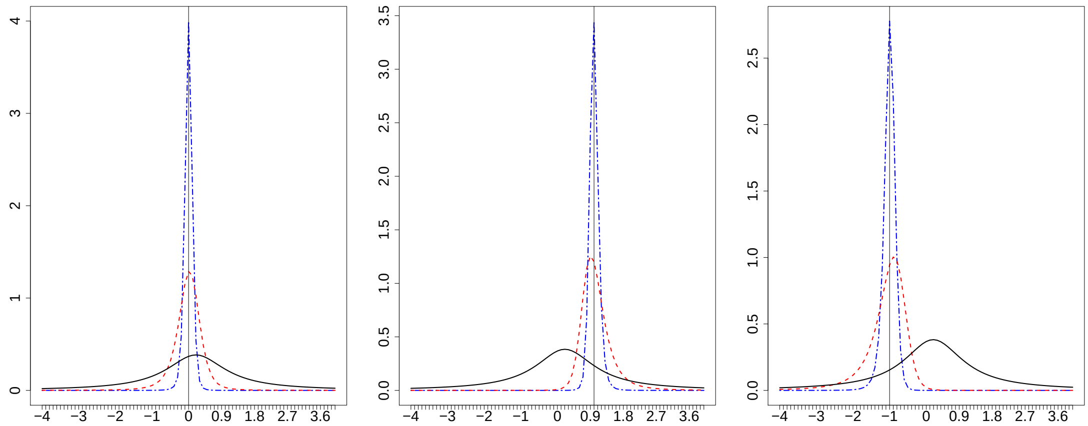
Figure 1: Influence of g on marginal prior of β with $g = 0.25$, $g = 0.01$ and $g = 0.0025$.



2.1: $\beta_0 = 0$ 2.2: $\beta_0 = 1$ 2.3: $\beta_0 = -1$

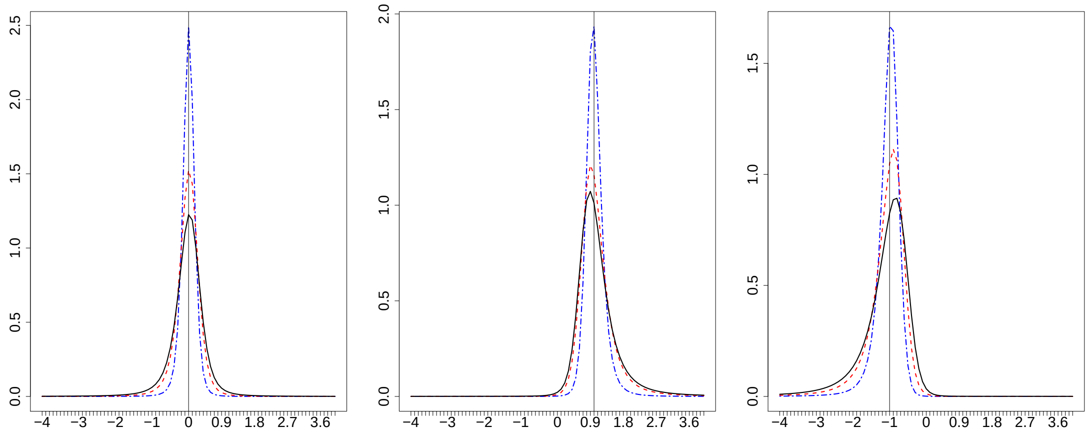
Figure 2: Influence of μ_0 on marginal prior of β with $\mu_0 = 25$, $\mu_0 = 10$ and $\mu_0 = 2$.

assumptions of how strong the instruments are. If we assume that ω_{22} is small, then there is much information in the instruments z_1 and z_8 about the endogenous variable y_2, which helps, in turn, to identify β more precisely. The influence of $\omega_{12,0}$ on the marginal prior of β is more difficult to pin down. Figure 6 shows that the fact whether the distribution of β is more peaked when $\omega_{12,0}$ increases seems to depend on the prior choice of β_0. When β_0 is positive, a larger covariance $\omega_{12,0}$ leads to a more peaked marginal prior of β and, for β_0 being negative, the relationship is the other way around.



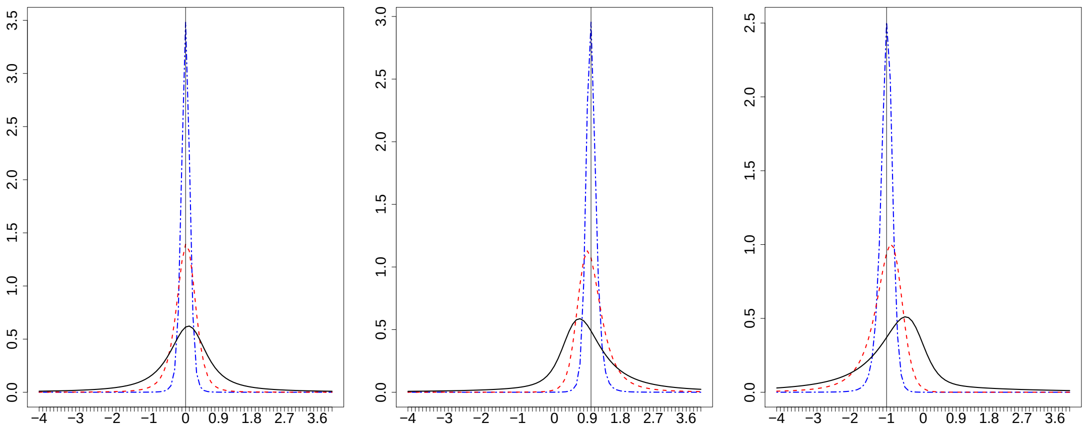
3.1: $\beta_0 = 0$ 3.2: $\beta_0 = 1$ 3.3: $\beta_0 = -1$

Figure 3: Influence of $\delta_0 = \binom{\delta_{11,0}}{\delta_{21,0}}$ on marginal prior of β with $\delta_0 = \binom{1.5}{1.5}$, $\delta_0 = \binom{0.5}{0.5}$ and $\delta_0 = \binom{0.01}{0.01}$.



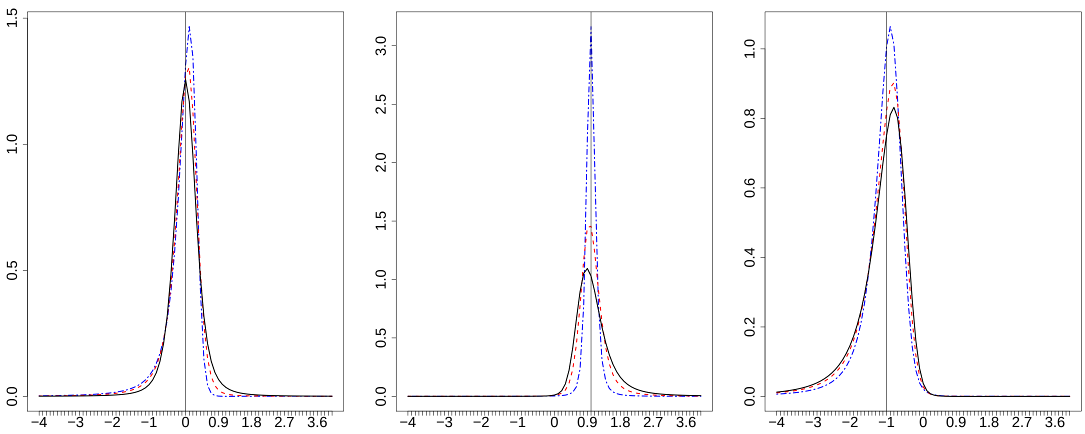
4.1: $\beta_0 = 0$ 4.2: $\beta_0 = 1$ 4.3: $\beta_0 = -1$

Figure 4: Influence of $\delta_0 = \binom{\delta_{11,0}}{0.5}$ on marginal prior of β with $\delta_{11,0} = 1.5$, $\delta_{11,0} = 0.5$ and $\delta_{11,0} = 0.01$.



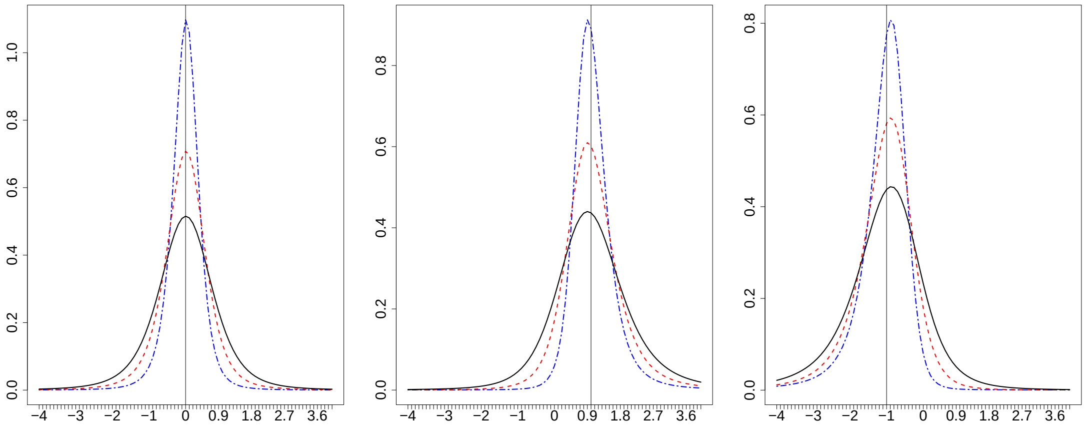
5.1: $\beta_0 = 0$ 5.2: $\beta_0 = 1$ 5.3: $\beta_0 = -1$

Figure 5: Influence of $\delta_0 = \binom{0.5}{\delta_{21,0}}$ on marginal prior of β with $\delta_{12,0} = 1.5$, $\delta_{12,0} = 0.5$ and $\delta_{12,0} = 0.01$.



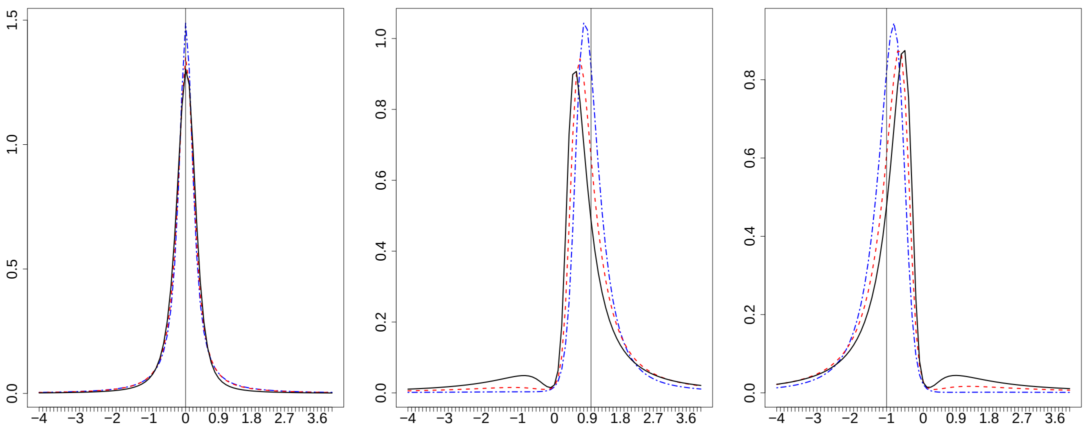
6.1: $\beta_0 = 0$ 6.2: $\beta_0 = 1$ 6.3: $\beta_0 = -1$

Figure 6: Influence of $\Omega_0 = \begin{pmatrix} 1 & \omega_{12,0} \\ \omega_{21,0} & 1 \end{pmatrix}$ on marginal prior of β. Under $\beta_0 = 0$ and $\beta_0 = 1$, $\omega_{12,0} = 0.9$, $\omega_{12,0} = 0.5$ and $\omega_{12,0} = 0.01$. Under $\beta_0 = -1$, $\omega_{12,0} = 0.01$, $\omega_{12,0} = 0.5$ and $\omega_{12,0} = 0.9$.



7.1: $\beta_0 = 0$ 7.2: $\beta_0 = 1$ 7.3: $\beta_0 = -1$

Figure 7: Influence of $\Omega_0 = \begin{pmatrix} \omega_{11} & 0.2 \\ 0.2 & 1 \end{pmatrix}$ on marginal prior of β with $\omega_{11,0} = 2$, $\omega_{11,0} = 5$ and $\omega_{11,0} = 10$.



8.1: $\beta_0 = 0$ 8.2: $\beta_0 = 1$ 8.3: $\beta_0 = -1$

Figure 8: Influence of $\Omega_0 = \begin{pmatrix} 1 & 0.2 \\ 0.2 & \omega_{22} \end{pmatrix}$ on marginal prior of β with $\omega_{22,0} = 2$, $\omega_{22,0} = 5$ and $\omega_{22,0} = 10$.

Posterior

Following Hoogerheide et al. (2007), the resulting joint posterior of β,δ,Ω using the G-prior in combination with the likelihood of the model in (6.4) is

$$\begin{aligned}p(\beta,\delta,\Omega|y_1,y_2) \propto & |\delta' Z' Z\delta|^{\frac{1}{2}}|(\beta:1)\Omega^{-1}(\beta:1)'|^{\frac{1}{2}(k-1)} \\ & \times|g(1+g)|^{\frac{1}{2}}|\Omega|^{-\frac{1}{2}(N+\mu_0+4)} \\ & \times\exp\Big[-\tfrac{1}{2}\text{tr}\{\Omega^{-1}[\Omega_0+(y_1:y_2)'(y_1:y_2)-(1+g)\hat{\Phi}_0' Z' Z\hat{\Phi}_0]\} \\ & -\tfrac{1}{2}\text{tr}\{(1+g)\Omega^{-1}(\delta(\beta:1)-\hat{\Phi}_0)' Z' Z(\delta(\beta:1)-\hat{\Phi}_0)\}\Big],\end{aligned} \tag{6.11}$$

with

$$\hat{\Phi}_0=\left(\tfrac{g}{1+g}\right)\delta_0(\beta_0:1)+\left(\tfrac{1}{1+g}\right)(Z'Z)^{-1}Z'(y_1:y_2).$$

The marginal posterior for (β,Ω) can be written as

$$p^g_{RRF}(\beta,\Omega|y_1,y_2)=p^g_{RRF}(\beta|\Omega,y_1,y_2)q^g_{RRF}(\Omega|y_1,y_2), \tag{6.12}$$

with

$$\begin{aligned}p^g_{RRF}(\beta|\Omega,y_1,y_2) & \propto|(\beta:1)\Omega^{-1}(\beta:1)'|^{-1}\sum_{h=0}^{\infty}\Big(\frac{1}{2}\xi\Big)^h 2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2h+1))}{\Gamma(\frac{1}{2}(k+2h))h!}, \\ q^g_{RRF}(\Omega|y_1,y_2) & \propto|\Omega|^{-\frac{1}{2}(N+\mu_0+3)}\exp\Big[-\tfrac{1}{2}\text{tr}[\Omega^{-1}(\Omega_0+(y_1:y_2)'(y_1:y_2) \\ & \quad +g(\beta_0:1)'\delta_0' Z' Z\delta_0(\beta_0:1))]\Big],\end{aligned}$$

where

$$\xi=(1+g)(\beta:1)\Omega^{-1}(\beta:1)'\bar{\delta}' Z' Z\bar{\delta}\ ,$$

$$\bar{\delta}=[\tfrac{1}{1+g}(Z'Z)^{-1}Z'(y_1:y_2)+\tfrac{g}{1+g}\delta_0(\beta_0:1)]\Omega^{-1}(\beta:1)'[(\beta:1)\Omega^{-1}(\beta:1)']^{-1}.$$

We can see that the marginal posterior of β,Ω has a similar form as the marginal prior of β,Ω. The data now updates the value of $\delta_0(\beta_0:1)$ and, hence, the weights in the sum in $p^g_{RRF}(\beta|\Omega,y_1,y_2)$. Further, the data also updates parts of $q^g_{RRF}(\Omega|y_1,y_2)$, which eventually influences the scale matrix and the degrees of freedom of the distribution of Ω. However, one should note that the tails of the marginal posterior of $\beta|\Omega$ are still Cauchy. This means that the posterior mean is not defined. Therefore, we do not learn from the data with regards to the tails of the posterior distribution in the case of dealing with an IV model.

Algorithm to Calculate the Marginal Posterior Density of β

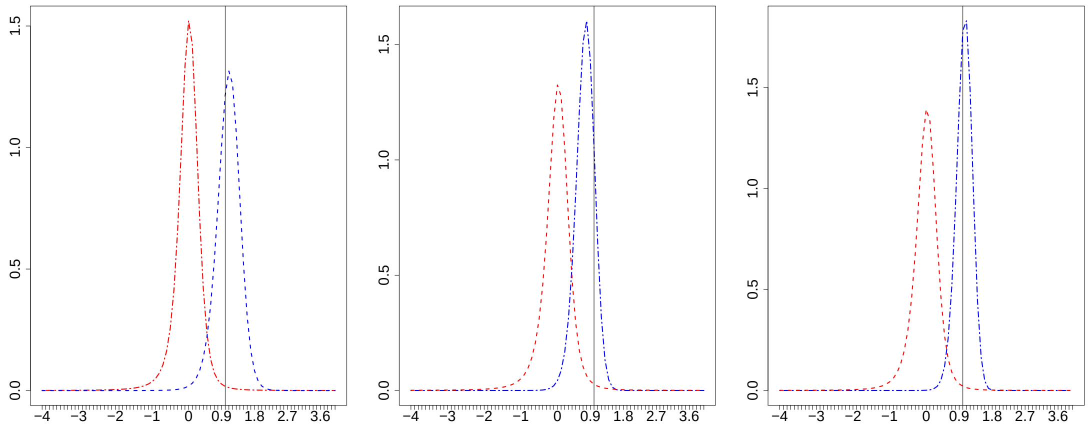
Since the marginal posterior of β,Ω has a similar form as its prior counterpart, Hoogerheide et al. (2007) suggest to use an analogous algorithm:

1. Specify a grid of values of $\beta, with \beta^1, ..., \beta^L$.
2. Generate T values of Ω^t,$t=1,...,T$, from an inverted-Wishart distribution with scale matrix $\Omega_0 + (y_1:y_2)'(y_1:y_2) + g(\beta_0:1)'\delta_0' Z' Z \delta_0(\beta_0:1)$ and $N+\mu_0+2$ degrees of freedom.
3. Compute $p^g_{RRF}(\beta^j|\Omega, y_1, y_2) = \frac{1}{T}\sum_{t=1}^{T} p^g_{RRF}(\beta^l|\Omega^t, y_1, y_2)$ for $l=1,...,L$.
4. Compute the marginal posterior of β, with $p^g_{RRF}(\beta^j|y_1,y_2) = \frac{1}{c} p^g_{RRF}(\beta^l|\Omega, y_1, y_2)$ for $l=1,...,L$, with $c=\sum_{i=1}^{L}(\beta^{i+1}-\beta^i)p^g_{RRF}(\beta^i|\Omega,y_1,y_2)$.

It is of fundamental interest that the resulting marginal posterior of β reflects the information about the true value of β with more and more data, implying consistency[36]. For the posterior analysis, we first choose a setting in which the marginal prior of β is quite informative (peaked) and afterwards we choose values for the hyper-parameters such that the marginal prior of β is relatively uninformative (flatter). In both cases, we repeat the analysis for different sample sizes, namely $N=10$,$N=100$ and $N=500$. In the following graphs, the vertical line represents the true value of β, which is equal to one. For the relatively informative prior, we choose the hyper-parameters to be $g=\frac{1}{N}$, $\mu_0 = 10$, $\Omega_0 = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$ and $\delta_0 = \binom{0.5}{0.5}$, whereas, for the relatively uninformative prior, we set the parameters to $g=\frac{1}{N}$, $\mu_0=2$, $\Omega_0 = \begin{pmatrix} 2 & 0.9 \\ 0.9 & 2 \end{pmatrix}$ and $\delta_0 = \binom{0.1}{0.1}$.

We can summarise the findings as follows. First, under $\beta_0 = 0$, the effects in terms of location and how peaked the marginal posterior is are the same under the more informative and the less informative prior settings (see Figure 9 and Figure 10). Further, a growing sample size leads to a more peaked posterior in both scenarios considered with the changes being of comparable magnitude.

[36]In the given setting, differences in the marginal posteriors are not overly large even though sample sizes differ considerably. This might be related to the fact that the marginal posterior of β has Cauchy tails.

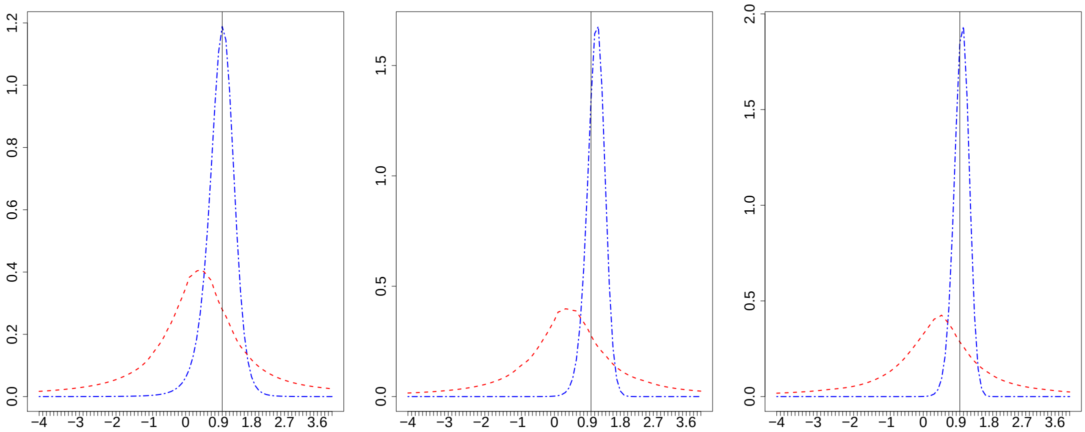


9.1: $N = 10$ | 9.2: $N = 100$ | 9.3: $N = 500$

Figure 9: Marginal prior of β (relatively informative) and marginal posterior of β under $\beta_0 = 0$ for different sample sizes.

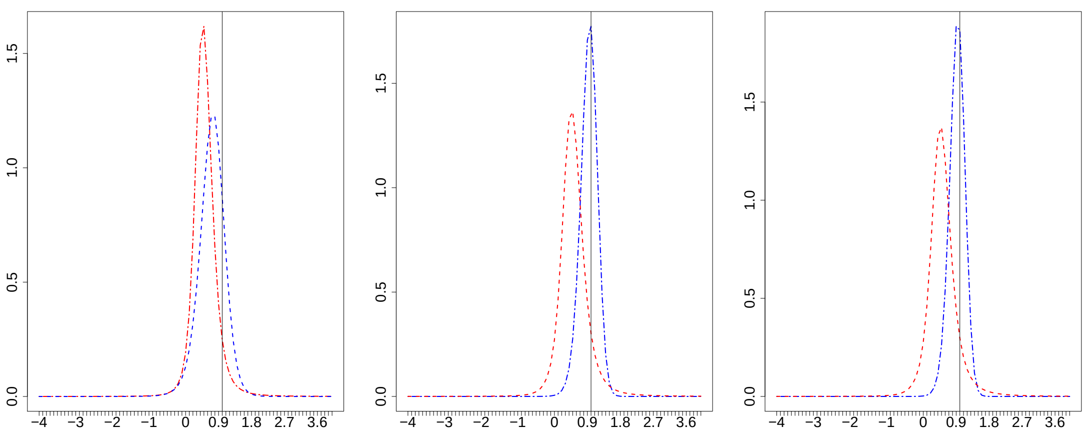
Second, under $\beta_0 = 0.5$, in the more informative setting, the marginal posterior of β is somewhat more peaked and centred around the true value of β for smaller sample sizes in comparison to the uninformative setting (see Figure 11 and Figure 12). This is to be expected since $\beta_0 = 0.5$ is closer to the true value than $\beta_0 = 0$ in this example. However, we observe the exactly opposite effect if our chosen hyper-parameter β_0 is not close to the true value, such that the prior information is in conflict with the data. In this context, see Figure 13 and Figure 14 for $\beta_0 = -0.5$.

In summary, for a more informative choice of the hyper-parameters, the marginal posterior will be more peaked, although this effect is limited. At the same time, however, the more our prior beliefs about β_0 differ from the true value of β, the more will an informative prior lead us astray, i.e. we need more data for the posterior to be centred around the true value. Therefore, if we do not have much information about our structural parameter in advance, e.g., through theoretical considerations, it is reasonable to choose the hyper-parameters in such a way that our prior is not that informative, especially if the sample is not large. However, in the most common applications of the given framework, we do possess samples of a considerable size ($N > 100$), in which case the posterior is not overly sensitive to hyper-parameter choices. The latter mitigates the problem of having to elicit a high number of hyper-parameters in advance.



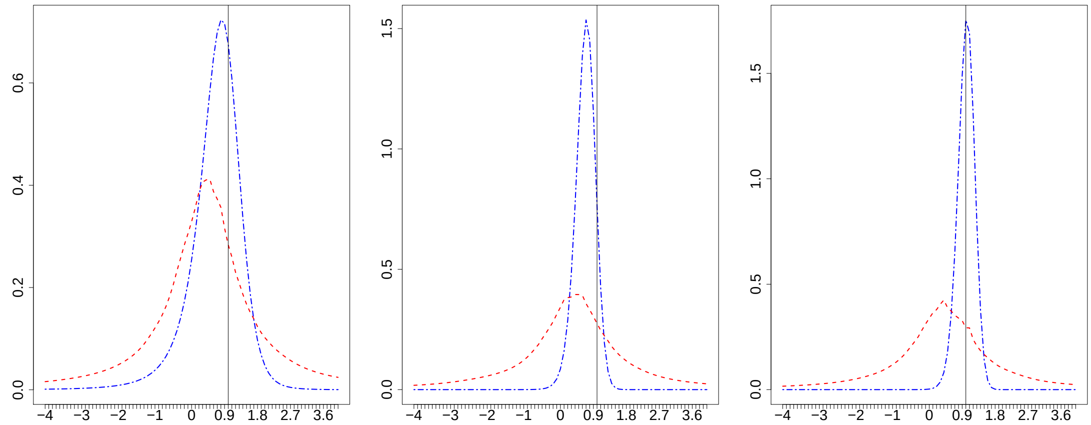
10.1: $N = 10$ 10.2: $N = 100$ 10.3: $N = 500$

Figure 10: Marginal prior of β (relatively uninformative) and marginal posterior of β under $\beta_0 = 0$ for different sample sizes.



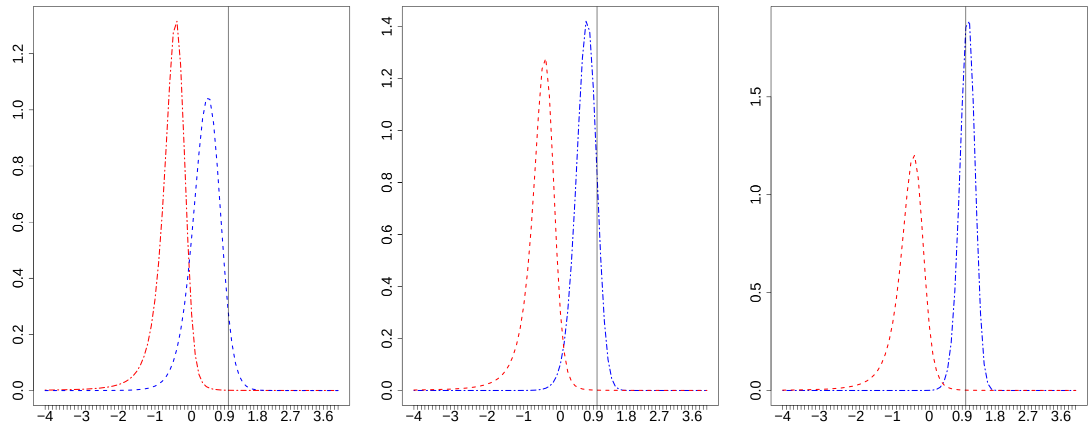
11.1: $N = 10$ 11.2: $N = 100$ 11.3: $N = 500$

Figure 11: Marginal prior of β (relatively informative) and marginal posterior of β under $\beta_0 = 0.5$ for different sample sizes.



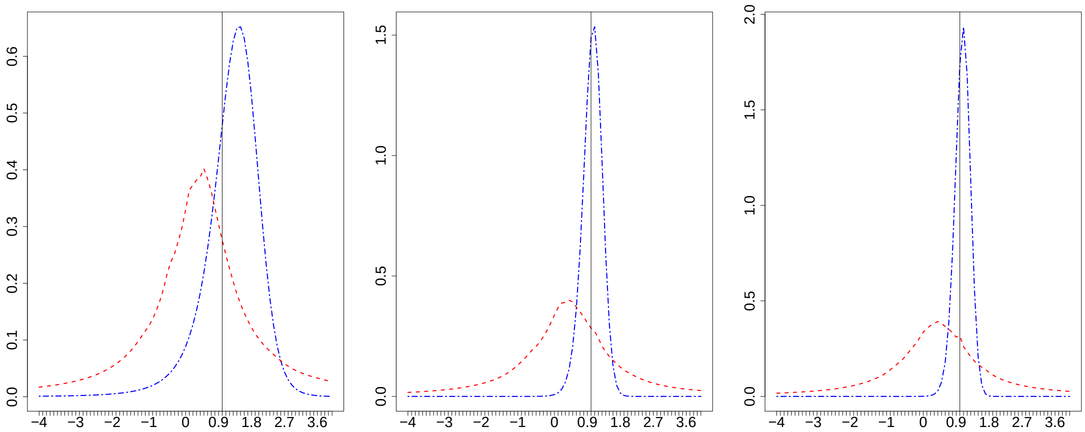
12.1: $N = 10$ 12.2: $N = 100$ 12.3: $N = 500$

Figure 12: Marginal prior of β (relatively uninformative) and marginal posterior of β under $\beta_0 = 0.5$ for different sample sizes.



13.1: $N = 10$ 13.2: $N = 100$ 13.3: $N = 500$

Figure 13: Marginal prior of β (relatively informative) and marginal posterior of β under $\beta_0 = -0.5$ for different sample sizes.



14.1: $N = 10$ 14.2: $N = 100$ 14.3: $N = 500$

Figure 14: Marginal prior of β (relatively uninformative) and marginal posterior of β under $\beta_0 = -0.5$ for different sample sizes.

Including Exogenous Regressors and Model Uncertainty

To make this approach somewhat more relevant to potential empirical applications, we now include exogenous regressors and we introduce model uncertainty as we have defined it in section 4.1. Just as a reminder and to make it easier to keep track of the notation used, we repeat the relevant equations below:

in **SF** with

$$\begin{aligned} y_1 &= y_2\beta + X_i\gamma_i + \epsilon_1, \\ y_2 &= Z_i\Delta_i + X_i\Lambda_i + V_2, \end{aligned}$$

in **RRF** with

$$\begin{aligned} y_1 &= Z_i(\Delta_i\beta) + X_i(\gamma_i + \Lambda_i\beta) + v_1, \\ Y_2 &= Z_i\Delta_i + X_i\Lambda_i + V_2, \end{aligned}$$

in **URF** with

$$\begin{aligned} y_1 &= Z_i\tau_i + X_i\xi_i + v_1, \\ Y_2 &= Z_i\Delta_i + X_i\Lambda_i + V_2. \end{aligned}$$

As we want to concentrate on β, we rewrite the **RRF** slightly, such that we base the analysis on a kind of **hybrid** between **RRF** and **URF**:

$$\begin{aligned} y_1 &= Z_i(\Delta_i\beta) + X_i\xi_i + v_1 \\ Y_2 &= Z_i\Delta_i + X_i\Lambda_i + V_2. \end{aligned} \tag{6.13}$$

The resulting likelihood from the system of equations (6.13) is

$$\begin{aligned} p(y_1, Y_2|\beta, \Delta_i, \Lambda_i, \xi_i, \Omega) =& (2\pi)^{-\frac{N(m+1)}{2}} |\Omega|^{-\frac{N}{2}} \\ & \times \exp\Big[-\tfrac{1}{2}\mathrm{tr}\{\Omega((y_1 : Y_2) - Z_i(\Delta_i\beta : \Delta_i) - X_i(\xi_i : \Lambda_i))' \\ & \times ((y_1 : Y_2) - Z_i(\Delta_i\beta : \Delta_i) - X_i(\xi_i : \Lambda_i))\}\Big]. \end{aligned} \tag{6.14}$$

Using this paramterization of the model implies that we are not interested in the structural parameter γ.

Derivation of the Natural Conjugate Prior for the Slightly more Advanced Setting

We basically follow the same four steps suggested by Hoogerheide et al. (2007) in order to be able to define a prior on the parameters of the hybrid system in (6.13). Hence, we start with a prior specification on the parameters of the **URF**. Let us assume that ξ_i and Λ_i are a priori independent from Δ_i, τ_i and Ω. Then

$$p_{URF}(\tau_i, \Delta_i, \Omega, \xi_i, \Lambda_i) = p_{URF}(\tau_i, \Delta_i, \Omega) p(\xi_i, \Lambda_i) = p_{URF}(\tau_i, \Delta_i|\Omega) p(\Omega) p(\xi_i, \Lambda_i).$$

Further, we assume a matrix normal prior for ξ_i and Λ_i. Hence,

$$\begin{aligned} p(\xi_i, \Lambda_i) =& (2\pi)^{\frac{-k_{xi}(m+1)}{2}} |cX_i'X_i|^{\frac{m+1}{2}} \\ & \times \exp\Big[-\tfrac{1}{2}\mathrm{tr}\{[(\xi_i : \Lambda_i) - (\xi_0 : \Lambda_0)]' cX_i'X_i [(\xi_i : \Lambda_i) - (\xi_0 : \Lambda_0)]\}\Big]. \end{aligned}$$

If we now set $\xi_0 = 0$ and $\Lambda_0 = 0$, we have

$$p(\xi_i, \Lambda_i) = (2\pi)^{\frac{-k_{xi}(m+1)}{2}} |cX_i'X_i|^{\frac{m+1}{2}} \exp\Big[-\tfrac{1}{2}\mathrm{tr}\{(\xi_i : \Lambda_i)' cX_i'X_i (\xi_i : \Lambda_i)\}\Big].$$

With $p_{URF}(\tau_i, \Delta|\Omega)$ being a matrix normal prior on (τ_i, Δ_i) given Ω and $p(\Omega)$ being an inverted-Wishart prior on Ω, we have

$$\begin{aligned} p_{URF}(\tau_i, \Delta_i|\Omega) =& (2\pi)^{-\frac{1}{2}k_{zi}(m+1)} |A_{0,i}|^{\frac{1}{2}m} |\Omega|^{-\frac{1}{2}k_{zi}} \exp\Big[-\tfrac{1}{2}\mathrm{tr}\{\Omega^{-1}((\tau_i : \Delta_i) - \Delta_{0,i}(\beta_0 : I_m))' \\ & \times A_{0,i}((\tau_i : \Delta_i) - \Delta_{0,i}(\beta_0 : I_m))\}\Big], \\ p(\Omega) =& \frac{|\Omega_0|^{\frac{\mu_0}{2}}}{2^{\frac{\mu_0(m+1)}{2}} \Gamma_{m+1}\left(\frac{\mu_0}{2}\right)} |\Omega|^{-\frac{1}{2}(\mu_0+m+2)} \exp\Big[-\tfrac{1}{2}\mathrm{tr}(\Omega^{-1}\Omega_0)\Big], \\ p(\xi_i, \Lambda_i) =& (2\pi)^{\frac{-k_{xi}(m+1)}{2}} |cX_i'X_i|^{\frac{m+1}{2}} \exp\Big[-\tfrac{1}{2}\mathrm{tr}\{(\xi_i : \Lambda_i)' cX_i'X_i (\xi_i : \Lambda_i)\}\Big], \end{aligned}$$

where, again, $A_{0,i}$ is a $k_{zi} \times k_{zi}$ scale matrix. β_0 is a vector location hyperparameter that reflects prior information on the structural parameter β, Δ_0 is a $k_{zi} \times m$ matrix with prior information on the reduced form parameter matrix Δ_i, Ω_0 is a $(m+1) \times (m+1)$ scale matrix reflecting the prior information on the covariance matrix Ω and μ_0 are the prior degrees of freedom in the distribution of Ω.

In the second step, the parameter matrix $(\tau_i : \Delta_i)$ is normalized with

$$\Theta_i = Q_i(\tau_i : \Delta_i)W, \tag{6.15}$$

such that $\Theta_i = (A_{0,i} + Z_i'Z_i)^{\frac{1}{2}}(\tau_i : \Delta_i)\Omega^{-\frac{1}{2}}$ and $\Theta_{0,i} = (A_{0,i} + Z_i'Z_i)^{\frac{1}{2}}\Delta_{0,i}(\beta_0 : I_m)\Omega^{-\frac{1}{2}}$. W is a 2×2 scale matrix with $W = \Omega^{-\frac{1}{2}}$. Then, the prior from step one can be reformulated as

$$p_{URF}(\Theta_i, \Omega, \xi_i, \Lambda_i) = p_{URF}(\Theta_i|\Omega)p(\Omega)p(\xi_i, \Lambda_i),$$

with

$$\begin{aligned} p_{URF}(\Theta_i|\Omega) =& p_{URF}(\Theta_i(\tau_i, \Delta_i)|\Omega)|J((\tau_i, \Delta_i), \Theta_i)|, \\ =&(2\pi)^{-\frac{1}{2}k_{zi}m}|(A_{0,i} + Z_i'Z_i)^{-\frac{1}{2}'}A_{0,i}(A_{0,i} + Z_i'Z_i)^{-\frac{1}{2}}|^{\frac{1}{2}m} \\ &\times \exp\Big[-\tfrac{1}{2}\mathrm{tr}\{(\Theta_i - \Theta_{0,i})'(A_{0,i} + Z_i'Z_i)^{-\frac{1}{2}'}A_{0,i}(A_{0,i} + Z_i'Z_i)^{-\frac{1}{2}}(\Theta_i - \Theta_{0,i})\}\Big], \end{aligned}$$

where $J(a, b)$ is the Jacobian of the transformation from $a \to b$.

Following step three from the process discussed at the beginning of subsection 6.2, using SVD, Θ_i can now be expressed in terms of β, Δ_i, ζ, with ζ_i being a $k_{zi} - 1$ dimensional vector which equals the smallest singular value pre-multiplied by a matrix

$$\Theta_i = Q_i[\Delta_i(\beta : 1) + \Delta_{\perp,i}\zeta_i(\beta : 1)_\perp]W, \tag{6.16}$$

where $(\beta : 1)_\perp$ is of dimension 1×2 and $\Delta_{\perp,i}$ is of dimension $k_{zi} \times (k_{zi} - 1)$, such that:

$$\begin{gathered} (\beta : 1)_\perp WW'(\beta : 1)' \equiv 0, \text{ and } (\beta : 1)_\perp WW'(\beta : 1)_\perp' \equiv 1, \\ \Delta_{\perp,i}'Q_iQ_i'\Delta_i \equiv 0, \text{ and } \Delta_{\perp,i}'Q_iQ_i'\Delta_{\perp,i} \equiv I_{k_{zi}-1}. \end{gathered}$$

Then,

$$p_{URF}(\beta, \Delta_i, \zeta_i, \Omega, \xi_i, \Lambda_i) = p_{URF}(\Theta_i(\Delta_i, \beta, \zeta_i)|\Omega)|J(\Theta_i(\Delta_i, \beta, \zeta_i))|p(\Omega)p(\xi_i, \Lambda_i).$$

In the fourth step, the prior for $(\beta, \Delta_i, \Omega, \xi_i, \Lambda_i)$ is obtained by setting $\zeta_i = 0$ in $p_{URF}(\beta, \Delta_i, \zeta_i, \Omega, \xi_i, \Lambda_i)$. Using Hoogerheide et al. (2007) equation (25), we can calculate

$$\begin{aligned} p_{RRF}(\beta, \Delta_i, \Omega) \propto& |\Delta_i'(A_{0,i} + Z_i'Z_i)\Delta_i|^{\frac{1}{2}}|(\beta : I_m)\Omega^{-1}(\beta : I_m)'|^{\frac{1}{2}(k_{zi}-m)}|A_{0,i}|^{\frac{1}{2}m}|\Omega|^{-\frac{1}{2}(\mu_0+2m+2)} \\ &\times \exp\Big[-\tfrac{1}{2}\mathrm{tr}(\Omega^{-1}\Omega_0) - \tfrac{1}{2}\mathrm{tr}\{\Omega^{-1}(\Delta_i(\beta : I_m) \\ &- \Delta_{0,i}(\beta_0 : I_m))'A_{0,i}(\Delta_i(\beta : I_m) - \Delta_{0,i}(\beta_0 : I_m))\}\Big]. \end{aligned}$$

We have that, $p(\beta,\Delta_i,\Omega,\xi_i,\Lambda_i)=p_{RRF}(\beta,\Delta_i|\Omega)p(\Omega)p(\xi_i,\Lambda_i)$ (see page 71 in Hoogerheide et al. (2007)). Therefore, the resulting prior is

$$\begin{aligned}
p(\beta,\Delta_i,\Omega,\xi_i,\Lambda_i) =&(2\pi)^{-\frac{mk_{zi}+k_{xi}(m+1)}{2}}|\Delta_i'(A_{0,i}+Z_i'Z_i)\Delta_i|^{\frac{1}{2}}|(\beta:I_m)\Omega^{-1}(\beta:I_m)'|^{\frac{1}{2}(k_{zi}-m)}|A_{0,i}|^{\frac{1}{2}m}\\
&\times\frac{|\Omega_0|^{\frac{\mu_0}{2}}}{2^{\frac{\mu_0(m+1)}{2}}\Gamma_{m+1}\left(\frac{\mu_0}{2}\right)}|\Omega|^{-\frac{1}{2}(\mu_0+2m+2)}|cX_i'X_i|^{\frac{m+1}{2}}\\
&\times\exp\Big[-\tfrac{1}{2}\mathrm{tr}\{\Omega^{-1}(\Delta_i(\beta:I_m)-\Delta_{0,i}(\beta_0:I_m))'\\
&\times A_{0,i}(\Delta_i(\beta:I_m)-\Delta_{0,i}(\beta_0:I_m))\}\Big]\exp\Big[-\tfrac{1}{2}\mathrm{tr}(\Omega^{-1}\Omega_0)\Big]\\
&\times\exp\Big[-\tfrac{1}{2}\mathrm{tr}\{(\xi_i:\Lambda_i)'cX_i'X_i(\xi_i:\Lambda_i)\}\Big].
\end{aligned} \tag{6.17}$$

Combining the prior with the likelihood from (6.14), we get the following posterior

$$\begin{aligned}
p(\beta,\Delta_i,\Omega,\xi_i,\Lambda_i|y_1,Y_2) =&(2\pi)^{-\frac{1}{2}(k_{zi}m+(k_{xi}+N)(m+1))}|\Delta_i'(A_{0,i}+Z_i'Z_i)\Delta_i|^{\frac{1}{2}}\\
&\times|(\beta:I_m)\Omega^{-1}(\beta:I_m)'|^{\frac{1}{2}(k_{zi}-m)}|A_{0,i}|^{\frac{1}{2}m}\\
&\times\frac{|\Omega_0|^{\frac{\mu_0}{2}}}{2^{\frac{\mu_0(m+1)}{2}}\Gamma_{m+1}\left(\frac{\mu_0}{2}\right)}|\Omega|^{-\frac{1}{2}(\mu_0+2m+2+N)}|cX_i'X_i|^{\frac{m+1}{2}}\\
&\times\exp\Big[-\tfrac{1}{2}\mathrm{tr}\{\Omega^{-1}(\Delta_i(\beta:I_m)-\Delta_{0,i}(\beta_0:I_m))'\\
&\times A_{0,i}(\Delta_i(\beta:I_m)-\Delta_{0,i}(\beta_0:I_m))\}\Big]\\
&\times\exp\Big[-\tfrac{1}{2}\mathrm{tr}(\Omega^{-1}\Omega_0)\Big]\exp\Big[-\tfrac{1}{2}\mathrm{tr}\{(\xi_i:\Lambda_i)'cX_i'X_i(\xi_i:\Lambda_i)\}\Big]\\
&\times\exp\Big[-\tfrac{1}{2}\mathrm{tr}\{\Omega((y_1:Y_2)-Z_i(\Delta_i\beta:\Delta_i)-X_i(\xi_i:\Lambda_i))'\\
&\times((y_1:Y_2)-Z_i(\Delta_i\beta:\Delta_i)-X_i(\xi_i:\Lambda_i))\}\Big].
\end{aligned} \tag{6.18}$$

Choosing a G-prior type of prior, we set $A_{0,i} = gZ_i'Z_i$ and $\Delta_{0,i} = 0$. Then, the posterior is equal to

$$\begin{aligned}
p(\beta,\Delta_i,\Omega,\xi_i,\Lambda_i|y_1,Y_2) =&(2\pi)^{-\frac{1}{2}(k_{zi}m+(k_{xi}+N)(m+1))}|\Delta_i'(gZ_i'Z_i+Z_i'Z_i)\Delta_i|^{\frac{1}{2}}\\
&\times|(\beta:I_m)\Omega^{-1}(\beta:I_m)'|^{\frac{1}{2}(k_{zi}-m)}|gZ_i'Z_i|^{\frac{1}{2}m}\\
&\times\frac{|\Omega_0|^{\frac{\mu_0}{2}}}{2^{\frac{\mu_0(m+1)}{2}}\Gamma_{m+1}\left(\frac{\mu_0}{2}\right)}|\Omega|^{-\frac{1}{2}(\mu_0+2m+2+N)}|cX_i'X_i|^{\frac{m+1}{2}}\\
&\times\exp\left[-\tfrac{1}{2}\mathrm{tr}\{\Omega^{-1}(\Delta_i(\beta:I_m))'gZ_i'Z_i(\Delta_i(\beta:I_m))\}\right]\\
&\times\exp\left[-\tfrac{1}{2}\mathrm{tr}(\Omega^{-1}\Omega_0)\right]\\
&\times\exp\left[-\tfrac{1}{2}\mathrm{tr}\{(\xi_i:\Lambda_i)'cX_i'X_i(\xi_i:\Lambda_i)\}\right]\\
&\times\exp\left[-\tfrac{1}{2}\mathrm{tr}\{\Omega((y_1:Y_2)-Z_i(\Delta_i\beta:\Delta_i)-X_i(\xi_i:\Lambda_i))'\right.\\
&\left.\times((y_1:Y_2)-Z_i(\Delta_i\beta:\Delta_i)-X_i(\xi_i:\Lambda_i))\}\right].
\end{aligned} \tag{6.19}$$

The posterior in (6.19) is more complex than the one in (6.11). It therefore will be difficult to integrate out Δ_i, ξ_i or Λ_i analytically even if we stick to the case where $m = 1$.

If we were using a model without exogenous regressors, but were interested in dealing with model uncertainty in terms of the choice of instruments, we could introduce model uncertainty in the same way as defined above. For $A_{0,i} = gZ_i'Z_i$ and $\Delta_{0,i} = 0$, the posterior then equals

$$\begin{aligned}
p(\beta,\Delta_i,\Omega|y_1,Y_2) =&\frac{|\Omega_0|^{\frac{\mu_0}{2}}}{2^{\frac{\mu_0(m+1)}{2}}\Gamma_{m+1}\left(\frac{\mu_0}{2}\right)}|\Omega|^{-\frac{1}{2}(N+\mu_0+2m+2)}|g\Delta_i'Z_i'Z_i\Delta_i|^{\frac{1}{2}}|gZ_i'Z_i|^{\frac{1}{2}m}\\
&\times|(\beta:I_m)\Omega^{-1}(\beta:I_m)'|^{\frac{1}{2}(k_{zi}-m)}\\
&\times\exp\left[-\tfrac{1}{2}\mathrm{tr}(\Omega^{-1}\Omega_0)\right]\exp\left[-\tfrac{1}{2}\mathrm{tr}\{g\Omega^{-1}(\Delta_i(\beta:I_m))'Z_i'Z_i(\Delta_i(\beta:I_m))\}\right]\\
&\times\exp\left[-\tfrac{1}{2}\mathrm{tr}\{\Omega((y_1:Y_2)-Z(\Delta_i\beta:\Delta_i))'((y_1:Y_2)-Z_i(\Delta_i\beta:\Delta_i))\}\right].
\end{aligned}$$

By defining $\bar{\Delta}_i = (\frac{1}{1+c})(Z_i'Z_i)^{-1}Z_i'(y_1 : Y_2)\Omega^{-1}(\beta : I_m)'[(\beta : I_m)\Omega^{-1}(\beta : I_m)']^{-1}$ and $\psi_i = (1+g)(\beta : I_m)\Omega^{-1}(\beta : I_m)'\bar{\Delta}_i{}'Z_i'Z_i\bar{\Delta}_i$, the posterior can be rewritten[37], such that

$$
\begin{aligned}
p(\beta, \Delta_i, \Omega | y_1, Y_2) = & \frac{|\Omega_0|^{\frac{\mu_0}{2}}}{2^{\frac{\mu_0(m+1)}{2}}\Gamma_{m+1}\left(\frac{\mu_0}{2}\right)}(2\pi)^{-\frac{(N(m+1)+k_{zi}m)}{2}}|\Omega|^{-\frac{1}{2}(N+\mu_0+2m+2)}g^{\frac{1}{2}m(k_{zi}-1)}((1+g)g)^{\frac{1}{2}m} \\
& \times |Z_i'Z_i|^{\frac{1}{2}(m-1)}|Z_i'Z_i|^{\frac{1}{2}}|\Delta_i Z_i'Z_i\Delta_i|^{\frac{1}{2}}|(\beta : I_m)\Omega^{-1}(\beta : I_m)'|^{-\frac{1}{2}(m+1)} \\
& \times |(\beta : I_m)\Omega^{-1}(\beta : I_m)'|^{\frac{1}{2}}|(\beta : I_m)\Omega^{-1}(\beta : I_m)'|^{\frac{1}{2}k_{zi}} \\
& \times \exp\Big[-\tfrac{1}{2}\{\Omega^{-1}(\Omega_0 + (y_1 : Y_2)'(I_N - Z_i'Z_i(Z_i'Z_i)^{-1}Z_i')(y_1 : Y_2))\}\Big] \\
& \times \exp\Big[-\tfrac{1}{2}\{\Omega^{-1}(y_1 : Y_2)'Z_i(Z_i'Z_i)^{-1}Z_i'(y_1 : Y_2) + \psi_i\}\Big] \\
& \times \exp\Big[-\tfrac{1}{2}\{(1+g)(\beta : I_m)\Omega^{-1}(\beta : I_m)'(\Delta_i - \bar{\Delta}_i)'Z_i'Z_i(\Delta_i - \bar{\Delta}_i)\}\Big].
\end{aligned}
$$

When $m = 1$, we can follow the same analytical steps as in the case without model uncertainty and can integrate out δ_i:

$$
\begin{aligned}
p(\beta, \Omega | y_1, y_2) = & \frac{|\Omega_0|^{\frac{\mu_0}{2}}}{2^{\mu_0}\Gamma_2\left(\frac{\mu_0}{2}\right)}(2\pi)^{-\frac{(2N+k_{zi})}{2}}|\Omega|^{-\frac{1}{2}(N+\mu_0+4)}g^{\frac{1}{2}k_{zi}}(1+g)^{\frac{1}{2}} \\
& \times |(\beta : 1)\Omega^{-1}(\beta : 1)'|^{-1} \\
& \times \exp\Big[-\tfrac{1}{2}\{\Omega^{-1}(\Omega_0 + (y_1 : y_2)'(I_N - Z_i'Z_i(Z_i'Z_i)^{-1}Z_i')(y_1 : y_2))\}\Big] \\
& \times \exp\Big[-\tfrac{1}{2}\{\Omega^{-1}(y_1 : y_2)'Z_i(Z_i'Z_i)^{-1}Z_i'(y_1 : y_2) + \psi_i\}\Big] \\
& \times \int \Bigg(|(\beta : 1)\Omega^{-1}(\beta : 1)'\delta_i Z_i'Z_i\delta_i|^{\frac{1}{2}}|(\beta : 1)\Omega^{-1}(\beta : 1)'|^{\frac{1}{2}k_{zi}}|Z_i'Z_i|^{\frac{1}{2}} \\
& \times \exp\Big[-\tfrac{1}{2}\{(1+g)(\beta : 1)\Omega^{-1}(\beta : 1)'(\delta_i - \bar{\delta}_i)'Z_i'Z_i(\delta_i - \bar{\delta}_i)\}\Big]\Bigg)d\delta_i,
\end{aligned}
$$

and

$$
\begin{aligned}
p(\beta, \Omega | y_1, y_2) = & \frac{|\Omega_0|^{\frac{\mu_0}{2}}}{2^{\mu_0}\Gamma_2\left(\frac{\mu_0}{2}\right)}(2\pi)^{-\frac{(2N+k_{zi})}{2}}|\Omega|^{-\frac{1}{2}(N+\mu_0+4)}g^{\frac{1}{2}k_{zi}}(1+g)^{\frac{1}{2}} \\
& \times |(\beta : 1)\Omega^{-1}(\beta : 1)'|^{-1} \\
& \times \exp\Big[-\tfrac{1}{2}\{\Omega^{-1}(\Omega_0 + (y_1 : y_2)'(I_N - Z_i'Z_i(Z_i'Z_i)^{-1}Z_i')(y_1 : y_2))\}\Big] \\
& \times \exp\Big[-\tfrac{1}{2}\{\Omega^{-1}(y_1 : y_2)'Z_i(Z_i'Z_i)^{-1}Z_i'(y_1 : y_2) + \psi_i\}\Big] \\
& \times (2\pi)^{\frac{1}{2}k_{zi}}(1+g)^{\frac{1}{2}k_{zi}}\sum_{j=0}^{\infty}\frac{(\frac{1}{2}\psi_i)^j}{j!}\exp[\tfrac{1}{2}\psi_i]2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k_{zi}+2j+1))}{\Gamma(\frac{1}{2}(k_{zi}+2j))},
\end{aligned}
$$

[37]See appendix F.2. in Hoogerheide et al. (2007)

as well as

$$
\begin{aligned}
p(\beta,\Omega|y_1,y_2) =& \frac{|\Omega_0|^{\frac{\mu_0}{2}}}{2^{\frac{2\mu_0+1}{2}}\Gamma_2\big(\frac{\mu_0}{2}\big)}(2\pi)^{-\frac{(2N)}{2}}|\Omega|^{-\frac{1}{2}(N+\mu_0+4)}g^{\frac{1}{2}k_{zi}}(1+g)^{\frac{1+k_{zi}}{2}} \\
& \times |(\beta:1)\Omega^{-1}(\beta:1)'|^{-1} \\
& \times \exp\Big[-\tfrac{1}{2}\{\Omega^{-1}(\Omega_0+(y_1:y_2)'(I_N-Z_i'Z_i(Z_i'Z_i)^{-1}Z_i')(y_1:y_2))\}\Big] \\
& \times \exp\Big[-\tfrac{1}{2}\{\Omega^{-1}(y_1:y_2)'Z_i(Z_i'Z_i)^{-1}Z_i'(y_1:y_2)+\psi_i\}\Big] \\
& \times \sum_{j=0}^{\infty}\frac{(\frac{1}{2}\psi_i)^j}{j!}\exp[\tfrac{1}{2}\psi_i]\frac{\Gamma(\frac{1}{2}(k_{zi}+2j+1))}{\Gamma(\frac{1}{2}(k_{zi}+2j))}.
\end{aligned}
$$

We could then use the same algorithm stated in the previous subsection to obtain a sample of the marginal posterior of β and use the median as point estimate of β. Further, we would need to calculate the posterior model probabilities as weights.

Three possible approaches come to mind in order to achieve that. First, we could use a BIC approximation. The latter would probably be relatively fast to implement, but also has its disadvantages as already discussed in section 2.

The second option is to use the Gelfand-Dey method. Gelfand and Dey (1994) calculate the marginal likelihoods in the following way. Suppose we have a sample $\theta_{i,b}$, with $b = 1,...,B$, from the posterior $p(\beta,\Omega,|y_1,y_2) = p(\theta_i|y_1,y_2)$. Then, the estimator of the marginal likelihood is

$$
\hat{p}(y_1,y_2|M_i) = \Big[\frac{1}{B}\sum_{b=1}^{B}\frac{p^*(\theta_{i,b})}{p(y_1,y_2|\theta_{i,b},M_i)p(\theta_{i,b}|M_i)}\Big]^{-1},
$$

with $p^*(\theta_{i,b})$ being thinner tailed than $p(y_1,y_2|\theta_{i,b},M_i)p(\theta_{i,b}|M_i)$. How $p^*(\cdot)$ should be chosen is an open question. One could potentially start by using a normal distribution for the parameters β and a IW distribution for Ω with many degrees of freedom and centred on some estimate (e.g., the ML estimate).

The third option would be to use a generalized version of the Savage-Dickey density ratio to compute Bayes factors (see Verdinelli and Wassermann, 1995). Suppose $\theta = (\delta_i,\phi)$, with $\phi = (\beta,\Omega)$. Comparing two nested models can be seen as testing the hypothesis that some of the coefficients in δ_i are restricted to be zero, namely the ones for the instruments which are not part of the model M_i. Defining M_δ as the full model which contains all the possible instruments, we can formulate $M_0 : \delta_i = \delta_0$ vs.

$M_1 : \delta_i \neq \delta_0$. The Bayes factor for M_0 vs. M_1 equals

$$BF_{\delta_i,\delta} = \frac{\int L(\delta_0, \phi) p_0(\phi) \, d\phi}{\int \int L(\delta, \phi) p(\delta, \phi) \, d\delta \, d\phi},$$

where the prior under M_0 is $p_0(\phi) = p(\phi | \delta_i = \delta_0)$. If we then want to compare model M_i and M_j to each other, we can use

$$BF_{\delta_i,\delta_j} = BF_{\delta_i,\delta} \times BF^{-1}_{\delta_j,\delta}.$$

Which one of these possibilities would give us the most precise estimate of the posterior model probabilties is unclear and would need further investigation. The above possibilities for calculating posterior model probabilities could also be used when exogenous regressors are included in the model, presupposed we are able to obtain a sample from that more complex posterior in that case. Whether there is actually something to gain by using this framework, particularly with regards to the precision of the estimate of the posterior of β, is an open question and further research is needed.

7 Conclusion

The main goal of this thesis is to further develop and understand in more detail the methodology of a BMA-framework for IV models. The major contributions and findings can be summarized as follows.

We begin with a simulation study focusing on the effects of ignoring endogeneity in a BMA framework. This study shows that not only do we obtain biased posterior point estimates (false posterior means), but also wrong posterior inclusion probabilities, if endogeneity is not accounted for properly. Further, at the core of this thesis is the evaluation and advancement of the Karl and Lenkoski (2012) approach. In this context, we introduce adapted prior structures and investigate their influence on the outcome of the estimation procedure by means of simulation studies. Additionally, we develop a tool to deal with the classification of variables with regards to being endogenous or exogenous. To achieve the latter, we wrote an extended code base, which we are currently integrating into the ivbma R package in order to make it accessible to other researchers. Within an empirical application example, we demonstrate that the choices of the priors do matter. Furthermore, we can point out that using our proposed prior structures and endogeneity classification approach helps to obtain improved results in terms of out-of-sample-prediction accuracy. Lastly, we try to formulate the problem in a different way by extending the approach of Hoogerheide et al. (2007). Here, we conduct the first analytical steps towards a different solution for contextualizing model uncertainty within an IV model framework. However, a considerable amount of work is still left to be done in order to find ways to implement this approach.

Future research possibilities within the outlined methodological framework, e.g., include the investigation of different prior structures. More precisely, the IW-prior for Σ could potentially be replaced by a prior which is more interpretable, such that domain insights can be reflected and incorporated in prior information more easily. Also, it might be beneficial to investigate the structure of the prior and posterior in the Karl and Lenkoski (2012) setting in a more analytical manner in order to be more definitive regarding recommendations for applied researchers. Further, as to the latter approach, another interesting extension would be to put a hyper-prior on the prior parameter g when using the gprvar-g or g-g prior as well as to consider using a different prior on the model space. We know from the linear regression framework that these choices have tangible effects on the estimation outcome.

Another potential research avenue, particularly with respect to empirical applications investigating corruption, are spatial effects. Here, extending the approach along the lines of the work by Cuaresma and Feldkircher (2013) would be interesting. Finally, a natural extension to the modelling framework of this thesis would be a panel data setting as in Moral-Benito (2016).

References

BERNARDO, J. M. AND A. F. SMITH (1994): "Bayesian Theory," *Wiley*.

CLYDE, M. AND E. I. GEORGE (2004): "Model Uncertainty," *Statistical Science*, 19, 81–94.

CONLEY, T. G., C. B. HANSEN, R. E. MCCULLOCH, AND P. E. ROSSI (2008): "A Semi-Parametric Bayesian Approach to the Instrumental Variable Problem," *Journal of Econometrics*, 144, 276–305.

CUARESMA, J. C. AND G. DOPPELHOFER (2007): "Nonlinearities in Cross-Country Growth Regressions: A Bayesian Averaging of Thresholds (BAT) Approach," *Journal of Macroeconomics*, 29, 541–554.

CUARESMA, J. C. AND M. FELDKIRCHER (2013): "Spatial Filtering, Model Uncertainty and the Speed of Income Convergence in Europe," *Journal of Applied Econometrics*, 28, 720–741.

DICKEY, J. M. (1971): "The Weighted Likelihood Ratio, Linear Hypotheses on Normal Location Parameters," *The Annals of Mathematical Statistics*, 42, 204–223.

DRÈZE, J. H. (1976): "Bayesian Limited Information Analysis of the Simultaneous Equations Model," *Econometrica*, 44, 1045–1075.

——— (1977): "Bayesian Regression Analysis Using Poly-t Densities," *Journal of Econometrics*, 6, 329–354.

DRÈZE, J. H. AND J.-F. RICHARD (1983): "Bayesian Analysis of Simultaneous Equation Systems," *Handbook of Econometrics*, 1, 517–598.

DURLAUF, S. N., A. KOURTELLOS, AND C. M. TAN (2008): "Are Any Growth Theories Robust?" *The Economic Journal*, 118, 329–346.

——— (2012): "Is God in the Details? A Reexamination of the Role of Religion in Economic Growth," *Journal of Applied Econometrics*, 27, 1059–1075.

FERNÁNDEZ, C., E. LEY, AND M. F. STEEL (2001a): "Benchmark Priors for Bayesian Model Averaging," *Journal of Econometrics*, 100, 381–427.

——— (2001b): "Model Uncertainty in Cross-Country Growth Regressions," *Journal of Applied Econometrics*, 16, 563–576.

FOSTER, D. P. AND E. I. GEORGE (1994): "The Risk Inflation Criterion for Multiple Regression," *The Annals of Statistics*, 22, 1947–1975.

GELFAND, A. E. AND D. DEY (1994): "Bayesian Model Choice: Asymptotics and Exact Calculations," *Journal of the Royal Statistical Society. Series B (Methodological)*, 56, 501–514.

GEORGE, E. E. (2010): "Dilution Priors: Compensating for Model Space Redundancy," *IMS Collections Borrowing Strength: Theory Powering Applications- A Festschrift for Lawrence D. Brown*, 6, 158–165.

GNEITING, T. AND A. E. RAFTERY (2007): "Strictly Proper Scoring Rules, Prediction and Estimation," *Journal of the American Statistical Association*, 102, 359–378.

GOOD, I. J. (1952): "Rational Decision," *Journal of Royal Statistical Society Series B*, 14, 107–114.

HAYASHI, F. (2000): "Econometrics," *Princeton University Press*.

HEYWOOD, P. M. (2015): "Routledge Handbook of Political Corruption," *Routledge*.

HJORT, N. L. AND G. CLAESKENS (2003): "Frequentist Model Averaging Estimators," *Journal of the American Statistical Association*, 98, 879–899.

HOETING, J. A., D. MADIGAN, A. E. RAFTERY, AND C. T. VOLINSKY (1999): "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14, 382 – 417.

HOLMES, C., D. DENISON, AND B. MALLICK (2002): "Bayesian Model Order Determination and Basis Selection for Seemingly Unrelated Regression," *Journal of Computational and Graphical Statistics*, 11, 533 – 551.

HOOGERHEIDE, L., F. KLEIBERGEN, AND H. K. VAN DIJK (2007): "Natural Conjugate Priors for the Instrumental Variables Regression Model Applied to the Angrist-Krueger Data," *Journal of Econometrics*, 138, 63 – 103.

JETTER, M. AND C. PARMETER (2018): "Sorting Through Global Corruption Determinants: Institutions and Education Matter - Not Culture," *World Development*, 279–294.

KARL, A. AND A. LENKOSKI (2012): "Instrumental Variable Bayesian Model Averaging via Conditional Bayes Factors," *arXiv preprint arXiv:1202.5846*.

KASS, R. E. AND A. E. RAFTERY (1995): "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795.

KIM, J. (2002): "Limited Information Likelihood and Bayesian Analysis," *Journal of Econometrics*, 107, 175–193.

KLEIBERGEN, F. (1997): "Bayesian Simultaneous Equation Analysis Using Equality Restricted Random Variables." *American Statistical Association*, 141–146.

KLEIBERGEN, F. AND H. K. VAN DIJK (1998): "Bayesian Simultaneous Equations Analysis Using Reduced Rank Structures," *Econometric Theory*, 14, 701 – 743.

KLEIBERGEN, F. AND E. ZIVOT (2003): "Bayesian and Classical Approaches to Instrumental Variable Regression," *Journal of Econometrics*, 114, 29 – 72.

KOOP, G. (2003): "Bayesian Econometrics," *John Wiley & Sons Ltd.*

KOOP, G., R. LEON-GONZALEZ, AND R. STRACHAN (2012): "Bayesian Model Averaging in the Instrumental Variable Regression Model," *Journal of Econometrics*, 171, 237 – 250.

LEAMER, E. (1985): "Sensitivity Analysis Would Help," *American Economic Review*, 57, 308–313.

LENKOSKI, A., T. S. EICHER, AND A. E. RAFTERY (2014): "Two-Stage Bayesian Model Averaging in Endogenous Variable Models," *Econometric Reviews*, 33, 122–151.

LEY, E. AND M. F. J. STEEL (2007): "Jointness in Bayesian Variable Selction with Applications to Growth Regressions," *Journal of Macroeconomics*, 29, 476–493.

——— (2009): "On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regressions," *Journal of Applied Econometrics*, 24, 651–674.

——— (2012): "Mixture of g-Priors for Bayesian Model Averaging with Economic Applications," *Journal of Econometrics*, 171, 251–266.

LIANG, F., R. PAULO, G. MOLINA, M. A. CLYDE, AND J. O. BERGER (2008): "Mixtures of g-Priors for Bayesian Variable Selection," *Journal of the American Statistical Association*, 103, 410–423.

LOPES, H. F. AND N. G. POLSON (2014): "Bayesian Instrumental Variables: Priors and Likelihoods," *Econometric Reviews*, 33, 100–121.

MADDALA, G. S. (1976): "Weak Priors and Sharp Posteriors in Simultaneous Equation Models," *Econometrica*, 44, 345 – 351.

MADIGAN, D. AND J. YORK (1995): "Bayesian Graphical Models for Discrete Data," *International Statistics Review*, 63, 215 – 232.

MARINARI, E. AND G. PARISI (1992): "Simulated Tempering: A new Monte Carlo Scheme," *Europhysics Letters*, 19, 451 – 458.

MARUYAMA, Y. AND E. I. GEORGE (2011): "Fully Bayes Factors with a Generalized g-Prior," *The Annals of Statistics*, 39, 2740 – 2765.

MONTGOMERY, J. AND B. NYHAN (2010): "Bayesian Model Averaging Theoretical Developments and Practical Implications," *Political Analysis*, 18, 245 – 270.

MORAL-BENITO, E. (2012): "Determinants of Economic Growth: A Bayesian Panel Data Approach," *The Review of Economics and Statistics*, 94, 566 – 579.

——— (2015): "Model Averaging in Economics: An Overview," *Journal of Economic Surveys*, 29, 46–75.

——— (2016): "Growth Empirics in Panel Data Under Model Uncertainty and Weak Exogeneity," *Journal of Applied Econometrics*.

PARKINSON, D. AND A. R. LIDDLE (2013): "Bayesian Model Averaging in Astrophysics: A Review," *Statistical Analysis and Data Mining*, 6, 3–14.

RAFTERY, A. E. (1995): "Bayesian Model Selection in Social Research," *Sociological Methodology*, 25, 111–163.

RAFTERY, A. E., D. MADIGAN, AND J. A. HOETING (1997): "Bayesian Model Averaging for Linear Regression Models," *Journal of the American Statistical Association*, 92, 179 – 191.

ROBERT, C. P. (2007): "The Bayesian Choice : From Decision-Theoretic Foundations to Computational Implementation," *New York : Springer,*, 2nd ed.

ROSE-ACKERMAN, S. (1999): "Corruption and Government - Causes, Consequences and Reform," *Cambridge University Press*.

ROSSI, P., G. ALLENBY, AND R. MCCULLOCH (2006): "Bayesian Statistics and Marketing," *New York: Wiley*.

SALA-I-MARTIN, X. (1997): "I Just Ran Two Million Regressions," *The American Economic Review*, 87, 178–183.

SALA-I-MARTIN, X., G. DOPPELHOFER, AND R. I. MILLER (2004): "Determinants of Long Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," *The American Economic Review*, 94, 813–835.

STEEL, M. F. J. (2019): "Model Averaging and its Use in Economics," *Journal of Economic Literature*, forthcoming.

TROTTA, R. (2008): "Bayes in the Sky: Bayesian Inference and Model Selection in Cosmology," *Contemporary Physics*, 49, 71 – 104.

TSANGARIDES, C. G. (2004): "A Bayesian Approach to Model Uncertainty," *IMF Working Paper*.

TSURUMI, H. (1985): "Limited Information Bayesian Analysis of a Structural Coefficient in a Simultaneous Equations System," *Communications in Statistics-Theory and Methods*, 14, 1103 – 1120.

VERDINELLI, I. AND L. WASSERMANN (1995): "Computing Bayes Factors Using a Generalisation of the Savage-Dickey Density Ratio," *Journal of the American Statistical Association*, 90, 614 – 618.

WAGENMAKERS, E.-J., T. LODEWYCKX, H. KURIYAL, AND R. GRASMAN (2010): "Bayesian Hypothesis Testing for Psychologists: A Tutorial on the Savage-Dickey Method," *Cognitive Psychology*, 60, 158–189.

WANG, H., X. ZHANG, AND G. ZOU (2009): "Frequentist Model Averaging Estimators: A Review," *Journal of Systems Science and Complexity*, 22, 732 – 748.

YEUNG, K. Y., R. E. BUMGARNER, AND A. E. RAFTERY (2005): "Bayesian Model Averaging: Development of an Improved Multi-Class, Gene Selection and Classification Tool for Microarray Data," *Bioinformatics*, 21, 2394 – 2402.

ZELLNER, A. (1970): "Estimation of Regression Relationships Containing Unobservable Independent Variables," *International Economic Review*, 11, 441 – 454.

——— (1986): "On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions," *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno De Finetti*, 6, 233 – 243.

ZEUGNER, S. AND M. FELDKIRCHER (2015): "Bayesian Model Averaging Employing Fixed and Flexible Priors: The BMS Package for R," *Journal of Statistical Software*, 68, 1–37.

A Some Distributions

The Matrix Normal Distribution

Let X be a $N \times p$ real random matrix. Further, let M be a $N \times p$ real location matrix, V a real positive-definite $p \times p$ scale matrix and U a real positive-definite $N \times N$ scale matrix. If $X \sim MN_{N,p}(M, U, V)$, i.e. X follows a matrix normal distribution, then the probability density function of X reads

$$p(X|M,U,V) = \frac{\exp\left(-\frac{1}{2}\mathrm{tr}\left[V^{-1}(X-M)'U^{-1}(X-M)\right]\right)}{(2\pi)^{\frac{Np}{2}}|V|^{\frac{N}{2}}|U|^{\frac{p}{2}}},$$

where tr denotes the trace operator.

There is an equivalence between the multivariate normal distribution and the matrix normal distribution. $X \sim MN_{N,p}(M, U, V)$ holds if and only if $\mathrm{vec}(X) \sim N_{Np}(\mathrm{vec}(M), V \otimes U)$, i.e. $\mathrm{vec}(X)$ follows a multivariate normal distribution with mean vector $\mathrm{vec}(M)$ and covariance matrix $V \otimes U$. That is, the probability density function of $\mathrm{vec}(X)$ reads

$$p(\mathrm{vec}(X)|\mathrm{vec}(M), V \otimes U) = \frac{\exp\left(-\frac{1}{2}(\mathrm{vec}(X)-\mathrm{vec}(M))'(V \otimes U)^{-1}(\mathrm{vec}(X)-\mathrm{vec}(M))\right)}{|2\pi(V \otimes U)|^{\frac{1}{2}}}.$$

The Inverse-Wishart Distribution

Let Ω be a $p \times p$ real positive-definite random matrix. Further, let Ω_0 be a $p \times p$ real positive-definite scale matrix and μ_0 a real scalar with $\mu_0 > p - 1$, denoting the degrees of freedom. If $\Omega \sim W_p^{-1}(\Omega_0, \mu_0)$, i.e. Ω follows an inverted Wishart distribution, then the probability density function of Ω reads

$$p(\Omega|\Omega_0, \mu_0) = \frac{|\Omega_0|^{\frac{\mu_0}{2}}}{2^{\frac{\mu_0 p}{2}}\Gamma_p(\frac{\mu_0}{2})}|\Omega|^{-\frac{\mu_0+p+1}{2}}\exp(-\tfrac{1}{2}\mathrm{tr}[\Omega_0\Omega^{-1}]),$$

where $\Gamma_p(\cdot)$ is the multivariate Gamma function, defined as

$$\Gamma_p(a) = \pi^{\frac{p(p-1)}{4}}\prod_{j=1}^{p}\Gamma\left[a + \tfrac{(1-j)}{2}\right].$$

The Inverse-Gamma Distribution

Let x be a univariate random variable with support $x > 0$ and probability distribution Inverse-Gamma. Then, the probability density function of x reads

$$p(x|\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{-\alpha-1} \exp\{-\tfrac{\beta}{x}\},$$

where $\Gamma(\cdot)$ is the univariate Gamma function, α is a real shape parameter with $\alpha > 0$ and β is a real scale parameter with $\beta > 0$.

B Additional Results: Simulation Study - Ignoring Endogeneity

Variable	$PIP_{N=100}$	$PIP_{N=10000}$	True Value	$Mean_{N=100}$	$Mean_{N=10000}$
$\mathbf{y_2}$	1.00	1.00	1	1.47	1.5
$\mathbf{x_1}$	0.83	1.00	-1.8	-1.59	-1.79
$\mathbf{x_2}$	0.68	1.00	1.5	1.14	1.51
x_3	0.08	0.01	0	-0.01	0.00
x_4	0.06	0.02	0	0.01	0.00
x_5	0.09	0.01	0	-0.03	0.00
x_6	0.08	0.01	0	-0.01	0.00
x_7	0.06	0.01	0	0.00	0.00
x_8	0.08	0.02	0	-0.01	0.00
x_9	0.09	0.01	0	0.02	0.00
x_{10}	0.08	0.01	0	0.02	0.00
$\mathbf{x_{11}}$	1.00	1.00	1	1.00	1.00
$\mathbf{x_{12}}$	1.00	1.00	-1.5	-1.47	-1.5
x_{13}	0.07	0.01	0	0.00	0.00
x_{14}	0.09	0.01	0	0.01	0.00
x_{15}	0.06	0.01	0	0.00	0.00

Table A.1: Estimation results for sample sizes $N = 100$ and $N = 10000$ for version one, based on (4.8). Variables that truly are part of the model are marked in bold blue.

Variable	$\text{PIP}_{N=100}$	$\text{PIP}_{N=10000}$	True Value	$\text{Mean}_{N=100}$	$\text{Mean}_{N=10000}$
$\mathbf{y_2}$	1.00	1.00	1	1.44	1.50
$\mathbf{x_1}$	0.75	1.00	-1.8	-1.39	-1.80
$\mathbf{x_2}$	0.67	1.00	1.5	1.10	1.50
x_3	0.11	0.01	0	-0.01	0.00
x_4	0.13	0.01	0	-0.01	0.00
x_5	0.09	0.01	0	-0.04	0.00
x_6	0.09	0.01	0	0.02	0.00
x_7	0.10	0.01	0	0.02	0.00
x_8	0.09	0.01	0	0.03	0.00
x_9	0.12	0.01	0	0.00	0.00
x_{10}	0.52	1.00	0	0.99	1.66
$\mathbf{x_{11}}$	1.00	1.00	1	0.99	1.00
$\mathbf{x_{12}}$	1.00	1.00	-1.5	-1.46	-1.50
x_{13}	0.09	0.03	0	0.01	0.00
x_{14}	0.10	0.02	0	0.00	0.00
x_{15}	0.12	0.02	0	-0.02	0.00

Table A.2: Estimation results for sample sizes $N = 100$ and $N = 10000$ for version two, based on (4.9). Variables that truly are part of the model are marked in bold blue.

Variable	$\text{PIP}_{N=100}$	$\text{PIP}_{N=10000}$	True Value	$\text{Mean}_{N=100}$	$\text{Mean}_{N=10000}$
$\mathbf{y_2}$	1.00	1.00	1	1.40	1.50
$\mathbf{x_1}$	0.83	1.00	-1.8	-1.55	-1.79
$\mathbf{x_2}$	0.71	1.00	1.5	1.17	1.49
x_3	0.15	0.02	0	-0.12	0.00
x_4	0.08	0.02	0	-0.02	0.00
x_5	0.08	0.01	0	0.00	0.00
x_6	0.09	0.02	0	-0.01	0.00
x_7	0.10	0.02	0	0.00	0.00
x_8	0.10	0.03	0	-0.04	0.00
x_9	0.12	0.04	0	0.09	0.00
x_{10}	0.60	1.00	0	1.11	1.65
$\mathbf{x_{11}}$	1.00	1.00	1	1.05	1.00
$\mathbf{x_{12}}$	1.00	1.00	-1.5	-1.88	-2.00
x_{13}	0.09	0.02	0	0.00	0.00
x_{14}	0.09	0.01	0	0.00	0.00
x_{15}	0.11	0.01	0	0.01	0.00

Table A.3: Estimation results for sample sizes $N = 100$ and $N = 10000$ for version three, based on (4.10). Variables that truly are part of the model are marked in bold blue.

C Additional Results: Simulation Study - Different Prior Settings

Full Tables of Results from Section 5

	PIP_a	PIP_b	PIP_c	PIP_d	True Value	$Mean_a$	$Mean_b$	$Mean_c$	$Mean_d$
$\mathbf{y_2}$	1.00	1.00	1.00	1.00	1	0.99	1.00	0.99	0.99
$\mathbf{y_3}$	1.00	1.00	1.00	1.00	-1	-0.97	-1.01	-0.95	-0.97
$\mathbf{x_1}$	0.96	0.94	0.94	0.97	-1.8	-1.45	-1.70	-1.70	-1.72
$\mathbf{x_2}$	0.91	0.87	0.85	0.91	1.5	1.15	1.34	1.30	1.37
x_3	0.38	0.08	0.18	0.17	0	-0.01	-0.02	-0.01	-0.03
x_4	0.38	0.08	0.18	0.16	0	0.01	0.00	0.01	0.01
x_5	0.36	0.06	0.15	0.14	0	-0.01	0.00	0.00	0.00
x_6	0.37	0.08	0.18	0.17	0	0.02	0.01	0.01	0.02
x_7	0.35	0.06	0.16	0.15	0	-0.02	-0.01	-0.01	-0.01
x_8	0.36	0.07	0.17	0.16	0	-0.03	-0.01	-0.02	-0.01
x_9	0.37	0.08	0.18	0.18	0	0.01	0.00	0.01	0.01
x_{10}	0.35	0.06	0.16	0.14	0	0.01	0.00	0.01	0.01
$\mathbf{x_{11}}$	1.00	1.00	1.00	1.00	1	1.01	1.02	1.02	1.02
$\mathbf{x_{12}}$	1.00	1.00	1.00	1.00	-1.5	-1.49	-1.52	-1.51	-1.51
x_{13}	0.19	0.03	0.19	0.19	0	0.00	0.00	0.00	0.00
x_{14}	0.17	0.02	0.16	0.15	0	0.00	0.00	0.00	0.00
x_{15}	0.19	0.02	0.17	0.16	0	-0.01	0.00	0.00	0.00

Table A.4: Estimated posterior inclusion probabilities (PIP) and posterior mean (Mean) for variables in structural equation (5.6) for different prior settings (a-d) for correctly specified system. Based on an average of 100 samples of size $N = 100$ and $\nu = 4$ in prior for Σ.

True Value	$q_{0.05,\mathbf{a}}$	$q_{0.95,\mathbf{a}}$	$q_{0.05,\mathbf{b}}$	$q_{0.95,\mathbf{b}}$	$q_{0.05,\mathbf{c}}$	$q_{0.95,\mathbf{c}}$	$q_{0.05,\mathbf{d}}$	$q_{0.95,\mathbf{d}}$
1	0.88	1.09	0.91	1.09	0.88	1.09	0.90	1.08
-1	-1.22	-0.75	-1.27	-0.77	-1.19	-0.73	-1.21	-0.75
-1.8	-2.14	-0.68	-2.37	-0.98	-2.48	-0.85	-2.37	-1.04
1.5	0.42	1.85	0.65	2.02	0.49	2.10	0.72	2.03
0	-0.46	0.45	-0.15	0.06	-0.26	0.27	-0.29	0.17
0	-0.45	0.46	-0.09	0.11	-0.25	0.27	-0.18	0.24
0	-0.48	0.44	-0.06	0.05	-0.23	0.22	-0.18	0.19
0	-0.40	0.44	-0.07	0.08	-0.22	0.28	-0.18	0.24
0	-0.44	0.38	-0.11	0.05	-0.27	0.22	-0.22	0.17
0	-0.47	0.36	-0.09	0.08	-0.29	0.19	-0.23	0.17
0	-0.39	0.44	-0.08	0.10	-0.22	0.26	-0.18	0.24
0	-0.38	0.43	-0.05	0.08	-0.20	0.25	-0.14	0.20
1	0.80	1.22	0.84	1.21	0.81	1.23	0.83	1.20
-1.5	-1.72	-1.26	-1.72	-1.32	-1.74	-1.28	-1.71	-1.30
0	-0.09	0.07	-0.01	0.01	-0.08	0.07	-0.08	0.06
0	-0.07	0.07	0.00	0.01	-0.07	0.07	-0.05	0.06
0	-0.08	0.07	-0.01	0.00	-0.08	0.07	-0.07	0.06

Table A.5: Estimated lower 0.05 quantile ($q_{0.05}$) and upper 0.95 quantile ($q_{0.95}$) for coefficients of variables in structural equation (5.6) for correctly specified system for different prior settings (a-d). Based on an average of 100 samples of size $N = 100$ and $\nu = 4$ in prior for Σ.

	$\text{PIP}_\mathbf{a}$	$\text{PIP}_\mathbf{b}$	$\text{PIP}_\mathbf{c}$	$\text{PIP}_\mathbf{d}$	$\text{PIP}_\mathbf{BMS}$	True Value	$\text{Mean}_\mathbf{a}$	$\text{Mean}_\mathbf{b}$	Mean	$\text{Mean}_\mathbf{d}$	$\text{Mean}_\mathbf{BMS}$
$\mathbf{y_2}$	1.00	1.00	1.00	1.00	1.00	1	0.98	0.99	0.98	0.98	0.98
$\mathbf{y_3}$	0.99	0.99	0.99	0.99	0.99	-1	-0.97	-0.98	-0.97	-0.97	-0.98
$\mathbf{x_1}$	0.90	0.77	0.83	0.83	0.73	-1.8	-1.28	-1.49	-1.55	-1.54	-1.43
$\mathbf{x_2}$	0.83	0.64	0.71	0.71	0.58	1.5	1.01	1.09	1.16	1.17	1.01
x_3	0.42	0.11	0.17	0.18	0.08	0	0.01	0.00	0.01	0.00	0.00
x_4	0.42	0.11	0.17	0.17	0.09	0	0.02	0.02	0.02	0.03	0.02
x_5	0.40	0.08	0.14	0.14	0.06	0	-0.02	-0.01	-0.01	-0.01	-0.01
x_6	0.41	0.09	0.16	0.16	0.07	0	0.02	0.01	0.02	0.02	0.02
x_7	0.40	0.09	0.15	0.16	0.07	0	-0.02	-0.01	-0.01	-0.02	-0.01
x_8	0.40	0.10	0.16	0.16	0.08	0	-0.04	-0.02	-0.03	-0.03	-0.02
x_9	0.41	0.10	0.17	0.17	0.09	0	0.02	-0.01	0.00	0.01	-0.01
x_{10}	0.39	0.08	0.15	0.15	0.06	0	0.02	0.01	0.01	0.01	0.01
$\mathbf{x_{11}}$	0.99	0.99	0.98	0.99	0.99	1	1.01	1.04	1.03	1.03	1.04
$\mathbf{x_{12}}$	1.00	1.00	1.00	1.00	1.00	-1.5	-1.48	-1.53	-1.52	-1.52	-1.52
x_{13}	0.24	0.04	0.18	0.18	0.09	0	0.00	0.00	0.00	0.00	0.00
x_{14}	0.22	0.03	0.15	0.15	0.08	0	0.00	0.00	0.00	0.00	0.00
x_{15}	0.23	0.04	0.17	0.17	0.08	0	-0.01	0.00	0.01	-0.01	0.00

Table A.6: Estimated posterior inclusion probabilities (PIP) and posterior means (Mean) for variables in structural equation (5.6) for incorrectly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 100$ and $\nu = 4$ in prior for Σ.

True Value	$q_{0.05,\mathbf{a}}$	$q_{0.95,\mathbf{a}}$	$q_{0.05,\mathbf{b}}$	$q_{0.95,\mathbf{b}}$	$q_{0.05,\mathbf{c}}$	$q_{0.95,\mathbf{c}}$	$q_{0.05,\mathbf{d}}$	$q_{0.95,\mathbf{d}}$
1	0.85	1.10	0.86	1.12	0.85	1.11	0.85	1.11
-1	-1.22	-0.72	-1.25	-0.72	-1.23	-0.71	-1.23	-0.71
-1.8	-2.12	-0.45	-2.53	-0.56	-2.56	-0.63	-2.55	-0.62
1.5	0.22	1.85	0.27	2.08	0.31	2.16	0.32	2.15
0	-0.53	0.60	-0.17	0.23	-0.27	0.36	-0.28	0.38
0	-0.57	0.58	-0.19	0.21	-0.30	0.32	-0.31	0.33
0	-0.61	0.56	-0.16	0.12	-0.31	0.27	-0.30	0.27
0	-0.50	0.55	-0.15	0.21	-0.26	0.32	-0.26	0.33
0	-0.56	0.49	-0.19	0.15	-0.33	0.26	-0.33	0.27
0	-0.60	0.45	-0.23	0.13	-0.37	0.21	-0.36	0.22
0	-0.49	0.56	-0.13	0.19	-0.25	0.31	-0.24	0.32
0	-0.48	0.55	-0.11	0.15	-0.23	0.28	-0.23	0.29
1	0.74	1.28	0.77	1.31	0.76	1.30	0.76	1.30
-1.5	-1.76	-1.19	-1.81	-1.24	-1.80	-1.23	-1.80	-1.23
0	-0.14	0.12	-0.02	0.03	-0.10	0.09	-0.10	0.09
0	-0.12	0.12	-0.01	0.01	-0.09	0.09	-0.08	0.09
0	-0.14	0.11	-0.03	0.02	-0.11	0.08	-0.10	0.08

Table A.7: Estimated lower 0.05 quantile ($q_{0.05}$) and upper 0.95 quantile ($q_{0.95}$) for coefficients of variables in structural equation (5.6) for incorrectly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 100$ and $\nu = 4$ in prior for Σ.

	$\text{PIP}_\mathbf{a}$	$\text{PIP}_\mathbf{b}$	$\text{PIP}_\mathbf{c}$	$\text{PIP}_\mathbf{d}$	Mean*	$\text{Mean}_\mathbf{a}$	$\text{Mean}_\mathbf{b}$	Mean	$\text{Mean}_\mathbf{d}$
$\mathbf{y_2}$	1.00	1.00	1.00	1.00	0.76	0.76	0.76	0.75	0.75
$\mathbf{y_3}$	1.00	1.00	1.00	1.00	-0.45	-0.43	-0.43	-0.42	-0.42
$\mathbf{x_1}$	0.91	0.75	0.95	0.94	-0.18	-0.17	-0.14	-0.17	-0.16
$\mathbf{x_2}$	0.79	0.65	0.87	0.90	0.15	0.12	0.11	0.13	0.13
x_3	0.08	0.01	0.16	0.16	0.00	0.00	0.00	0.00	0.00
x_4	0.08	0.01	0.16	0.16	0.00	0.00	0.00	0.00	0.00
x_5	0.06	0.01	0.14	0.13	0.00	0.00	0.00	0.00	0.00
x_6	0.07	0.01	0.16	0.16	0.00	0.00	0.00	0.00	0.00
x_7	0.06	0.01	0.16	0.15	0.00	0.00	0.00	0.00	0.00
x_8	0.07	0.01	0.15	0.15	0.00	0.00	0.00	0.00	0.00
x_9	0.08	0.01	0.16	0.17	0.00	0.00	0.00	0.00	0.00
x_{10}	0.06	0.01	0.14	0.13	0.00	0.00	0.00	0.00	0.00
$\mathbf{x_{11}}$	1.00	1.00	1.00	1.00	0.34	0.35	0.35	0.34	0.34
$\mathbf{x_{12}}$	1.00	1.00	1.00	1.00	-0.51	-0.51	-0.51	-0.51	-0.50
x_{13}	0.09	0.01	0.18	0.17	0.00	0.00	0.00	0.00	0.00
x_{14}	0.07	0.01	0.15	0.14	0.00	0.00	0.00	0.00	0.00
x_{15}	0.08	0.01	0.16	0.15	0.00	0.00	0.00	0.00	0.00

Table A.8: Estimated posterior inclusion probabilities (PIP) and posterior means (Mean) for variables in structural equation (5.6) for correctly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 100$ and $\nu = 4$ in prior for Σ and **scaled** data.

Mean*	$q_{0.05,\mathbf{a}}$	$q_{0.95,\mathbf{a}}$	$q_{0.05,\mathbf{b}}$	$q_{0.95,\mathbf{b}}$	$q_{0.05,\mathbf{c}}$	$q_{0.95,\mathbf{c}}$	$q_{0.05,\mathbf{d}}$	$q_{0.95,\mathbf{d}}$
0.76	0.68	0.83	0.68	0.83	0.67	0.83	0.68	0.82
-0.45	-0.54	-0.33	-0.55	-0.33	-0.52	-0.32	-0.53	-0.32
-0.18	-0.24	-0.08	-0.22	-0.07	-0.24	-0.09	-0.23	-0.09
0.15	0.05	0.20	0.04	0.18	0.05	0.20	0.07	0.20
0.00	-0.01	0.01	0.00	0.00	-0.03	0.02	-0.03	0.02
0.00	-0.01	0.01	0.00	0.00	-0.02	0.02	-0.02	0.02
0.00	-0.01	0.00	0.00	0.00	-0.02	0.02	-0.02	0.02
0.00	-0.01	0.01	0.00	0.00	-0.02	0.03	-0.02	0.02
0.00	-0.01	0.01	0.00	0.00	-0.02	0.02	-0.02	0.02
0.00	-0.01	0.01	0.00	0.00	-0.02	0.02	-0.02	0.02
0.00	-0.01	0.01	0.00	0.00	-0.02	0.03	-0.02	0.02
0.00	-0.01	0.01	0.00	0.00	-0.02	0.02	-0.01	0.02
0.34	0.28	0.42	0.28	0.41	0.27	0.41	0.28	0.41
-0.51	-0.59	-0.43	-0.58	-0.44	-0.58	-0.43	-0.58	-0.43
0.00	-0.01	0.01	0.00	0.00	-0.03	0.02	-0.03	0.02
0.00	-0.01	0.01	0.00	0.00	-0.02	0.02	-0.02	0.02
0.00	-0.01	0.01	0.00	0.00	-0.02	0.02	-0.02	0.02

Table A.9: Estimated lower 0.05 quantile ($q_{0.05}$) and upper 0.95 quantile ($q_{0.95}$) for coefficients of variables in structural equation (5.6) for correctly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 100$ and $\nu = 4$ in prior for Σ and **scaled** data.

	$\text{PIP}_\mathbf{a}$	$\text{PIP}_\mathbf{b}$	$\text{PIP}_\mathbf{c}$	$\text{PIP}_\mathbf{d}$	Mean*	$\text{Mean}_\mathbf{a}$	$\text{Mean}_\mathbf{b}$	$\text{Mean}_\mathbf{c}$	$\text{Mean}_\mathbf{d}$
$\mathbf{y_2}$	1.00	1.00	1.00	1.00	0.76	0.69	0.69	0.68	0.68
$\mathbf{y_3}$	1.00	0.98	1.00	1.00	-0.45	-0.41	-0.40	-0.40	-0.40
$\mathbf{x_1}$	0.75	0.48	0.82	0.81	-0.18	-0.13	-0.09	-0.14	-0.14
$\mathbf{x_2}$	0.61	0.33	0.69	0.70	0.15	0.09	0.06	0.10	0.10
x_3	0.10	0.01	0.17	0.18	0.00	0.00	0.00	0.00	0.00
x_4	0.11	0.02	0.17	0.17	0.00	0.00	0.00	0.00	0.00
x_5	0.08	0.01	0.14	0.14	0.00	0.00	0.00	0.00	0.00
x_6	0.09	0.01	0.16	0.16	0.00	0.00	0.00	0.00	0.00
x_7	0.08	0.01	0.15	0.16	0.00	0.00	0.00	0.00	0.00
x_8	0.09	0.01	0.16	0.16	0.00	0.00	0.00	0.00	0.00
x_9	0.10	0.02	0.17	0.17	0.00	0.00	0.00	0.00	0.00
x_{10}	0.07	0.01	0.14	0.15	0.00	0.00	0.00	0.00	0.00
$\mathbf{x_{11}}$	1.00	0.99	1.00	1.00	0.34	0.32	0.32	0.32	0.32
$\mathbf{x_{12}}$	1.00	1.00	1.00	1.00	-0.51	-0.47	-0.47	-0.47	-0.47
x_{13}	0.11	0.01	0.18	0.18	0.00	0.00	0.00	0.00	0.00
x_{14}	0.09	0.01	0.15	0.15	0.00	0.00	0.00	0.00	0.00
x_{15}	0.10	0.02	0.16	0.16	0.00	0.00	0.00	0.00	0.00

Table A.10: Estimated posterior inclusion probabilities (PIP) and posterior means (Mean) for variables in structural equation (5.6) for incorrectly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 100$ and $\nu = 4$ in prior for Σ and **scaled** data.

Mean*	$q_{0.05,\mathbf{a}}$	$q_{0.95,\mathbf{a}}$	$q_{0.05,\mathbf{b}}$	$q_{0.95,\mathbf{b}}$	$q_{0.05,\mathbf{c}}$	$q_{0.95,\mathbf{c}}$	$q_{0.05,\mathbf{d}}$	$q_{0.95,\mathbf{d}}$
0.76	0.60	0.78	0.59	0.78	0.59	0.77	0.59	0.77
-0.45	-0.51	-0.30	-0.52	-0.29	-0.51	-0.29	-0.51	-0.29
-0.18	-0.23	-0.05	-0.19	-0.03	-0.23	-0.05	-0.23	-0.05
0.15	0.02	0.19	0.01	0.14	0.03	0.19	0.03	0.19
0.00	-0.01	0.02	0.00	0.00	-0.02	0.03	-0.03	0.03
0.00	-0.02	0.02	0.00	0.01	-0.03	0.03	-0.03	0.03
0.00	-0.01	0.01	0.00	0.00	-0.03	0.02	-0.03	0.02
0.00	-0.01	0.02	0.00	0.00	-0.02	0.03	-0.02	0.03
0.00	-0.01	0.01	0.00	0.00	-0.03	0.02	-0.03	0.02
0.00	-0.02	0.01	0.00	0.00	-0.03	0.02	-0.03	0.02
0.00	-0.01	0.02	0.00	0.00	-0.02	0.03	-0.02	0.03
0.00	-0.01	0.01	0.00	0.00	-0.02	0.03	-0.02	0.03
0.34	0.24	0.41	0.24	0.41	0.23	0.40	0.23	0.40
-0.51	-0.56	-0.38	-0.56	-0.38	-0.56	-0.38	-0.56	-0.38
0.00	-0.02	0.02	0.00	0.00	-0.03	0.03	-0.03	0.03
0.00	-0.01	0.02	0.00	0.00	-0.03	0.03	-0.03	0.03
0.00	-0.02	0.01	0.00	0.00	-0.03	0.02	-0.03	0.02

Table A.11: Estimated lower 0.05 quantile ($q_{0.05}$) and upper 0.95 quantile ($q_{0.95}$) for coefficients of variables in structural equation (5.6) for incorrectly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 100$ and $\nu = 4$ in prior for Σ and **scaled** data.

Tables Displaying the Median for Results from Section 5

	PIP_a	PIP_b	PIP_c	PIP_d	True Value	Med_a	Med_b	Med_c	Med_d
$\mathbf{y_2}$	1.00	1.00	1.00	1.00	1	0.99	1.00	0.84	0.99
$\mathbf{y_3}$	1.00	1.00	1.00	1.00	-1	-0.96	-0.92	-0.77	-0.96
$\mathbf{x_1}$	0.96	0.94	0.94	0.97	-1.8	-1.47	-1.75	-1.49	-1.72
$\mathbf{x_2}$	0.91	0.87	0.85	0.91	1.5	1.16	1.47	1.17	1.37
x_3	0.38	0.08	0.18	0.17	0	-0.01	0.11	-0.01	-0.02
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{10}	0.35	0.06	0.16	0.14	0	0.01	0.01	0.01	0.00
$\mathbf{x_{11}}$	1.00	1.00	1.00	1.00	1	1.02	1.03	0.90	1.02
$\mathbf{x_{12}}$	1.00	1.00	1.00	1.00	-1.5	-1.49	-1.55	-1.30	-1.50
x_{13}	0.19	0.03	0.19	0.19	0	0.00	-0.01	0.00	0.00
x_{14}	0.17	0.02	0.16	0.15	0	0.00	0.00	0.00	0.00
x_{15}	0.19	0.02	0.17	0.16	0	-0.01	-0.02	-0.01	-0.01

Table A.12: Estimated posterior inclusion probabilities (PIP) and posterior median (Med) for variables in structural equation (5.6) for correctly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 100$. $\nu = 4$ in prior for Σ.

	PIP_a	PIP_b	PIP_c	PIP_d	True Value	Med_a	Med_b	Med_c	Med_d
$\mathbf{y_2}$	1.00	1.00	1.00	1.00	1	0.98	0.99	0.98	0.98
$\mathbf{y_3}$	0.99	0.99	0.99	0.99	-1	-0.97	-0.99	-0.97	-0.97
$\mathbf{x_1}$	0.90	0.77	0.83	0.83	-1.8	-1.28	-1.48	-1.55	-1.55
$\mathbf{x_2}$	0.83	0.64	0.71	0.71	1.5	1.00	1.04	1.12	1.13
x_3	0.42	0.11	0.17	0.18	0	0.00	-0.01	-0.01	-0.02
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{10}	0.39	0.08	0.15	0.15	0	0.01	0.00	0.01	0.01
$\mathbf{x_{11}}$	0.99	0.99	0.98	0.99	1	1.01	1.04	1.03	1.03
$\mathbf{x_{12}}$	1.00	1.00	1.00	1.00	-1.5	-1.48	-1.53	-1.52	-1.52
x_{13}	0.24	0.04	0.18	0.18	0	0.00	0.00	0.00	0.00
x_{14}	0.22	0.03	0.15	0.15	0	0.00	0.00	0.00	0.00
x_{15}	0.23	0.04	0.17	0.17	0	0.00	0.00	0.00	0.00

Table A.13: Estimated posterior inclusion probabilities (PIP) and posterior median (Med) for variables in structural equation (5.6) for incorrectly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 100$ and $\nu = 4$ in prior for Σ.

	PIP_a	PIP_b	PIP_c	PIP_d	$Med_{N=5000}$	Med_a	Med_b	Med_c	Med_d
$\mathbf{y_2}$	1.00	1.00	1.00	1.00	0.76	0.76	0.76	0.75	0.75
$\mathbf{y_3}$	1.00	1.00	1.00	1.00	-0.45	-0.43	-0.43	-0.41	-0.42
$\mathbf{x_1}$	0.91	0.75	0.95	0.94	-0.18	-0.17	-0.14	-0.17	-0.16
$\mathbf{x_2}$	0.79	0.65	0.87	0.90	0.15	0.12	0.10	0.13	0.13
x_3	0.08	0.01	0.16	0.16	0.00	0.00	0.00	0.00	0.00
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_{10}	0.06	0.01	0.14	0.13	0.00	0.00	0.00	0.00	0.00
$\mathbf{x_{11}}$	1.00	1.00	1.00	1.00	0.34	0.35	0.35	0.34	0.34
$\mathbf{x_{12}}$	1.00	1.00	1.00	1.00	-0.51	-0.51	-0.51	-0.51	-0.50
x_{13}	0.09	0.01	0.18	0.17	0.00	0.00	0.00	0.00	0.00
x_{14}	0.07	0.01	0.15	0.14	0.00	0.00	0.00	0.00	0.00
x_{15}	0.08	0.01	0.16	0.15	0.00	0.00	0.00	0.00	0.00

Table A.14: Estimated posterior inclusion probabilities (PIP) and posterior median (Med) for variables in structural equation (5.6) for correctly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 100$. $\nu = 4$ in prior for Σ and **scaled** data.

	PIP_a	PIP_b	PIP_c	PIP_d	$Med_{N=5000}$	Med_a	Med_b	Med_c	Med_d
$\mathbf{y_2}$	1.00	1.00	1.00	1.00	0.70	0.69	0.69	0.68	0.68
$\mathbf{y_3}$	1.00	0.98	1.00	1.00	-0.42	-0.41	-0.41	-0.40	-0.40
$\mathbf{x_1}$	0.75	0.48	0.82	0.81	-0.16	-0.13	-0.09	-0.14	-0.14
$\mathbf{x_2}$	0.61	0.33	0.69	0.70	0.13	0.09	0.05	0.10	0.10
x_3	0.10	0.01	0.17	0.18	0.00	0.00	0.00	0.00	0.00
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_{10}	0.07	0.01	0.14	0.15	0.00	0.00	0.00	0.00	0.00
$\mathbf{x_{11}}$	1.00	0.99	1.00	1.00	0.31	0.32	0.32	0.32	0.31
$\mathbf{x_{12}}$	1.00	1.00	1.00	1.00	-0.47	-0.47	-0.47	-0.47	-0.47
x_{13}	0.11	0.01	0.18	0.18	0.00	0.00	0.00	0.00	0.00
x_{14}	0.09	0.01	0.15	0.15	0.00	0.00	0.00	0.00	0.00
x_{15}	0.10	0.02	0.16	0.16	0.00	0.00	0.00	0.00	0.00

Table A.15: Estimated posterior inclusion probabilities (PIP) and posterior median (Med) for variables in structural equation (5.6) for an incorrectly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 100$ and $\nu = 4$ in prior for Σ and **scaled** data.

Results for Approach from Section 5 with $\nu = 4$ and $N = 500$

	PIP_a	PIP_b	PIP_c	PIP_d	True Value	$Mean_a$	$Mean_b$	$Mean_c$	$Mean_d$
$\mathbf{y_2}$	1.00	1.00	1.00	1.00	1	1.00	1.00	1.00	1.00
$\mathbf{y_3}$	1.00	1.00	1.00	1.00	-1	-1.00	-1.01	-0.99	-1.00
$\mathbf{x_1}$	1.00	1.00	1.00	1.00	-1.8	-1.73	-1.79	-1.79	-1.79
$\mathbf{x_2}$	1.00	1.00	1.00	1.00	1.5	1.43	1.50	1.48	1.49
x_3	0.25	0.02	0.10	0.09	0	0.01	0.00	0.01	0.01
x_4	0.25	0.02	0.10	0.10	0	0.00	0.00	0.00	0.00
x_5	0.26	0.02	0.10	0.09	0	-0.01	0.00	-0.01	0.00
x_6	0.24	0.02	0.10	0.10	0	0.00	0.00	0.00	0.00
x_7	0.23	0.02	0.09	0.09	0	0.00	0.00	0.00	0.00
x_8	0.23	0.02	0.09	0.09	0	0.00	0.00	0.00	0.00
x_9	0.23	0.01	0.09	0.08	0	0.01	0.00	0.01	0.01
x_{10}	0.24	0.02	0.10	0.09	0	-0.01	0.00	0.00	0.00
$\mathbf{x_{11}}$	1.00	1.00	1.00	1.00	1	0.99	0.99	0.99	0.99
$\mathbf{x_{12}}$	1.00	1.00	1.00	1.00	-1.5	-1.50	-1.50	-1.50	-1.50
x_{13}	0.09	0.01	0.09	0.09	0	0.00	0.00	0.00	0.00
x_{14}	0.09	0.01	0.09	0.09	0	0.00	0.00	0.00	0.00
x_{15}	0.08	0.00	0.07	0.08	0	0.00	0.00	0.00	0.00

Table A.16: Estimated posterior inclusion probabilities (PIP) and posterior mean (Mean) for variables in structural equation (5.6) for correctly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 500$ and $\nu = 4$ in prior for Σ.

True Value	$q_{0.05,\mathbf{a}}$	$q_{0.95,\mathbf{a}}$	$q_{0.05,\mathbf{b}}$	$q_{0.95,\mathbf{b}}$	$q_{0.05,\mathbf{c}}$	$q_{0.95,\mathbf{c}}$	$q_{0.05,\mathbf{d}}$	$q_{0.95,\mathbf{d}}$
1	0.95	1.04	0.96	1.04	0.95	1.04	0.96	1.04
-1	-1.11	-0.89	-1.12	-0.90	-1.10	-0.89	-1.11	-0.89
-1.8	-2.03	-1.43	-2.06	-1.52	-2.09	-1.49	-2.06	-1.52
1.5	1.12	1.73	1.23	1.76	1.17	1.78	1.23	1.76
0	-0.16	0.16	-0.01	0.01	-0.07	0.08	-0.05	0.06
0	-0.17	0.16	-0.01	0.02	-0.07	0.06	-0.06	0.07
0	-0.18	0.15	-0.01	0.01	-0.08	0.05	-0.06	0.05
0	-0.14	0.13	-0.01	0.01	-0.06	0.05	-0.06	0.04
0	-0.13	0.15	-0.02	0.01	-0.05	0.06	-0.04	0.05
0	-0.10	0.16	0.00	0.00	-0.04	0.06	-0.04	0.06
0	-0.11	0.15	-0.01	0.01	-0.04	0.06	-0.03	0.06
0	-0.17	0.12	-0.01	0.01	-0.08	0.05	-0.06	0.05
1	0.90	1.08	0.92	1.07	0.90	1.08	0.91	1.07
-1.5	-1.59	-1.41	-1.59	-1.42	-1.59	-1.41	-1.59	-1.41
0	-0.01	0.02	0.00	0.00	-0.01	0.02	-0.01	0.02
0	-0.02	0.02	0.00	0.00	-0.02	0.01	-0.02	0.01
0	-0.01	0.01	0.00	0.00	-0.01	0.01	-0.01	0.01

Table A.17: Estimated lower 0.05 quantile ($q_{0.05}$) and upper 0.95 quantile ($q_{0.95}$) for coefficients of variables in structural equation (5.6) for correctly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 500$ and $\nu = 4$ in prior for Σ.

$\hat{\Sigma}_a$	$\hat{\Sigma}_b$	$\hat{\Sigma}_c$	$\hat{\Sigma}_d$
$\begin{pmatrix} 24.72 & 0.94 & 10.38 \\ 0.94 & 3.94 & 1.72 \\ 10.38 & 1.72 & 8.88 \end{pmatrix}$	$\begin{pmatrix} 24.93 & 0.96 & 10.47 \\ 0.96 & 3.97 & 1.75 \\ 10.47 & 1.75 & 8.93 \end{pmatrix}$	$\begin{pmatrix} 24.65 & 0.94 & 10.35 \\ 0.94 & 3.94 & 1.74 \\ 10.35 & 1.74 & 8.89 \end{pmatrix}$	$\begin{pmatrix} 24.75 & 0.95 & 10.41 \\ 0.95 & 3.97 & 1.75 \\ 10.41 & 1.75 & 8.93 \end{pmatrix}$

Table A.18: Estimated posterior means of covariance matrices in structural equation (5.6) for correctly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 500$ and $\nu = 4$ in prior for Σ.

	$\text{PIP}_\mathbf{a}$	$\text{PIP}_\mathbf{b}$	$\text{PIP}_\mathbf{c}$	$\text{PIP}_\mathbf{d}$	True Value	$\text{Mean}_\mathbf{a}$	$\text{Mean}_\mathbf{b}$	$\text{Mean}_\mathbf{c}$	$\text{Mean}_\mathbf{d}$
$\mathbf{y_2}$	1.00	1.00	1.00	1.00	1	1.00	1.00	1.00	1.00
$\mathbf{y_3}$	1.00	1.00	1.00	1.00	-1	-1.00	-1.00	-1.00	-1.00
$\mathbf{x_1}$	1.00	1.00	1.00	1.00	-1.8	-1.71	-1.79	-1.79	-1.79
$\mathbf{x_2}$	1.00	1.00	1.00	1.00	1.5	1.40	1.47	1.47	1.47
x_3	0.30	0.03	0.10	0.10	0	0.00	0.01	0.01	0.01
x_4	0.29	0.02	0.08	0.08	0	-0.01	0.00	0.00	0.00
x_5	0.30	0.03	0.09	0.09	0	-0.02	0.00	-0.01	-0.01
x_6	0.28	0.03	0.09	0.09	0	-0.01	0.00	0.00	0.00
x_7	0.28	0.02	0.09	0.09	0	0.00	0.00	0.00	0.00
x_8	0.27	0.02	0.09	0.09	0	0.01	0.00	0.00	0.00
x_9	0.28	0.02	0.09	0.09	0	0.01	0.00	0.01	0.01
x_{10}	0.29	0.03	0.10	0.10	0	-0.02	0.00	-0.01	-0.01
$\mathbf{x_{11}}$	1.00	1.00	1.00	1.00	1	0.99	0.99	0.99	0.99
$\mathbf{x_{12}}$	1.00	1.00	1.00	1.00	-1.5	-1.49	-1.51	-1.50	-1.50
x_{13}	0.12	0.01	0.09	0.09	0	0.00	0.00	0.00	0.00
x_{14}	0.12	0.01	0.09	0.09	0	0.00	0.00	0.00	0.00
x_{15}	0.11	0.01	0.08	0.07	0	0.00	0.00	0.00	0.00

Table A.19: Estimated posterior inclusion probabilities (PIP) and posterior mean (Mean) for variables in structural equation (5.6) for incorrectly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 500$ and $\nu = 4$ in prior for Σ.

True Value	$q_{0.05,\mathbf{a}}$	$q_{0.95,\mathbf{a}}$	$q_{0.05,\mathbf{b}}$	$q_{0.95,\mathbf{b}}$	$q_{0.05,\mathbf{c}}$	$q_{0.95,\mathbf{c}}$	$q_{0.05,\mathbf{d}}$	$q_{0.95,\mathbf{d}}$
1	0.94	1.05	0.94	1.05	0.94	1.05	0.94	1.05
-1	-1.11	-0.89	-1.11	-0.89	-1.11	-0.89	-1.11	-0.89
-1.8	-2.07	-1.34	-2.17	-1.42	-2.17	-1.41	-2.16	-1.41
1.5	1.04	1.77	1.09	1.84	1.09	1.84	1.09	1.84
0	-0.22	0.22	-0.02	0.04	-0.08	0.08	-0.08	0.09
0	-0.22	0.23	-0.02	0.01	-0.08	0.07	-0.08	0.07
0	-0.26	0.21	-0.03	0.01	-0.10	0.06	-0.10	0.06
0	-0.21	0.19	-0.03	0.02	-0.08	0.06	-0.08	0.06
0	-0.18	0.21	-0.02	0.02	-0.06	0.07	-0.06	0.07
0	-0.14	0.24	-0.02	0.02	-0.05	0.07	-0.05	0.08
0	-0.17	0.22	-0.01	0.03	-0.04	0.08	-0.04	0.08
0	-0.24	0.17	-0.03	0.01	-0.10	0.06	-0.10	0.06
1	0.87	1.11	0.88	1.11	0.88	1.11	0.88	1.11
-1.5	-1.62	-1.37	-1.63	-1.39	-1.62	-1.38	-1.62	-1.38
0	-0.03	0.04	0.00	0.00	-0.02	0.02	-0.02	0.02
0	-0.03	0.02	0.00	0.00	-0.03	0.01	-0.03	0.01
0	-0.03	0.02	0.00	0.00	-0.02	0.01	-0.02	0.01

Table A.20: Estimated lower 0.05 quantile ($q_{0.05}$) and upper 0.95 quantile ($q_{0.95}$) for coefficients of variables in structural equation (5.6) for incorrectly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 500$ and $\nu = 4$ in prior for Σ.

$\hat{\Sigma}_a$	$\hat{\Sigma}_b$	$\hat{\Sigma}_c$	$\hat{\Sigma}_d$
$\begin{pmatrix} 24.68 & -0.03 & 0.01 \\ -0.03 & 3.95 & 1.76 \\ 0.01 & 1.76 & 8.87 \end{pmatrix}$	$\begin{pmatrix} 24.77 & -0.04 & 0.03 \\ -0.04 & 3.98 & 1.77 \\ 0.03 & 1.77 & 8.93 \end{pmatrix}$	$\begin{pmatrix} 24.72 & -0.04 & 0.00 \\ -0.04 & 3.95 & 1.76 \\ 0.00 & 1.76 & 8.87 \end{pmatrix}$	$\begin{pmatrix} 24.72 & -0.03 & 0.01 \\ -0.03 & 3.98 & 1.77 \\ 0.01 & 1.77 & 8.93 \end{pmatrix}$

Table A.21: Estimated posterior means of covariance matrices in structural equation (5.6) for incorrectly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 500$ and $\nu = 4$ in prior for Σ.

	PIP_a	PIP_b	PIP_c	PIP_d	Mean*	$Mean_a$	$Mean_b$	$Mean_c$	$Mean_d$
$\mathbf{y_2}$	1.00	1.00	1.00	1.00	0.76	0.76	0.76	0.76	0.76
$\mathbf{y_3}$	1.00	1.00	1.00	1.00	-0.45	-0.45	-0.45	-0.45	-0.45
$\mathbf{x_1}$	1.00	1.00	1.00	1.00	-0.18	-0.18	-0.18	-0.18	-0.18
$\mathbf{x_2}$	1.00	1.00	1.00	1.00	0.15	0.15	0.15	0.15	0.15
x_3	0.04	0.00	0.09	0.09	0.00	0.00	0.00	0.00	0.00
x_4	0.04	0.00	0.09	0.09	0.00	0.00	0.00	0.00	0.00
x_5	0.04	0.00	0.09	0.09	0.00	0.00	0.00	0.00	0.00
x_6	0.04	0.00	0.09	0.09	0.00	0.00	0.00	0.00	0.00
x_7	0.04	0.00	0.09	0.09	0.00	0.00	0.00	0.00	0.00
x_8	0.04	0.00	0.09	0.09	0.00	0.00	0.00	0.00	0.00
x_9	0.03	0.00	0.08	0.08	0.00	0.00	0.00	0.00	0.00
x_{10}	0.04	0.00	0.09	0.09	0.00	0.00	0.00	0.00	0.00
$\mathbf{x_{11}}$	1.00	1.00	1.00	1.00	0.34	0.34	0.34	0.34	0.34
$\mathbf{x_{12}}$	1.00	1.00	1.00	1.00	-0.51	-0.51	-0.51	-0.51	-0.51
x_{13}	0.04	0.00	0.08	0.09	0.00	0.00	0.00	0.00	0.00
x_{14}	0.04	0.00	0.09	0.09	0.00	0.00	0.00	0.00	0.00
x_{15}	0.03	0.00	0.07	0.08	0.00	0.00	0.00	0.00	0.00

Table A.22: Estimated posterior inclusion probabilities (PIP) and posterior means (Mean) for variables in structural equation (5.6) for correctly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 500$ and $\nu = 4$ in prior for Σ and **scaled** data.

Mean*	$q_{0.05,a}$	$q_{0.95,a}$	$q_{0.05,b}$	$q_{0.95,b}$	$q_{0.05,c}$	$q_{0.95,c}$	$q_{0.05,d}$	$q_{0.95,d}$
0.76	0.73	0.79	0.73	0.79	0.73	0.79	0.73	0.79
-0.45	-0.50	-0.40	-0.50	-0.40	-0.50	-0.40	-0.50	-0.40
-0.18	-0.21	-0.15	-0.20	-0.15	-0.21	-0.15	-0.20	-0.15
0.15	0.12	0.18	0.12	0.18	0.12	0.18	0.12	0.17
0.00	0.00	0.00	0.00	0.00	-0.01	0.01	-0.01	0.01
0.00	0.00	0.00	0.00	0.00	-0.01	0.01	-0.01	0.01
0.00	0.00	0.00	0.00	0.00	-0.01	0.00	-0.01	0.01
0.00	0.00	0.00	0.00	0.00	-0.01	0.00	-0.01	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01
0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01
0.00	0.00	0.00	0.00	0.00	-0.01	0.01	-0.01	0.01
0.34	0.31	0.37	0.31	0.36	0.31	0.37	0.31	0.36
-0.51	-0.54	-0.48	-0.54	-0.48	-0.54	-0.48	-0.54	-0.48
0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01
0.00	0.00	0.00	0.00	0.00	-0.01	0.00	-0.01	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table A.23: Estimated lower 0.05 quantile ($q_{0.05}$) and upper 0.95 quantile ($q_{0.95}$) for coefficients of variables in structural equation (5.6) for correctly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 500$ and $\nu = 4$ in prior for Σ and **scaled** data.

$\hat{\Sigma}_a$	$\hat{\Sigma}_b$	$\hat{\Sigma}_c$	$\hat{\Sigma}_d$
$\begin{pmatrix} 0.24 & 0.01 & 0.22 \\ 0.01 & 0.07 & 0.05 \\ 0.22 & 0.05 & 0.43 \end{pmatrix}$	$\begin{pmatrix} 0.24 & 0.01 & 0.22 \\ 0.01 & 0.07 & 0.05 \\ 0.22 & 0.05 & 0.43 \end{pmatrix}$	$\begin{pmatrix} 0.24 & 0.01 & 0.22 \\ 0.01 & 0.07 & 0.05 \\ 0.22 & 0.05 & 0.43 \end{pmatrix}$	$\begin{pmatrix} 0.24 & 0.01 & 0.22 \\ 0.01 & 0.07 & 0.05 \\ 0.22 & 0.05 & 0.43 \end{pmatrix}$

Table A.24: Estimated posterior means of covariance matrices in structural equation (5.6) for correctly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 500$ and $\nu = 4$ in prior for Σ and **scaled** data.

	$\text{PIP}_\mathbf{a}$	$\text{PIP}_\mathbf{b}$	$\text{PIP}_\mathbf{c}$	$\text{PIP}_\mathbf{d}$	Mean*	$\text{Mean}_\mathbf{a}$	$\text{Mean}_\mathbf{b}$	$\text{Mean}_\mathbf{c}$	$\text{Mean}_\mathbf{d}$
$\mathbf{y_2}$	1.00	1.00	1.00	1.00	0.76	0.70	0.70	0.70	0.70
$\mathbf{y_3}$	1.00	1.00	1.00	1.00	-0.45	-0.42	-0.42	-0.42	-0.41
$\mathbf{x_1}$	1.00	1.00	1.00	1.00	-0.18	-0.16	-0.16	-0.16	-0.16
$\mathbf{x_2}$	1.00	0.99	1.00	1.00	0.15	0.13	0.13	0.13	0.13
x_3	0.05	0.00	0.05	0.10	0.00	0.00	0.00	0.00	0.00
x_4	0.04	0.00	0.04	0.08	0.00	0.00	0.00	0.00	0.00
x_5	0.05	0.00	0.05	0.09	0.00	0.00	0.00	0.00	0.00
x_6	0.05	0.00	0.05	0.09	0.00	0.00	0.00	0.00	0.00
x_7	0.04	0.00	0.04	0.09	0.00	0.00	0.00	0.00	0.00
x_8	0.05	0.00	0.05	0.09	0.00	0.00	0.00	0.00	0.00
x_9	0.04	0.00	0.04	0.09	0.00	0.00	0.00	0.00	0.00
x_{10}	0.05	0.00	0.05	0.10	0.00	0.00	0.00	0.00	0.00
$\mathbf{x_{11}}$	1.00	1.00	1.00	1.00	0.34	0.31	0.31	0.31	0.31
$\mathbf{x_{12}}$	1.00	1.00	1.00	1.00	-0.51	-0.47	-0.47	-0.47	-0.47
x_{13}	0.05	0.00	0.05	0.09	0.00	0.00	0.00	0.00	0.00
x_{14}	0.05	0.00	0.05	0.09	0.00	0.00	0.00	0.00	0.00
x_{15}	0.04	0.00	0.04	0.08	0.00	0.00	0.00	0.00	0.00

Table A.25: Estimated posterior inclusion probabilities (PIP) and posterior means (Mean) for variables in structural equation (5.6) for incorrectly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 500$ and $\nu = 4$ in prior for Σ and **scaled** data.

Mean*	$q_{0.05,\mathbf{a}}$	$q_{0.95,\mathbf{a}}$	$q_{0.05,\mathbf{b}}$	$q_{0.95,\mathbf{b}}$	$q_{0.05,\mathbf{c}}$	$q_{0.95,\mathbf{c}}$	$q_{0.05,\mathbf{d}}$	$q_{0.95,\mathbf{d}}$
0.76	0.66	0.74	0.66	0.74	0.66	0.74	0.66	0.74
-0.45	-0.46	-0.37	-0.46	-0.37	-0.46	-0.37	-0.46	-0.37
-0.18	-0.20	-0.13	-0.20	-0.13	-0.20	-0.13	-0.20	-0.13
0.15	0.10	0.17	0.10	0.17	0.10	0.17	0.10	0.17
0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	0.01
0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	0.01
0.00	-0.01	0.00	0.00	0.00	-0.01	0.00	-0.01	0.01
0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	0.01
0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	0.01
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	0.01
0.34	0.27	0.35	0.27	0.34	0.27	0.35	0.27	0.35
-0.51	-0.51	-0.43	-0.51	-0.43	-0.51	-0.43	-0.51	-0.43
0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	0.01
0.00	-0.01	0.00	0.00	0.00	-0.01	0.00	-0.01	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	0.00

Table A.26: Estimated lower 0.05 quantile ($q_{0.05}$) and upper 0.95 quantile ($q_{0.95}$) for coefficients of variables in structural equation (5.6) for incorrectly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 500$ and $\nu = 4$ in prior for Σ and **scaled** data.

$\hat{\Sigma}_a$	$\hat{\Sigma}_b$	$\hat{\Sigma}_c$	$\hat{\Sigma}_d$
$\begin{pmatrix} 0.21 & 0.00 & 0.00 \\ 0.00 & 0.07 & 0.05 \\ 0.00 & 0.05 & 0.43 \end{pmatrix}$	$\begin{pmatrix} 0.21 & 0.00 & 0.00 \\ 0.00 & 0.07 & 0.05 \\ 0.00 & 0.05 & 0.43 \end{pmatrix}$	$\begin{pmatrix} 0.21 & 0.00 & 0.00 \\ 0.00 & 0.07 & 0.05 \\ 0.00 & 0.05 & 0.43 \end{pmatrix}$	$\begin{pmatrix} 0.21 & 0.00 & 0.00 \\ 0.00 & 0.07 & 0.05 \\ 0.00 & 0.05 & 0.43 \end{pmatrix}$

Table A.27: Estimated posterior means of covariance matrices in structural equation (5.6) for incorrectly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 500$ and $\nu = 4$ in prior for Σ and **scaled** data.

Results for Approach from Section 5 with $\nu = 8$ and $N = 100$

	PIP_a	PIP_b	PIP_c	PIP_d	True Value	Mean_a	Mean_b	Mean_c	Mean_d
$\mathbf{y_2}$	1.00	1.00	1.00	1.00	1	0.99	1.00	0.99	0.99
$\mathbf{y_3}$	1.00	1.00	1.00	1.00	-1	-0.97	-1.00	-0.95	-0.97
$\mathbf{x_1}$	0.96	0.95	0.95	0.97	-1.8	-1.46	-1.72	-1.71	-1.73
$\mathbf{x_2}$	0.91	0.87	0.86	0.92	1.5	1.16	1.34	1.31	1.38
x_3	0.38	0.08	0.18	0.18	0	0.00	-0.02	-0.01	-0.03
x_4	0.38	0.08	0.18	0.17	0	0.01	0.00	0.01	0.01
x_5	0.36	0.06	0.15	0.14	0	-0.01	0.00	0.00	0.00
x_6	0.37	0.07	0.18	0.18	0	0.02	0.01	0.01	0.02
x_7	0.35	0.06	0.17	0.16	0	-0.02	-0.01	-0.01	-0.01
x_8	0.36	0.07	0.17	0.16	0	-0.03	-0.01	-0.02	-0.01
x_9	0.37	0.08	0.18	0.18	0	0.02	0.00	0.01	0.01
x_{10}	0.35	0.06	0.16	0.15	0	0.02	0.00	0.01	0.01
$\mathbf{x_{11}}$	1.00	1.00	0.99	1.00	1	1.01	1.03	1.02	1.02
$\mathbf{x_{12}}$	1.00	1.00	1.00	1.00	-1.5	-1.49	-1.52	-1.51	-1.50
x_{13}	0.19	0.03	0.19	0.19	0	0.00	0.00	0.00	0.00
x_{14}	0.17	0.02	0.16	0.15	0	0.00	0.00	0.00	0.00
x_{15}	0.19	0.03	0.18	0.17	0	0.00	0.00	0.00	0.00

Table A.28: Estimated posterior inclusion probabilities (PIP) and posterior mean (Mean) for variables in structural equation (5.6) for correctly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 100$ and $\nu = 8$ in prior for Σ.

True Value	$q_{0.05,a}$	$q_{0.95,a}$	$q_{0.05,b}$	$q_{0.95,b}$	$q_{0.05,c}$	$q_{0.95,c}$	$q_{0.05,d}$	$q_{0.95,d}$
1	0.89	1.09	0.91	1.08	0.88	1.09	0.90	1.08
-1	-1.22	-0.75	-1.26	-0.77	-1.19	-0.74	-1.20	-0.75
-1.8	-2.14	-0.72	-2.37	-1.02	-2.47	-0.90	-2.37	-1.07
1.5	0.44	1.85	0.68	2.01	0.52	2.09	0.74	2.02
0	-0.45	0.45	-0.15	0.06	-0.27	0.27	-0.29	0.17
0	-0.44	0.46	-0.08	0.11	-0.25	0.27	-0.18	0.24
0	-0.47	0.43	-0.06	0.06	-0.24	0.23	-0.19	0.19
0	-0.39	0.44	-0.08	0.11	-0.22	0.29	-0.18	0.24
0	-0.43	0.37	-0.11	0.05	-0.28	0.22	-0.22	0.17
0	-0.46	0.35	-0.11	0.08	-0.29	0.19	- 0.24	0.17
0	-0.39	0.43	-0.07	0.10	-0.22	0.26	-0.18	0.25
0	-0.37	0.43	-0.06	0.08	-0.20	0.25	-0.15	0.21
1	0.81	1.22	0.84	1.21	0.82	1.22	0.83	1.20
-1.5	-1.72	-1.27	-1.72	-1.32	-1.74	-1.29	-1.70	-1.31
0	-0.09	0.07	-0.01	0.01	-0.08	0.07	-0.08	0.06
0	-0.07	0.07	-0.01	0.01	-0.06	0.07	-0.06	0.06
0	-0.08	0.07	-0.01	0.00	-0.08	0.07	-0.07	0.06

Table A.29: Estimated lower 0.05 quantile ($q_{0.05}$) and upper 0.95 quantile ($q_{0.95}$) for coefficients of variables in structural equation (5.6) for correctly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 100$ and $\nu = 8$ in prior for Σ.

$\hat{\Sigma}_a$	$\hat{\Sigma}_b$	$\hat{\Sigma}_c$	$\hat{\Sigma}_d$
$\begin{pmatrix} 23.09 & 0.81 & 9.43 \\ 0.81 & 3.67 & 1.53 \\ 9.43 & 1.53 & 8.00 \end{pmatrix}$	$\begin{pmatrix} 24.22 & 0.82 & 9.93 \\ 0.82 & 4.03 & 1.62 \\ 9.93 & 1.62 & 8.32 \end{pmatrix}$	$\begin{pmatrix} 22.73 & 0.80 & 9.28 \\ 0.80 & 3.67 & 1.54 \\ 9.28 & 1.54 & 8.00 \end{pmatrix}$	$\begin{pmatrix} 23.18 & 0.81 & 9.59 \\ 0.81 & 4.03 & 1.62 \\ 9.59 & 1.62 & 8.32 \end{pmatrix}$

Table A.30: Estimated posterior means of covariance matrices in structural equation (5.6) for correctly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 100$ and $\nu = 8$ in prior for Σ.

	PIP_a	PIP_b	PIP_c	PIP_d	True Value	Mean_a	Mean_b	Mean_c	Mean_d
$\mathbf{y_2}$	1.00	1.00	1.00	1.00	1	0.98	0.99	0.98	0.98
$\mathbf{y_3}$	1.00	0.99	0.99	0.99	-1	-0.97	-0.99	-0.97	-0.97
$\mathbf{x_1}$	0.91	0.79	0.84	0.84	-1.8	-1.30	-1.51	-1.57	-1.56
$\mathbf{x_2}$	0.83	0.66	0.72	0.73	1.5	1.03	1.11	1.18	1.18
x_3	0.42	0.11	0.18	0.18	0	0.01	0.00	0.01	0.01
x_4	0.42	0.12	0.18	0.18	0	0.02	0.02	0.03	0.03
x_5	0.40	0.08	0.14	0.14	0	-0.02	-0.01	-0.01	-0.01
x_6	0.41	0.10	0.17	0.17	0	0.02	0.01	0.02	0.02
x_7	0.40	0.09	0.16	0.16	0	-0.02	-0.01	-0.02	-0.02
x_8	0.41	0.10	0.16	0.17	0	-0.04	-0.02	-0.03	-0.03
x_9	0.41	0.10	0.17	0.17	0	0.02	0.00	0.00	0.01
x_{10}	0.39	0.08	0.15	0.15	0	0.02	0.01	0.01	0.01
$\mathbf{x_{11}}$	1.00	0.99	0.99	0.99	1	1.01	1.04	1.03	1.03
$\mathbf{x_{12}}$	1.00	1.00	1.00	1.00	-1.5	-1.48	-1.53	-1.52	-1.52
x_{13}	0.24	0.04	0.18	0.19	0	0.00	0.00	0.00	0.00
x_{14}	0.21	0.03	0.16	0.16	0	0.00	0.00	0.00	0.00
x_{15}	0.23	0.04	0.17	0.18	0	-0.01	0.00	-0.01	-0.01

Table A.31: Estimated posterior inclusion probabilities (PIP) and posterior mean (Mean) for variables in structural equation (5.6) for incorrectly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 100$ and $\nu = 8$ in prior for Σ.

True Value	$q_{0.05,\mathbf{a}}$	$q_{0.95,\mathbf{a}}$	$q_{0.05,\mathbf{b}}$	$q_{0.95,\mathbf{b}}$	$q_{0.05,\mathbf{c}}$	$q_{0.95,\mathbf{c}}$	$q_{0.05,\mathbf{d}}$	$q_{0.95,\mathbf{d}}$
1	0.85	1.10	0.87	1.12	0.86	1.11	0.86	1.11
-1	-1.22	-0.72	-1.24	-0.73	-1.22	-0.72	-1.22	-0.72
-1.8	-2.12	-0.48	-2.52	-0.62	-2.55	-0.65	-2.54	-0.64
1.5	0.24	1.85	0.29	2.08	0.35	2.14	0.34	2.14
0	-0.52	0.59	-0.17	0.23	-0.27	0.36	-0.28	0.38
0	-0.56	0.58	-0.19	0.21	-0.31	0.33	-0.31	0.32
0	-0.61	0.54	-0.16	0.13	-0.30	0.27	-0.30	0.27
0	-0.50	0.54	-0.15	0.21	-0.26	0.32	-0.26	0.33
0	-0.56	0.48	-0.19	0.15	-0.33	0.26	-0.33	0.27
0	-0.60	0.44	-0.23	0.14	-0.37	0.21	-0.35	0.22
0	-0.48	0.56	-0.13	0.19	-0.25	0.32	-0.24	0.33
0	-0.47	0.54	-0.11	0.15	-0.23	0.28	-0.23	0.30
1	0.74	1.27	0.77	1.30	0.76	1.30	0.76	1.29
-1.5	-1.76	-1.20	-1.81	-1.25	-1.80	-1.23	-1.80	-1.23
0	-0.13	0.12	-0.02	0.02	-0.10	0.09	-0.11	0.09
0	-0.12	0.12	-0.01	0.01	-0.09	0.09	-0.09	0.09
0	-0.14	0.11	-0.03	0.02	-0.10	0.08	-0.11	0.08

Table A.32: Estimated lower 0.05 quantile ($q_{0.05}$) and upper 0.95 quantile ($q_{0.95}$) for coefficients of variables in structural equation (5.6) for incorrectly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 100$ and $\nu = 8$ in prior for Σ.

$\hat{\Sigma}_a$	$\hat{\Sigma}_b$	$\hat{\Sigma}_c$	$\hat{\Sigma}_d$
$\begin{pmatrix} 23.79 & -0.01 & -0.13 \\ -0.01 & 3.68 & 1.61 \\ -0.13 & 1.61 & 7.85 \end{pmatrix}$	$\begin{pmatrix} 24.33 & -0.05 & 0.00 \\ -0.05 & 4.05 & 1.67 \\ 0.00 & 1.67 & 8.23 \end{pmatrix}$	$\begin{pmatrix} 23.93 & -0.03 & -0.15 \\ -0.03 & 3.68 & 1.61 \\ -0.15 & 1.61 & 7.85 \end{pmatrix}$	$\begin{pmatrix} 23.95 & -0.04 & -0.12 \\ -0.04 & 4.05 & 1.67 \\ -0.12 & 1.67 & 8.23 \end{pmatrix}$

Table A.33: Estimated posterior means of covariance matrices in structural equation (5.6) for incorrectly specified system and different prior settings (a-d). Based on an average of 100 samples of size $N = 100$ and $\nu = 8$ in prior for Σ.

D Additional Results: SD-Density Ratio for Endogeneity

Additional Simulation Results for a Two-Equation System

Based on the simulation setup discussed in subsection 4.2, we design the two-equation system

$$y_1 = y_2 - 1.8x_1 + 1.5x_2 + x_{11} - 1.5x_{12} + \epsilon_1$$
$$y_2 = 1.1x_1 - 0.5x_3 + 0.75x_{12} + 0.75z_1 - 2z_8 + \epsilon_2.$$

The error terms ϵ_1 and ϵ_2 are drawn from a multivariate normal distribution with

$$\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 25 & \sigma_{12} \\ \sigma_{12} & 9 \end{pmatrix}\right).$$

The corresponding correlation matrix is then given by

$$\begin{pmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{pmatrix}.$$

All the following results are obtained by using the standard IW-prior with $IW \sim (I_2, \nu)$.

	ρ_{12}							
ν	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
2	17.46	14.72	7.91	2.71	0.67	0.09	∼0	∼0
3	17.86	15.07	8.08	2.75	0.68	0.09	∼0	∼0
4	18.34	15.40	8.25	2.80	0.69	0.09	∼0	∼0
8	19.95	16.81	8.92	2.98	0.73	0.09	∼0	∼0

Table A.34: Density value at zero for the marginal posterior density of σ_{12} for a two-equation system with $N = 100$ observations and I-I prior.

	ρ_{12}							
ν	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
2	35.63	30.04	16.14	5.53	1.37	0.18	∼0	∼0
3	14.06	11.87	6.36	2.17	0.54	0.07	∼0	∼0
4	8.15	6.84	3.67	1.24	0.31	0.04	∼0	∼0
8	2.61	2.20	1.17	0.39	0.09	0.01	∼0	∼0

Table A.35: SD-ratio value for a two-equation system with $N = 100$ observations and I-I prior.

	ρ_{12}							
ν	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
2	17.11	14.95	8.29	2.81	0.73	0.14	∼0	∼0
3	17.30	15.11	8.39	2.84	0.74	0.14	∼0	∼0
4	17.70	15.46	8.56	2.88	0.75	0.14	∼0	∼0
8	19.33	16.86	9.25	3.04	0.79	0.15	∼0	∼0

Table A.36: Density value at zero for the marginal posterior density of σ_{12} for a two-equation system with $N = 100$ observations and g-g prior.

	ρ_{12}							
ν	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
2	34.92	30.51	16.92	5.73	1.49	0.29	∼0	∼0
3	13.62	11.90	6.61	2.24	0.58	0.11	∼0	∼0
4	7.87	6.87	3.80	1.28	0.33	0.06	∼0	∼0
8	2.53	2.20	1.21	0.40	0.10	0.02	∼0	∼0

Table A.37: SD-ratio value for the marginal posterior density of σ_{12} for a two-equation system with $N = 100$ observations and g-g prior.

	ρ_{12}							
ν	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
2	17.12	14.31	7.63	2.54	0.68	0.13	∼0	∼0
3	17.53	14.63	7.79	2.58	0.69	0.13	∼0	∼0
4	17.93	14.96	7.94	2.61	0.70	0.13	∼0	∼0
8	19.59	16.30	8.56	2.76	0.74	0.14	∼0	∼0

Table A.38: Density value at zero for the marginal posterior density of σ_{12} for a two-equation system with $N = 100$ observations and gprvar-g prior.

	ρ_{12}							
ν	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
2	34.94	29.20	15.57	5.18	1.39	0.26	∼0	∼0
3	13.80	11.52	6.13	2.03	0.54	0.10	∼0	∼0
4	7.97	6.65	3.53	1.16	0.31	0.06	∼0	∼0
8	2.56	2.13	1.12	0.36	0.10	0.02	∼0	∼0

Table A.39: SD-ratio value for the marginal posterior density of σ_{12} for a two-equation system with $N = 100$ observations and gprvar-g prior.

	ρ_{12}					
ν	0	0.1	0.2	0.3	0.4	0.5
2	582.85	60.55	0.09	∼ 0	∼ 0	∼ 0
3	584.22	60.67	0.09	∼ 0	∼ 0	∼ 0
4	585.58	60.73	0.09	∼ 0	∼ 0	∼ 0
8	590.71	61.05	0.09	∼ 0	∼ 0	∼ 0

Table A.40: Density value at zero for the marginal posterior density of σ_{12} for a two-equation system with $N = 1000$ observations and I-I prior.

	ρ_{12}					
ν	0	0.1	0.2	0.3	0.4	0.5
2	1189.49	123.57	0.18	~ 0	~ 0	~ 0
3	460.01	47.77	0.07	~ 0	~ 0	~ 0
4	260.26	26.99	0.04	~ 0	~ 0	~ 0
8	77.22	7.98	0.01	~ 0	~ 0	~ 0

Table A.41: SD-ratio value for a two-equation system with $N = 1000$ observations and I-I prior.

	ρ_{12}							
ν	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
2	585.90	63.30	0.09	~ 0	~ 0	~ 0	~ 0	~ 0
3	587.96	63.40	0.09	~ 0	~ 0	~ 0	~ 0	~ 0
4	588.53	63.46	0.09	~ 0	~ 0	~ 0	~ 0	~ 0
8	593.82	63.79	0.09	~ 0	~ 0	~ 0	~ 0	~ 0

Table A.42: Density value at zero for the marginal posterior density of σ_{12} for a two-equation system with $N = 1000$ observations and g-g prior.

	ρ_{12}							
ν	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
2	1195.71	129.18	0.18	~ 0	~ 0	~ 0	~ 0	~ 0
3	462.96	49.92	0.07	~ 0	~ 0	~ 0	~ 0	~ 0
4	261.57	28.20	0.04	~ 0	~ 0	~ 0	~ 0	~ 0
8	77.62	8.34	0.01	~ 0	~ 0	~ 0	~ 0	~ 0

Table A.43: SD-ratio value for the marginal posterior density of σ_{12} for a two-equation system with $N = 1000$ observations and g-g prior.

	ρ_{12}							
ν	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
2	585.43	61.22	0.08	~ 0	~ 0	~ 0	~ 0	~ 0
3	586.75	61.30	0.08	~ 0	~ 0	~ 0	~ 0	~ 0
4	588.06	61.38	0.08	~ 0	~ 0	~ 0	~ 0	~ 0
8	593.78	61.70	0.08	~ 0	~ 0	~ 0	~ 0	~ 0

Table A.44: Density value at zero for the marginal posterior density of σ_{12} for a two-equation system with $N = 1000$ observations and gprvar-g prior.

	ρ_{12}							
ν	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
2	1194.76	124.94	0.16	~ 0	~ 0	~ 0	~ 0	~ 0
3	462.01	48.26	0.06	~ 0	~ 0	~ 0	~ 0	~ 0
4	261.36	27.28	0.04	~ 0	~ 0	~ 0	~ 0	~ 0
8	77.62	8.06	0.01	~ 0	~ 0	~ 0	~ 0	~ 0

Table A.45: SD-ratio value for the marginal posterior density of σ_{12} for a two-equation system with $N = 1000$ observations and gprvar-g prior.

Marginal Posterior Simulation Results for Three-Equation System

Setting with $\rho_{12} = -0.7$ and $\rho_{13} = -0.6$							
		marginal posterior at zero			SD-ratio		
prior for Σ	ν	I-I	g-g	gprvar-g	I-I	g-g	gprvar-g
$\Sigma \sim IW(I_3, \nu)$	3	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	6	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	10	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
$\Sigma \sim IW(Q_{3,c}, \nu)$	3	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	6	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	10	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0

Table A.46: Density value of the marginal posterior of σ_{13} at zero and **SD-ratio** for a three-equation system and a sample size of $N = 100$.

Setting with $\rho_{12} = -0.4$ and $\rho_{13} = 0$							
		marginal posterior at zero			SD-ratio		
prior for Σ	ν	I-I	g-g	gprvar-g	I-I	g-g	gprvar-g
$\Sigma \sim IW(I_3, \nu)$	3	17.33	16.65	16.78	34.66	33.30	33.56
	6	18.51	17.77	17.90	5.48	5.26	5.30
	10	20.13	19.29	19.45	2.18	2.07	2.09
$\Sigma \sim IW(Q_{3,c}, \nu)$	3	17.33	16.65	16.79	45.61	43.82	44.18
	6	18.50	17.76	17.91	11.21	10.76	10.85
	10	20.15	19.29	19.45	7.90	7.56	7.63

Table A.47: Density value of the marginal posterior of σ_{13} at zero and **SD-ratio** for a three-equation system and a sample size of $N = 100$.

Setting with $\rho_{12} = -0.7$ and $\rho_{13} = -0.3$							
		marginal posterior at zero			SD-ratio		
prior for Σ	ν	I-I	g-g	gprvar-g	I-I	g-g	gprvar-g
	3	2.85	3.04	2.74	5.70	6.08	5.48
$\Sigma \sim IW(I_3, \nu)$	6	3.00	3.19	2.87	0.89	0.94	0.85
	10	3.19	3.39	3.05	0.34	0.36	0.33
	3	2.90	3.09	2.78	7.63	8.13	7.32
$\Sigma \sim IW(Q_{3,c}, \nu)$	6	3.05	3.24	2.91	1.85	1.96	1.76
	10	3.25	3.44	3.10	1.27	1.35	1.22

Table A.48: Density value of the marginal posterior of σ_{13} at zero and **SD-ratio** for a three-equation system and a sample size of $N = 100$.

Setting with $\rho_{12} = 0.7$ and $\rho_{13} = 0.6$							
		marginal posterior at zero			SD-ratio		
prior for Σ	ν	I-I	g-g	gprvar-g	I-I	g-g	gprvar-g
	3	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
$\Sigma \sim IW(I_3, \nu)$	6	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	10	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	3	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
$\Sigma \sim IW(Q_{3,c}, \nu)$	6	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	10	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0

Table A.49: Density value of the marginal posterior of σ_{13} at zero and **SD-ratio** for a three-equation system and a sample size of $N = 500$.

Setting with $\rho_{12} = -0.7$ and $\rho_{13} = -0.6$							
		marginal posterior at zero			SD-ratio		
prior for Σ	ν	I-I	g-g	gprvar-g	I-I	g-g	gprvar-g
	3	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
$\Sigma \sim IW(I_3, \nu)$	6	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	10	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	3	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
$\Sigma \sim IW(Q_{3,c}, \nu)$	6	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	10	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0

Table A.50: Density value of the marginal posterior of σ_{13} at zero and **SD-ratio** for a three-equation system and a sample size of $N = 500$.

Setting with $\rho_{12} = -0.7$ and $\rho_{13} = -0.3$							
		marginal posterior at zero			SD-ratio		
prior for Σ	ν	I-I	g-g	gprvar-g	I-I	g-g	gprvar-g
	3	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
$\Sigma \sim IW(I_3, \nu)$	6	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	10	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	3	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
$\Sigma \sim IW(Q_{3,c}, \nu)$	6	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
	10	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0

Table A.51: Density value of the marginal posterior of σ_{13} at zero and **SD-ratio** for a three-equation system and a sample size of $N = 500$.

E Additional Results: Empirical Example

Variable	Description
CPI	composite index which measures the perceived level of corruption, between 0 and 10 with higher values indicating less corruption
property rights	between 0 and 100 with higher values representing better protection of property rights
rule of law	rule of law estimate
absence of press freedom	between 0 and 100 with under 30: free, between 31-60: partly free and between 61 and 100: not free
trade freedom	between 0 and 100 with higher numbers indicating more freedom
government size	general government final consumption expenditure in % of GDP
level of democracy	between -10 (totally autocratic) and 10 (total democracy)
absence of political rights	measures whether people can participate freely in political process, between 1 (most free) and 7 (least free)
government effectiveness	between approximately -2.5 (weak) and 2.5 (strong)
ln(GDP/cap)	ln of GDP per capita in 2000 US$
trade openness	exports plus imports as % of GDP
imports as share of GDP	imports of goods an services as % of GDP
FDI	foreign direct investment as % of GDP
years democratic	consecutive years with *level of democracy*>0 before 2001
federal	dummy for federal structure
British	dummy for former colony
Spanish	dummy for former colony
Portuguese	dummy for former colony
French	dummy for former colony
Dutch	dummy for former colony
Europe	continental dummy
Africa	continental dummy
Asia	continental dummy
South America	continental dummy, also comprises the Caribbean
North America	continental dummy
seats held by women	proportion of seats held by women in national parliaments
duration secondary education	duration of secondary education in years
duration primary education	duration of primary education in years
sec. education enrollment rate	secondary school enrollment rate in %
common law	dummy for English common law system
ln(population)	ln of total population
ethnic fractionalisation	between 0 (low ethnic fractionalisation) and 1 (high ethnic fractionalisation)
language fractionalisation	between 0 (low language fractionalisation) and 1 (high language fractionalisation)
religious fractionalisation	between 0 (low religious fractionalisation) and 1 (high religious fractionalisation)
natural resources rents	total natural resources rents in % of GDP
share of Protestants	share of Protestants
urbanization rate	share of urban population in society

Table A.52: Description of variables used in the empirical application based on Table A1 in Jetter and Parmeter (2018).

Country	CPI	Country	CPI	Country	CPI
Albania	2.81	Gambia, The	2.60	Nepal	2.50
Algeria	2.89	Georgia	2.94	Netherlands	8.83
Angola	1.96	Ghana	3.66	New Zealand	9.46
Argentina	2.86	Greece	4.23	Nicaragua	2.55
Armenia	2.89	Guatemala	2.76	Niger	2.54
Australia	8.69	Guinea	1.84	Norway	8.63
Austria	8.13	Guinea-Bissau	2.03	Oman	5.64
Azerbaijan	2.10	Guyana	2.58	Pakistan	2.34
Bahrain	5.48	Honduras	2.51	Panama	3.40
Bangladesh	1.70	Hungary	5.02	Paraguay	2.11
Belarus	2.89	India	3.08	Peru	3.64
Benin	2.87	Indonesia	2.28	Philippines	2.52
Bolivia	2.53	Iran, Islamic Rep.	2.54	Poland	4.14
Botswana	5.82	Ireland	7.54	Portugal	6.30
Brazil	3.72	Israel	6.48	Qatar	6.24
Bulgaria	3.90	Italy	4.89	Romania	3.22
Burkina Faso	3.28	Jamaica	3.46	Russian Federation	2.41
Burundi	2.12	Japan	7.33	Saudi Arabia	3.81
Cambodia	2.05	Jordan	4.98	Senegal	3.16
Cameroon	2.17	Kazakhstan	2.47	Slovak Republic	4.28
Canada	8.70	Kenya	2.07	Slovenia	6.19
Chad	1.73	Korea, Rep.	4.92	South Africa	4.69
Chile	7.22	Kuwait	4.58	Spain	6.73
China	3.45	Kyrgyz Republic	2.08	Sri Lanka	3.29
Colombia	3.76	Lao PDR	2.32	Swaziland	3.15
Congo, Dem. Rep.	1.94	Latvia	4.24	Sweden	9.22
Congo, Rep.	2.13	Lesotho	3.32	Switzerland	8.87
Costa Rica	4.72	Libya	2.45	Syrian Arab Republic	2.84
Cote d'Ivoire	2.16	Lithuania	4.78	Tajikistan	2.04
Croatia	3.84	Madagascar	2.83	Thailand	3.44
Denmark	9.43	Malawi	2.94	Togo	2.52
Djibouti	2.98	Malaysia	4.93	Trinidad and Tobago	4.02
Dominican Republic	3.06	Mali	2.90	Tunisia	4.66
Ecuador	2.27	Mauritania	2.66	Turkey	3.79
Egypt, Arab Rep.	3.18	Mauritius	4.78	Uganda	2.44
Equatorial Guinea	1.88	Mexico	3.48	Ukraine	2.42
Estonia	6.20	Moldova	2.79	United Kingdom	8.29
Fiji	4.00	Mongolia	2.89	Uruguay	6.14
Finland	9.47	Morocco	3.37	Uzbekistan	2.14
France	6.98	Mozambique	2.71	Venezuela, RB	2.24
Gabon	3.04	Namibia	4.62	Vietnam	2.59

Table A.53: List of countries and their corresponding CPI based on Table A4 in Jetter and Parmeter (2018).

Variable	PIP	Mean	Lower	Median	Upper
property rights	0.01	0	0	0	0
rule of law	1	1.37	0.79	1.34	2.06
absence of press freedom	0.04	0	0	0	0.01
trade freedom	0.49	0.01	0	0	0.04
government size	0.03	0	0	0	0
level of democracy	0.33	0.03	0	0	0.13
absence of political rights	0.79	0.2	0	0.16	0.54
government effectiveness	0.86	0.57	0	0.6	1.16
ln(GDP/cap)	0.33	0.06	0	0	0.33
trade openness	0.01	0	0	0	0
imports as share of GDP	0.01	0	0	0	0
FDI	0.04	0	-0.01	0	0
years democratic	0.04	0	0	0	0.01
federal	0.15	-0.01	-0.24	0	0.08
British	0.21	-0.03	-0.34	0	0.01
Spanish	0.19	0.02	-0.13	0	0.32
Portuguese	0.31	0.07	-0.03	0	0.5
French	0.14	0	-0.12	0	0.16
Dutch	0.53	-0.19	-0.68	-0.07	0
Europe	0.23	-0.04	-0.47	0	0.11
Africa	0.21	-0.03	-0.4	0	0.14
Asia	0.43	-0.13	-0.58	0	0
South America	0.22	0.04	-0.04	0	0.41
North America	0.19	-0.01	-0.31	0	0.21
seats held by women	0.85	0.02	0	0.02	0.03
duration secondary education	0.23	0.03	0	0	0.22
duration primary education	0.85	0.18	0	0.19	0.36
sec. education enrollment rate	0	0	0	0	0
common law	0.14	-0.01	-0.19	0	0.11
ln(population)	0.05	0	-0.02	0	0
ethnic fractionalisation	0.28	0.06	-0.09	0	0.55
language fractionalisation	0.37	0.13	-0.02	0	0.71
religious fractionalisation	0.6	0.24	0	0.21	0.74
natural resources rents	0	0	0	0	0
share of Protestants	0.45	0.21	-0.11	0	1.05
urbanization rate	0.88	0.77	0	0.83	1.51

Table A.54: Posterior estimation results of the full system for the structural equation using the original I-I prior for model parameters while setting the prior of Σ to $\Sigma \sim IW(I_{13}, 14)$. PIP represents the posterior inclusion probability and Mean represents the posterior mean. Lower, Median and Upper are the 0.05, 0.5 and 0.95 posterior quantile, respectively. Variables marked in bold blue are assumed to be endogenous.

Variable	PIP	Mean	Lower	Median	Upper
property rights	0.16	0	-0.01	0	0.01
absence of press freedom	0.29	0	0	0	0.03
level of democracy	0.7	0.05	0	0.06	0.14
ln(GDP/cap)	0.36	0.06	0	0	0.31
trade openness	0.42	0	-0.01	0	0
FDI	0.52	-0.01	-0.17	0	0.21
rule of law	1	1.81	1.24	1.84	2.24
trade freedom	0.44	0	0	0	0.02
government size	0.18	0	0	0	0.02
absence of political rights	0.87	0.28	0	0.3	0.56
government effectiveness	0.44	0.19	0	0	0.77
imports as share of GDP	0.31	0	-0.02	0	0.02
years democratic	0.41	0	0	0	0.02
federal	0.11	-0.01	-0.2	0	0.05
British	0.16	-0.02	-0.29	0	0
Spanish	0.1	0.01	-0.09	0	0.19
Portuguese	0.23	0.06	0	0	0.5
French	0.1	-0.01	-0.14	0	0.05
Dutch	0.23	-0.07	-0.5	0	0
Europe	0.11	-0.01	-0.21	0	0.07
Africa	0.22	-0.05	-0.42	0	0
Asia	0.13	-0.02	-0.28	0	0.04
South America	0.14	0.02	0	0	0.34
North America	0.09	0.01	-0.05	0	0.2
seats held by women	1	0.02	0.01	0.02	0.03
duration secondary education	0.39	0.04	0	0	0.19
duration primary education	0.83	0.15	0	0.16	0.32
sec. education enrollment rate	0.11	0	0	0	0
common law	0.12	-0.01	-0.22	0	0.01
ln(population)	0.22	0.01	-0.05	0	0.09
ethnic fractionalisation	0.12	0.02	-0.03	0	0.37
language fractionalisation	0.3	0.11	0	0	0.68
religious fractionalisation	0.46	0.21	0	0	0.85
natural resources rents	0.25	0	-0.01	0	0
share of Protestants	0.21	0.11	0	0	0.94
urbanization rate	0.81	0.81	0	0.93	1.56

Table A.55: Posterior estimation results for the tested-down system for the structural equation using the original I-I prior for model parameters while setting the prior of Σ to $\Sigma \sim IW(I_7, 8)$. PIP represents the posterior inclusion probability and Mean represents the posterior mean. Lower, Median and Upper are the 0.05, 0.5 and 0.95 posterior quantile, respectively. Variables marked in bold blue are assumed to be endogenous.

Variable	PIP	Mean	Lower	Median	Upper
property rights	0.01	0	0	0	0
rule of law	1	1.39	0.84	1.39	1.96
absence of press freedom	0.08	0	0	0	0.02
trade freedom	0.42	0.01	0	0	0.03
government size	0.02	0	0	0	0
level of democracy	0.11	0	0	0	0.08
absence of political rights	0.51	0.08	0	0	0.37
government effectiveness	0.7	0.39	0	0.41	1.04
ln(GDP/cap)	0.45	0.1	0	0	0.36
trade openness	0.01	0	0	0	0
imports as share of GDP	0.01	0	0	0	0
FDI	0.05	0	-0.03	0	0
years democratic	0.03	0	0	0	0.01
federal	0.13	-0.01	-0.17	0	0.07
British	0.19	-0.03	-0.32	0	0.02
Spanish	0.17	0	-0.19	0	0.25
Portuguese	0.25	0.05	-0.03	0	0.44
French	0.14	0	-0.12	0	0.16
Dutch	0.63	-0.25	-0.73	-0.25	0
Europe	0.2	-0.03	-0.45	0	0.11
Africa	0.19	-0.03	-0.4	0	0.09
Asia	0.33	-0.09	-0.55	0	0
South America	0.25	0.05	-0.04	0	0.44
North America	0.22	-0.03	-0.44	0	0.17
seats held by women	0.92	0.02	0	0.02	0.03
duration secondary education	0.35	0.05	0	0	0.22
duration primary education	0.93	0.21	0	0.21	0.38
sec. education enrollment rate	0	0	0	0	0
common law	0.13	0	-0.15	0	0.1
ln(population)	0.04	0	0	0	0
ethnic fractionalisation	0.24	0.04	-0.16	0	0.48
language fractionalisation	0.46	0.18	0	0	0.76
religious fractionalisation	0.62	0.24	0	0.2	0.72
natural resources	0	0	0	0	0
share of Protestants	0.49	0.24	-0.09	0	1.07
urbanization rate	0.86	0.76	0	0.82	1.51

Table A.56: Posterior estimation results for the full system for the structural equation using the original I-I prior for model parameters while setting the prior of Σ to $\Sigma \sim IW(Q_{13,c}, 15)$. PIP represents the posterior inclusion probability and Mean represents the posterior mean. Lower, Median and Upper are the 0.05, 0.5 and 0.95 posterior quantile, respectively. Variables marked in bold blue are assumed to be endogenous.

Variable	PIP	Mean	Lower	Median	Upper
property rights	0.01	0	0	0	0
absence of political rights	0.09	0.01	0	0	0.08
rule of law	1	1.17	0.73	1.16	1.66
absence of press freedom	0.01	0	0	0	0
trade freedom	0.02	0	0	0	0
government size	0.08	0	0	0	0.02
level of democracy	0.02	0	0	0	0
government effectiveness	0.9	0.57	0	0.61	1.08
ln(GDP/cap)	0.33	0.07	0	0	0.37
trade openness	0	0	0	0	0
imports as share of GDP	0.01	0	0	0	0
FDI	0.01	0	0	0	0
years democratic	0.01	0	0	0	0
federal	0.13	0	-0.17	0	0.1
British	0.26	-0.05	-0.37	0	0.01
Spanish	0.17	-0.01	-0.27	0	0.17
Portuguese	0.19	0.02	-0.1	0	0.34
French	0.16	-0.02	-0.25	0	0.04
Dutch	0.46	-0.16	-0.68	0	0
Europe	0.15	0	-0.14	0	0.18
Africa	0.2	-0.03	-0.33	0	0.05
Asia	0.18	-0.03	-0.3	0	0.02
South America	0.2	0.03	-0.06	0	0.36
North America	0.21	-0.02	-0.36	0	0.2
seats held by women	0.64	0.01	0	0.02	0.03
duration secondary education	0.83	0.15	0	0.17	0.25
duration primary education	0.99	0.3	0.1	0.32	0.43
sec. education enrollment rate	0	0	0	0	0
common law	0.17	-0.01	-0.24	0	0.07
ln(population)	0.04	0	0	0	0
ethnic fractionalisation	0.23	0.03	-0.15	0	0.45
language fractionalisation	0.33	0.1	-0.03	0	0.63
religious fractionalisation	0.48	0.17	0	0	0.69
natural resources rents	0	0	0	0	0
share of Protestants	0.58	0.36	-0.02	0.23	1.25
urbanization rate	0.94	1.05	0	1.13	1.7

Table A.57: Posterior estimation results for the tested-down system for the structural equation using the original I-I prior for model parameters while setting the prior of Σ to $\Sigma \sim IW(Q_{3,c}, 5)$. PIP represents the posterior inclusion probability and Mean represents the posterior mean. Lower, Median and Upper are the 0.05, 0.5 and 0.95 posterior quantile, respectively. Variables marked in bold blue are assumed to be endogenous.

Variable	PIP	Mean	Lower	Median	Upper
property rights	0.05	0	0	0	0.02
rule of law	1	1.69	0.85	1.79	2.17
absence of press freedom	0	0	0	0	0
trade freedom	0.39	0.01	0	0	0.03
government size	0	0	0	0	0
level of democracy	0.64	0.06	0	0.08	0.14
absence of political rights	0.88	0.3	0	0.37	0.56
government effectiveness	0.32	0.2	0	0	1
ln(GDP/cap)	0.06	0.01	0	0	0.21
trade openness	0	0	0	0	0
imports as share of GDP	0	0	0	0	0
FDI	0	0	0	0	0
years democratic	0	0	0	0	0
federal	0.02	0	0	0	0
British	0.04	-0.01	-0.14	0	0
Spanish	0.03	0	0	0	0
Portuguese	0.07	0.02	0	0	0.39
French	0.01	0	0	0	0
Dutch	0.18	-0.08	-0.66	0	0
Europe	0.02	0	0	0	0
Africa	0.02	0	0	0	0
Asia	0.09	-0.03	-0.37	0	0
South America	0.03	0.01	0	0	0.03
North America	0.02	0	0	0	0
seats held by women	0.9	0.02	0	0.03	0.04
duration secondary education	0.03	0	0	0	0.09
duration primary education	0.89	0.19	0	0.2	0.35
sec. education enrollment rate	0	0	0	0	0
common law	0.02	0	0	0	0
ln(population)	0	0	0	0	0
ethnic fractionalisation	0.04	0.01	0	0	0.22
language fractionalisation	0.04	0.01	0	0	0.11
religious fractionalisation	0.42	0.22	0	0	0.82
natural resources rents	0	0	0	0	0
share of Protestants	0.2	0.15	0	0	1.17
urbanization rate	0.87	1	0	1.11	1.7

Table A.58: Posterior estimation results for the full system for the structural equation using the original g-g prior for model parameters while setting the prior of Σ to $\Sigma \sim IW(I_{13}, 14)$. PIP represents the posterior inclusion probability and Mean represents the posterior mean. Lower, Median and Upper are the 0.05, 0.5 and 0.95 posterior quantile, respectively. Variables marked in bold blue are assumed to be endogenous.

Variable	PIP	Mean	Lower	Median	Upper
property rights	0	0	0	0	0
absence of press freedom	0	0	0	0	0
ln(GDP/cap)	0.58	0.16	0	0.19	0.45
trade openness	0	0	0	0	0
rule of law	1	1.86	1.18	1.91	2.21
trade freedom	0.02	0	0	0	0
government size	0	0	0	0	0
level of democracy	0.8	0.07	0	0.08	0.14
absence of political rights	0.88	0.31	0	0.35	0.54
government effectiveness	0.13	0.07	0	0	0.78
imports as share of GDP	0	0	0	0	0
FDI	0	0	0	0	0
years democratic	0	0	0	0	0
federal	0.01	0	0	0	0
British	0.01	0	0	0	0
Spanish	0.02	0	0	0	0
Portuguese	0.06	0.02	0	0	0.3
French	0.01	0	0	0	0
Dutch	0.09	-0.03	-0.47	0	0
Europe	0.02	0	0	0	0
Africa	0.02	0	0	0	0
Asia	0.03	0	0	0	0
South America	0.04	0.01	0	0	0.26
North America	0.02	0	0	0	0
seats held by women	0.97	0.03	0	0.03	0.04
duration secondary education	0.11	0.02	0	0	0.24
duration primary education	0.45	0.09	0	0	0.33
sec. eduction enrollment rate	0	0	0	0	0
common law	0.01	0	0	0	0
ln(population)	0.01	0	0	0	0
ethnic fractionalisation	0.08	0.03	0	0	0.51
language fractionalisation	0.27	0.13	0	0	0.72
religious fractionalisation	0.11	0.04	0	0	0.56
natural resources rents	0	0	0	0	0
share of Protestants	0.13	0.08	0	0	0.98
urbanization rate	0.59	0.73	0	0.76	1.8

Table A.59: Posterior estimation results for the tested-down system for the structural equation using the original g-g prior for model parameters while setting the prior of Σ to $\Sigma \sim IW(I_5, 6)$. PIP represents the posterior inclusion probability and Mean represents the posterior mean. Lower, Median and Upper are the 0.05, 0.5 and 0.95 posterior quantile, respectively. Variables marked in bold blue are assumed to be endogenous.

Variable	PIP	Mean	Lower	Median	Upper
property rights	0.04	0	0	0	0.02
rule of law	1	1.64	0.96	1.7	2.11
absence of press freedom	0.01	0	0	0	0
trade freedom	0.57	0.01	0	0.02	0.03
government size	0	0	0	0	0
level of democracy	0.33	0.03	0	0	0.11
absence of political rights	0.85	0.18	0	0.13	0.46
government effectiveness	0.29	0.16	0	0	0.87
ln(GDP/cap)	0.12	0.03	0	0	0.32
trade openness	0	0	0	0	0
imports as share of GDP	0	0	0	0	0
FDI	0.01	0	0	0	0
years democratic	0	0	0	0	0
federal	0.01	0	0	0	0
British	0.03	0	0	0	0
Spanish	0.02	0	0	0	0
Portuguese	0.04	0.01	0	0	0.19
French	0.01	0	0	0	0
Dutch	0.15	-0.06	-0.57	0	0
Europe	0.02	0	0	0	0
Africa	0.02	0	0	0	0
Asia	0.06	-0.01	-0.26	0	0
South America	0.03	0.01	0	0	0.04
North America	0.02	0	0	0	0
seats held by women	0.94	0.02	0	0.03	0.04
duration of secondary education	0.03	0	0	0	0.05
duration of primary education	0.97	0.23	0	0.23	0.35
sec. education enrollment rate	0	0	0	0	0
common law	0.02	0	0	0	0
ln(population)	0	0	0	0	0
ethnic fractionalisation	0.03	0.01	0	0	0
language fractionalisation	0.09	0.04	0	0	0.58
religious fractionalisation	0.45	0.22	0	0	0.78
natural resources rents	0	0	0	0	0
share of Protestants	0.18	0.12	0	0	1.05
urbanization rate	0.91	1.05	0	1.12	1.71

Table A.60: Posterior estimation results for the full system for the structural equation using the original g-g prior for model parameters while setting the prior of Σ to $\Sigma \sim IW(Q_{13,c}, 15)$. PIP represents the posterior inclusion probability and Mean represents the posterior mean. Lower, Median and Upper are the 0.05, 0.5 and 0.95 posterior quantile, respectively. Variables marked in bold blue are assumed to be endogenous.

Variable	PIP	Mean	Lower	Median	Upper
property rights	0	0	0	0	0
rule of law	1	2.06	1.55	2.09	2.27
absence of press freedom	0	0	0	0	0
trade freedom	0	0	0	0	0
government size	0	0	0	0	0
level of democracy	0.97	0.1	0	0.1	0.13
absence of political rights	0.98	0.4	0.08	0.41	0.53
government effectiveness	0.09	0.04	0	0	0.58
ln(GDP/cap)	0.01	0	0	0	0
trade openness	0	0	0	0	0
imports as share of GDP	0	0	0	0	0
FDI	0	0	0	0	0
years democratic	0	0	0	0	0
federal	0.02	0	0	0	0
British	0.04	-0.01	-0.16	0	0
Spanish	0.02	0	0	0	0
Portuguese	0.03	0.01	0	0	0.08
French	0.01	0	0	0	0
Dutch	0.11	-0.04	-0.52	0	0
Europe	0.02	0	0	0	0
Africa	0.02	0	0	0	0
Asia	0.06	-0.01	-0.25	0	0
South America	0.03	0	0	0	0
North America	0.02	0	0	0	0
seats held by women	0.94	0.02	0	0.02	0.04
duration secondary education	0.03	0.01	0	0	0.11
duration primary education	0.98	0.18	0.08	0.18	0.3
sec. education enrollment rate	0	0	0	0	0
common law	0.02	0	0	0	0
ln(population)	0	0	0	0	0
ethnic fractionalisation	0.05	0.02	0	0	0.32
language fractionalisation	0.04	0.01	0	0	0.06
religious fractionalisation	0.73	0.46	0	0.54	0.98
natural resources	0	0	0	0	0
share of Protestants	0.19	0.15	0	0	1.22
urbanization rate	0.99	1.36	0.9	1.37	1.8

Table A.61: Posterior estimation results for the down-tested system for the structural equation using the original g-g prior for model parameters while setting the prior of Σ to $\Sigma \sim IW(Q_{2,c}, 4)$. PIP represents the posterior inclusion probability and Mean represents the posterior mean. Lower, Median and Upper are the 0.05, 0.5 and 0.95 posterior quantile, respectively. Variables marked in bold blue are assumed to be endogenous.

Variable	PIP	Mean	Lower	Median	Upper
property rights	0.33	0	0	0	0.02
rule of law	1	1.55	0.98	1.56	2.06
absence of press freedom	0.15	0	-0.01	0	0.01
trade freedom	0.9	0.02	0	0.02	0.03
government size	0.16	0	0	0	0.02
level of democracy	0.69	0.04	0	0.05	0.11
absence of political rights	0.98	0.3	0.08	0.3	0.54
government effectiveness	0.63	0.31	0	0.31	0.85
ln(GDP/cap)	0.22	0.02	-0.01	0	0.22
trade openness	0.31	0	-0.02	0	0
imports as share of GDP	0.23	0	-0.01	0	0
FDI	0.44	0.02	-0.12	0	0.33
years democratic	0.32	0	0	0	0.01
federal	0.12	-0.01	-0.22	0	0
British	0.15	-0.02	-0.28	0	0
Spanish	0.16	0.03	-0.01	0	0.32
Portuguese	0.32	0.1	0	0	0.56
French	0.1	0	-0.06	0	0.09
Dutch	0.38	-0.13	-0.65	0	0
Europe	0.16	-0.03	-0.4	0	0.04
Africa	0.14	-0.01	-0.29	0	0.1
Asia	0.33	-0.08	-0.47	0	0
South America	0.14	0.02	-0.02	0	0.3
North America	0.1	0.01	-0.06	0	0.2
seats held by women	1	0.02	0.01	0.02	0.04
duration secondary education	0.15	0	-0.03	0	0.09
duration primary education	0.89	0.16	0	0.17	0.29
sec. education enrollment rate	0.12	0	0	0	0
common law	0.11	-0.01	-0.17	0	0.06
ln(population)	0.31	-0.02	-0.09	0	0
ethnic fractionalisation	0.11	0.01	-0.1	0	0.31
language fractionalisation	0.17	0.04	0	0	0.47
religious fractionalisation	0.66	0.3	0	0.32	0.81
natural resources rents	0.11	0	0	0	0
share of Protestants	0.19	0.09	0	0	0.89
urbanization rate	0.87	0.79	0	0.86	1.48

Table A.62: Posterior estimation results for the full system for the structural equation using the original gprvar-g prior for model parameters while setting the prior of Σ to $\Sigma \sim IW(I_{13}, 14)$. PIP represents the posterior inclusion probability and Mean represents the posterior mean. Lower, Median and Upper are the 0.05, 0.5 and 0.95 posterior quantile, respectively. Variables marked in bold blue are assumed to be endogenous.

Variable	PIP	Mean	Lower	Median	Upper
property rights	0.15	0	-0.01	0	0.01
absence of press freedom	0.17	0	-0.01	0	0.02
ln(GDP/cap)	0.33	0.05	0	0	0.27
rule of law	1	1.89	1.33	1.94	2.24
trade freedom	0.44	0.01	0	0	0.02
government size	0.17	0	0	0	0.02
level of democracy	0.91	0.08	0	0.08	0.13
absence of political rights	0.96	0.35	0	0.38	0.56
government effectiveness	0.31	0.12	0	0	0.66
trade openness	0.31	0	-0.01	0	0
imports as share of GDP	0.22	0	-0.01	0	0
FDI	0.17	0	0	0	0.04
years democratic	0.28	0	0	0	0.01
federal	0.1	-0.01	-0.17	0	0.02
British	0.13	-0.02	-0.25	0	0
Spanish	0.12	0.02	-0.05	0	0.28
Portuguese	0.19	0.05	0	0	0.44
French	0.11	-0.01	-0.16	0	0.03
Dutch	0.27	-0.09	-0.57	0	0
Europe	0.12	-0.01	-0.22	0	0.11
Africa	0.18	-0.03	-0.39	0	0.02
Asia	0.19	-0.04	-0.37	0	0
South America	0.13	0.02	0	0	0.36
North America	0.09	0	-0.09	0	0.18
seats held by women	1	0.02	0.01	0.02	0.04
duration secondary education	0.24	0.02	-0.01	0	0.18
duration primary education	0.81	0.13	0	0.15	0.29
sec. education enrollment rate	0.11	0	0	0	0
common law	0.1	-0.01	-0.18	0	0.05
ln(population)	0.15	0	-0.06	0	0.01
ethnic fractionalisation	0.14	0.03	-0.03	0	0.43
language fractionalisation	0.22	0.07	0	0	0.59
religious fractionalisation	0.48	0.24	0	0	0.89
natural resources rents	0.1	0	0	0	0
share of Protestants	0.25	0.15	0	0	1.1
urbanization rate	0.86	0.9	0	1.01	1.6

Table A.63: Posterior estimation results for the tested-down system for the structural equation using the original gprvar-g prior for model parameters while setting the prior of Σ to $\Sigma \sim IW(I_4, 5)$. PIP represents the posterior inclusion probability and Mean represents the posterior mean. Lower, Median and Upper are the 0.05, 0.5 and 0.95 posterior quantile, respectively. Variables marked in bold blue are assumed to be endogenous.

Variable	PIP	Mean	Lower	Median	Upper
property rights	0.38	0	0	0	0.02
rule of law	1	1.58	1.01	1.61	1.98
absence of press freedom	0.17	0	-0.01	0	0.01
trade freedom	0.97	0.02	0	0.02	0.03
government size	0.19	0	0	0	0.02
level of democracy	0.63	0.03	0	0.03	0.09
absence of political rights	0.95	0.24	0	0.24	0.46
government effectiveness	0.39	0.15	0	0	0.69
ln(GDP/cap)	0.16	0	-0.12	0	0.1
trade openness	0.29	0	-0.01	0	0
imports as share of GDP	0.35	0	-0.02	0	0.01
FDI	0.5	-0.01	-0.13	0	0.19
years democratic	0.22	0	0	0	0.01
federal	0.09	0	-0.13	0	0.06
British	0.13	-0.01	-0.23	0	0.02
Spanish	0.15	0.02	-0.02	0	0.31
Portuguese	0.34	0.1	0	0	0.55
French	0.09	0	-0.06	0	0.08
Dutch	0.4	-0.14	-0.62	0	0
Europe	0.15	-0.03	-0.39	0	0.02
Africa	0.13	-0.01	-0.25	0	0.13
Asia	0.25	-0.05	-0.43	0	0
South America	0.15	0.03	0	0	0.34
North America	0.09	0	-0.07	0	0.15
seats held by women	1	0.02	0.01	0.02	0.03
duration secondary education	0.12	0	-0.03	0	0.05
duration primary education	0.94	0.17	0	0.18	0.29
sec. education enrollment rate	0.18	0	-0.01	0	0
common law	0.1	0	-0.12	0	0.05
ln(population)	0.2	-0.01	-0.07	0	0.01
ethnic fractionalisation	0.12	0.02	-0.06	0	0.34
language fractionalisation	0.24	0.07	0	0	0.52
religious fractionalisation	0.78	0.38	0	0.43	0.83
natural resources rents	0.12	0	0	0	0
share of Protestants	0.17	0.07	0	0	0.76
urbanization rate	0.86	0.74	0	0.79	1.45

Table A.64: Posterior estimation results for the full system for the structural equation using the original gprvar-g prior for model parameters while setting the prior of Σ to $\Sigma \sim IW(Q_{13,c}, 15)$. PIP represents the posterior inclusion probability and Mean represents the posterior mean. Lower, Median and Upper are the 0.05, 0.5 and 0.95 posterior quantile, respectively. Variables marked in bold blue are assumed to be endogenous.

Variable	PIP	Mean	Lower	Median	Upper
property rights	0.72	0.01	0	0.01	0.03
rule of law	1	1.42	0.81	1.43	1.98
absence of press freedom	0.14	0	-0.01	0	0.01
governments size	0.24	0	0	0	0.04
ln(GDP/cap)	0.41	0.07	0	0	0.31
imports as share of GDP	0.5	-0.04	-0.19	0	0.01
FDI	0.55	-0.01	-0.29	0	0.27
trade freedom	0.47	0.01	0	0	0.02
level of democracy	0.77	0.04	0	0.05	0.1
absence of political rights	0.98	0.28	0.09	0.28	0.45
government effectiveness	0.33	0.11	-0.19	0	0.69
trade openness	0.52	0.02	-0.01	0	0.1
years democratic	0.36	0	0	0	0.01
federal	0.11	-0.01	-0.2	0	0.01
British	0.14	-0.02	-0.26	0	0
Spanish	0.14	0.02	-0.03	0	0.33
Portuguese	0.18	0.04	0	0	0.43
French	0.11	-0.01	-0.15	0	0.02
Dutch	0.25	-0.07	-0.54	0	0
Europe	0.16	-0.03	-0.39	0	0.04
Africa	0.19	-0.05	-0.5	0	0.02
Asia	0.22	-0.05	-0.47	0	0
South America	0.18	0.04	0	0	0.45
North America	0.1	0	-0.11	0	0.2
seats held by women	0.99	0.02	0.01	0.02	0.03
duration secondary education	0.24	0.02	-0.01	0	0.15
duration primary education	0.79	0.14	0	0.15	0.31
sec. education enrollment rate	0.1	0	0	0	0
common law	0.12	-0.01	-0.21	0	0.03
ln(population)	0.23	-0.01	-0.08	0	0
ethnic fractionalisation	0.13	0.02	-0.08	0	0.36
language fractionalisation	0.37	0.16	0	0	0.78
religious fractionalisation	0.37	0.14	0	0	0.68
natural resources rents	0.4	-0.01	-0.07	0	0
share of Protestants	0.23	0.13	0	0	1
urbanization rate	0.82	0.85	0	0.94	1.61

Table A.65: Posterior estimation results for the tested-down system for the structural equation using the original gprvar-g prior for model parameters while setting the prior of Σ to $\Sigma \sim IW(Q_{8,c}, 10)$. PIP represents the posterior inclusion probability and Mean represents the posterior mean. Lower, Median and Upper are the 0.05, 0.5 and 0.95 posterior quantile, respectively. Variables marked in bold blue are assumed to be endogenous.