Bayesian estimation of the network autocorrelation model

Dittrich, D.; Leenders, R.T.A.J.; Mulder, J.

Published in:
Social Networks

DOI:
10.1016/j.socnet.2016.09.002

Publication date:
2017

Document Version
Peer reviewed version

Citation for published version (APA):
Dittrich, D., Leenders, R. T. A. J., & Mulder, J. (2017). Bayesian estimation of the network autocorrelation model. *Social Networks*, *48*, 213-236. https://doi.org/10.1016/j.socnet.2016.09.002

Bayesian estimation of the network autocorrelation model[☆]

Dino Dittrich[a,*], Roger Th.A.J. Leenders[b], Joris Mulder[a]

[a]*Department of Methodology and Statistics, Tilburg University, The Netherlands*
[b]*Department of Organization Studies, Tilburg University*

Abstract

The network autocorrelation model has been extensively used by researchers interested modeling social influence effects in social networks. The most common inferential method in the model is classical maximum likelihood estimation. This approach, however, has known problems such as negative bias of the network autocorrelation parameter and poor coverage of confidence intervals. In this paper, we develop new Bayesian techniques for the network autocorrelation model that address the issues inherent to maximum likelihood estimation. A key ingredient of the Bayesian approach is the choice of the prior distribution. We derive two versions of Jeffreys prior, the Jeffreys rule prior and the Independence Jeffreys prior, which have not yet been developed for the network autocorrelation model. These priors can be used for Bayesian analyses of the model when prior information is completely unavailable. Moreover, we propose an informative as well as a weakly informative prior for the network autocorrelation parameter that are both based on an extensive literature review of empirical applications of the network autocorrelation model across many fields. Finally, we provide new and efficient Markov Chain Monte Carlo algorithms to sample from the resulting posterior distributions. Simulation results suggest that the considered Bayesian estimators outperform the maximum likelihood estimator with respect to bias and frequentist coverage of credible and confidence intervals.

Keywords: Network autocorrelation model, Bayesian inference, Jeffreys rule prior, informative prior distribution, frequentist coverage

1. Introduction

Identifying and estimating network influence on individual behavior is a common and important challenge encountered in social network analysis. Throughout the last decades, a number of different models studying network influence effects have emerged, out of which the network autocorrelation model is probably the most popular one (Doreian, 1980; Leenders, 2002; Marsden and Friedkin, 1993; Plümper and Neumayer, 2010; Wang et al., 2014).

A traditional and widely used technique for parameter estimation in the network autocorrelation model is maximum likelihood estimation (Doreian, 1981; Ord, 1975) which has also been implemented in common statistical software packages such as R (Bivand and Piras, 2015; Butts, 2008;

[☆]Accepted for publication in Social Networks.
[*]Corresponding author
Email address: dino.dittrich@gmail.com (Dino Dittrich)

Leifeld et al., 2015; McMillen, 2013; Wilhelm and Godinho de Matos, 2015), MATLAB (LeSage, 1999), Python (Rey and Anselin, 2007), or Stata (Pisati, 2001). Despite the popularity and usefulness of maximum likelihood estimation, there are also important issues related to this estimation technique of the model. First, several simulation studies suggest that the maximum likelihood estimator for the network effect ρ is negatively biased under many different scenarios, that the underestimation of ρ becomes more severe for increasing network density, and that it occurs regardless of the network structure and the network size (Dow et al., 1982a; Mizruchi and Neuman, 2008; Neuman and Mizruchi, 2010; Smith, 2009). Second, maximum likelihood-based precision estimates, such as confidence intervals, rely heavily on asymptotic theory. Consequently, the coverage rates of the associated confidence intervals may be distorted for small to medium sample sizes which are often encountered in social science research, such as school classes, care teams, or members of an executive board. Notwithstanding the tremendous capability of the network autocorrelation model and the theoretical advances it has yielded for understanding the structure of social influence in social networks, the concerns regarding the maximum likelihood estimation approach may ultimately discourage researchers from utilizing the model at all.

In this paper, we develop Bayesian statistical estimation methods for the network autocorrelation model that may attenuate the issues which have been encountered with maximum likelihood estimation. The Bayesian approach has at least two attractive features that are not shared by classical methods. First, it allows researchers to incorporate external information about the model parameters via a *prior distribution*. For example, if previous research suggests that people in a certain network are positively influenced by each other, as is often the case in social networks, one could specify a prior distribution that assumes positive values for the network autocorrelation ρ to be more likely than negative ones. Indeed, as we will show in Section 4.4 of this paper, the vast empirical literature on the model suggests that network effects are much more likely, a priori, to be in certain intervals than in others. Second, Bayesian analysis provides "exact" inference without the need for asymptotic approximations (Berger, 2006; De Oliveira and Song, 2008). This characteristic is especially appealing for small to moderate-sized groups and can be seen as a distinct advantage of the Bayesian approach over classical, frequentist methods. In other words, when networks are small, Bayesian estimation of the network autocorrelation model is statistically preferable over frequentist estimation.

Bayesian statistics is a fundamentally different approach than classical statistics. In brief, a Bayesian data analysis is carried out as follows. First, a prior distribution, or simply *prior*, for the model parameters is needed, where the prior distribution reflects the prior knowledge about the model parameters before observing the data. If prior information is available, for example based on published literature, an *informative prior* can be specified. On the other hand, if such information is absent a so-called *non-informative prior*, or *default prior*, can be employed (Berger, 2006). After observing the data, Bayes' theorem is used to update the prior expectations with the information contained in the data to arrive at the *posterior distribution*, or *posterior*, for the model parameters. All inference is based on the posterior, and it is used to obtain Bayesian point estimates and *credible intervals*, the Bayesian equivalent to classical confidence intervals.

Although the specification of the prior distribution is one of the most important steps in any Bayesian analysis, it has not received much attention in the literature on Bayesian estimation of the network autocorrelation model (Han and Lee, 2013; Hepple, 1979, 1995a,b; Holloway et al., 2002; LeSage, 2000; LeSage and Pace, 2009), with the exception of LeSage (1997a) and LeSage and Parent (2007). In some cases, it is in fact difficult to elicit a prior, for example when prior information is absent, or a researcher would like to add as little prior information as possible to the analysis. For these situations, non-informative priors are typically used to carry out a Bayesian analysis. In this paper, we are the first to derive two versions of *Jeffreys prior* (Jeffreys, 1961), called the *Jeffreys rule prior* and the *Independence Jeffreys prior*, for the network autocorrelation model and to establish results on the propriety of the resulting posterior distributions. Jeffreys rule prior construes the concept of a non-informative prior in a formal way and is the most commonly used non-informative prior (De Oliveira and Song, 2008). Moreover, in several simulation studies of related autoregressive models, the Independence Jeffreys prior has been shown to result in superior inferences compared to those based on maximum likelihood estimation (De Oliveira, 2010, 2012; De Oliveira and Song, 2008). These findings serve as another motivation to consider these two versions of Jeffreys prior for the network autocorrelation model as well.

Furthermore, we provide a novel, informative prior for the network effect ρ based on an extensive literature review of empirical applications of the network autocorrelation model. To the best of our knowledge, this is the first, empirically justified informative prior for ρ to be found in the literature. Because of the empirical justification of this prior, it is a reasonable "entry point" for a Bayesian analysis of the network autocorrelation model, as it summarizes the currently available evidence about observed network autocorrelations from many different sources. Moreover, we introduce a related *weakly informative prior* for ρ which can be used by a researcher who agrees that past findings should not be dismissed but who is at the same time reluctant or deliberately refrains from including all available prior information.

In addition, we present efficient *Markov Chain Monte Carlo* (MCMC) algorithms to sample from the resulting posterior distributions which we find to be computationally superior compared to existing schemes (LeSage, 2000; LeSage and Pace, 2009). An R package will be made readily available by the time of publication of this paper and will allow researchers to run Bayesian analyses from within their familiar statistical environment.

We conduct a simulation study to investigate numerical properties of Bayesian inferences about the network effect ρ and the error variance σ^2 based on the proposed priors and to compare them to inferences coming from maximum likelihood estimation. As will be shown, the Bayesian estimator based on the informative prior performs overall the best when network effects are positive, while using the weakly informative prior eliminates virtually all the negative bias in the estimation of ρ in case of no or marginal network effects.

We proceed as follows: In Section 2, we discuss the network autocorrelation model in more detail. We continue with a short introduction of the Bayesian approach in regard to the model in Section 3. In Section 4, we derive two versions of Jeffreys prior and propose an informative as well as a weakly informative prior for the network autocorrelation parameter ρ based on reported network effects from the literature. Moreover, we state properties of these priors and their corresponding posteriors and provide comparisons between the priors. Section 5 presents efficient MCMC implementations for Bayesian estimation of the model. We assess the numerical performance of the Bayesian estimators and the maximum likelihood estimator in a simulation study in Section 6. Section 7 concludes.

2. Network Autocorrelation Model

Originally developed by geographers (Ord, 1975), the network autocorrelation model has been used to address the problem of structured dependence ever since. In contrast to a standard linear regression model, the network autocorrelation model does not assume the observations to be independent from each other but allows for dependence among them. In a social network context, this has the interpretation that ego's opinion may not solely depend on exogenous variables; instead, ego's opinion might be influenced by the opinions of other actors in the network as well. Thus, in the network autocorrelation model ego's opinion is viewed as a combination of interaction and exogenous variables, formally expressed as

$$\boldsymbol{y} = \rho W \boldsymbol{y} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N\left(0, \sigma^2 I_g\right) \tag{1}$$

where, as in standard linear regression, $\boldsymbol{y}$ is a vector of length g consisting of values of a dependent variable for the g network actors, X is a $(g \times k)$ matrix of values for the actors on k covariates (including a vector of ones in the first column for the intercept term), $\boldsymbol{\beta}$ is a vector of regression coefficients of length k, I_g symbolizes the $(g \times g)$ identity matrix, and $\boldsymbol{\varepsilon}$ is a vector of length g containing independent and identically normally distributed error terms with zero mean and variance of σ^2. Furthermore, W denotes a given $(g \times g)$ connectivity matrix representing social ties in a network, where each entry W_{ij} $(i, j \in \{1, ..., g\})$ stands for the degree of influence of actor j (alter) on actor i (ego). By convention, we exclude loops, i.e. relationships from an actor to himself, so $W_{ii} = 0$ for all $i \in \{1, ..., g\}$. Finally, ρ is a scalar termed the *network autocorrelation parameter*. It is the key parameter of the model and measures the level of network influence for given $\boldsymbol{y}$, W, and X. Note that when $\rho = 0$, the model reduces to the classical linear regression model.

The likelihood of the model in (1) is given by (Doreian, 1980)

$$f\left(\boldsymbol{y}; \rho, \sigma^2, \boldsymbol{\beta}\right) = \left|\det\left(A_\rho\right)\right| \left(2\pi\sigma^2\right)^{-\frac{g}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)^T \left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)\right), \tag{2}$$

where $A_\rho := I_g - \rho W$. To ensure that $\left|\det\left(A_\rho\right)\right|$ is non-zero and the model's likelihood function in (2) is well-defined, there are restrictions on the feasible values for ρ. In practice, the admissible range for ρ is usually chosen as the interval containing $\rho = 0$ for which A_ρ is non-singular (Hepple, 1995a; Holloway et al., 2002; LeSage, 1997a, 2000; LeSage and Pace, 2009; Smith, 2009). This interval is given by $\left(\lambda_g^{-1}, \lambda_1^{-1}\right)$, where $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_g$ are the ordered eigenvalues of W (Hepple, 1995a), and we will follow this convention in the remainder.[1,2] We denote the resulting set of model parameters as $\boldsymbol{\theta} := \left(\rho, \sigma^2, \boldsymbol{\beta}\right)$ and the associated parameter space as $\Omega := \Omega_\rho \times \Omega_{\sigma^2} \times \Omega_{\boldsymbol{\beta}} = \left(\lambda_g^{-1}, \lambda_1^{-1}\right) \times (0, \infty) \times \mathbb{R}^k$. Hence, the parameter space of the model has the remarkable property that it depends on properties, i.e. the eigenvalues, of the connectivity matrix W.

Throughout the literature the model is also referred to as mixed regressive-autoregressive model

[1] To avoid unnecessary complications, we restrict ourselves to connectivity matrices with real eigenvalues. These include e.g. all W that either are symmetric or *row standardizations*, i.e. where each row sums to unity, of symmetric matrices (Smith, 2009). Furthermore, we assume that $\lambda_1 > 0$ which includes all nonzero symmetric connectivity matrices (Smith, 2009), so $\lambda_g < 0 < \lambda_1$ since $\operatorname{tr}(W) = 0$. In the common case of row-standardized connectivity matrices, it follows that $\lambda_1 = 1$ (Anselin, 1982).

[2] It is mathematically not necessary to constrain the parameter space of ρ to $\left(\lambda_g^{-1}, \lambda_1^{-1}\right)$. It suffices to exclude the reciprocals of the eigenvalues of W, $\left\{{\lambda_1}^{-1}, {\lambda_2}^{-1}, ..., {\lambda_g}^{-1}\right\}$, from the domain of ρ as A_ρ is singular only for those values (Leenders, 1995). Some authors prefer to restrict ρ to $(-1, 1)$ as for $\rho \in (-1, 1)$, $A^{-1} = \sum_{k=0}^{\infty} \rho^k W^k$, which implies an underlying stationary process (Griffith, 1979). We choose the interval $\left(\lambda_g^{-1}, \lambda_1^{-1}\right)$ as admissible range for ρ, rather than $(-1, 1)$, as the latter choice might yield estimates for ρ at the lower boundary of the interval, and considering the whole parameter space $\mathbb{R} \backslash \left\{{\lambda_1}^{-1}, {\lambda_2}^{-1}, ..., {\lambda_g}^{-1}\right\}$ typically results in improper posterior distributions (see the remark in the proof of Corollary 1). Lastly, $\forall \rho \in \left(\lambda_g^{-1}, \lambda_1^{-1}\right)$: $\det\left(A_\rho\right) > 0$, so we can write $|A_\rho|$ for $\left|\det\left(A_\rho\right)\right|$ in the following.

(Ord, 1975), spatial effects model (Doreian, 1980), network effects model (Dow et al., 1982b), or spatial lag model (Anselin, 2002), and it has been applied in many different fields, such as sociology (Kirk and Papachristos, 2011; Land et al., 1991; Ruggles, 2007), political science (Beck et al., 2006; Shin and Ward, 1999; Tam Cho, 2003), criminology (Baller et al., 2001; Fornango, 2010; Tita and Radil, 2011), and geography (Fingleton, 2001; McMillen, 2010; Seldadyo et al., 2010).

3. Bayesian network autocorrelation modeling

The starting point of every Bayesian analysis is the formulation of prior expectations for the parameters in a statistical model. Formally, these prior expectations are expressed in terms of probability distributions, where the resulting prior distributions represent the available knowledge about the model parameters before observing data. We denote the joint prior distribution for all model parameters as $p(\boldsymbol{\theta})$. In general, prior expectations can come from the researcher's beliefs or from accumulated empirical evidence from previous studies in a field. Alternatively, one might also (purposely) stay vague and opt for a non-informative prior distribution. The idea of a non-informative prior is that it is completely dominated by the data, and different methods have been proposed how to construct such priors (Bernardo, 1979; Box and Tiao, 1973; Jeffreys, 1946; Kass and Wasserman, 1996).

After having specified a prior distribution, the sample $\boldsymbol{y}$ is observed. Since the sample contains information about the unknown parameters, it can be used to update the initial expectations. The information in the data for the model parameters is summarized by the likelihood function, $f(\boldsymbol{y}|\boldsymbol{\theta})$, given in equation (2). Linking information from the prior distribution and the data leads to the posterior distribution of the model parameters, given the observations $\boldsymbol{y}$. We denote the posterior by $p(\boldsymbol{\theta}|\boldsymbol{y})$. Applying elementary rules of probability theory, the posterior can be written by Bayes' theorem as

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\boldsymbol{y})}. \tag{3}$$

The denominator in (3) is called the marginal likelihood and serves as normalizing constant to ensure that the posterior integrates to unity. However, as the normalizing constant does not depend on the model parameters and does not affect parameter estimation, the expression in (3) can be simplified to

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}). \tag{4}$$

Hence, (4) means that the posterior distribution is proportional, with respect to the model parameters $\boldsymbol{\theta}$, to the prior distribution multiplied by the likelihood function. Formally, the normalizing constant can only be dropped if it is finite, so if the posterior is integrable and thus a proper probability distribution. For the network autocorrelation model this is the case when the network size, compared to the number of covariates, is large enough. We will come back to this in the following section.

The posterior distribution can then be used to derive point estimates for the model parameters $\boldsymbol{\theta}$ (e.g. the posterior mean or the posterior median), credible intervals (i.e. intervals in the domain of the posterior), or to get other statistics of interest, such as the probability that the network autocorrelation is positive for given data, $P(\rho > 0|\boldsymbol{y})$. The latter statistic is quite useful to quantify a researcher's belief that people in a network positively influence each other regarding the observations $\boldsymbol{y}$. However, such a probability cannot be obtained when using classical, frequentist methods but only when taking the Bayesian route.

4. Prior choices for the network autocorrelation model

The specification of the prior distribution is one of the most important steps in a Bayesian analysis. Despite its importance, prior specification in the network autocorrelation model has been largely neglected. Most of the previous work on Bayesian estimation of the model is based on using uniform priors for $\rho, \boldsymbol{\beta}$, and $\log\left(\sigma^2\right)$ (Hepple, 1979, 1995a,b; Holloway et al., 2002; LeSage, 1997a, 2000). Only recently, Han and Lee (2013) and LeSage and Pace (2009) considered the standard normal-inverse gamma priors for $\boldsymbol{\beta}$ and σ^2 from linear regression, resulting in a normal prior distribution for $\boldsymbol{\beta}$ conditional on σ^2, an inverse gamma prior for σ^2, and the standard uniform prior for ρ.

In this section, we briefly review the common uniform prior for ρ first, before deriving two versions of Jeffreys prior and proposing two novel, informative priors for the network effect ρ.

4.1. Flat prior

Using *flat*, or uniform, priors is the simplest and most intuitive way to quantify prior ignorance about model parameters. A uniform prior assigns equal, or uniform, probability to all possible values a parameter can attain, resulting in a flat prior density function. Applying this rationale to the network autocorrelation model means that all possible network effects ρ and regression coefficients $\boldsymbol{\beta}$ are considered as equally likely before observing the data $\boldsymbol{y}$. In mathematical notation, we denote the flat prior distributions for ρ and $\boldsymbol{\beta}$ as $p_{\mathrm{F}}(\rho) \propto 1$ and $p_{\mathrm{F}}(\boldsymbol{\beta}) \propto 1$, respectively. As noted in the previous section, for estimation purposes it suffices to give the prior distributions in these unnormalized forms. Furthermore, the error variance σ^2 is constrained to the positive axis by definition, and it is customary to consider its logarithm and assign a flat prior to this transformed variable (Fernandez et al., 2001; Kass and Wasserman, 1996). Retransforming the flat prior on $\log\left(\sigma^2\right)$ back in terms of σ^2 yields $p_{\mathrm{F}}\left(\sigma^2\right) \propto 1/\sigma^2$. Finally, note that under the flat prior all parameters are assumed to be a priori independent. The flat prior for $\boldsymbol{\theta} = \left(\rho, \sigma^2, \boldsymbol{\beta}\right)$ is then written as

$$p_{\mathrm{F}}(\boldsymbol{\theta}) = p_{\mathrm{F}}(\rho) \times p_{\mathrm{F}}\left(\sigma^2\right) \times p_{\mathrm{F}}(\boldsymbol{\beta}) \propto 1/\sigma^2.$$

This prior is sometimes also referred to as the *diffuse* prior in the literature (Hepple, 1979; LeSage, 1997a, 2000). While it is obvious that the flat prior itself is improper, i.e. the integral of $p_{\mathrm{F}}(\boldsymbol{\theta})$ over Ω is not finite, it is easy to verify that the resulting posterior distribution is proper under the very weak conditions stated in Corollary 1.

Corollary 1. *Consider the network autocorrelation model given in* (1)*. Then,*

(i) *The flat prior $p_F(\boldsymbol{\theta})$ is unbounded and not integrable on $\Omega = \left(\lambda_g^{-1}, \lambda_1^{-1}\right) \times (0, \infty) \times \mathbb{R}^k$.*

(ii) *The corresponding posterior $p_F(\boldsymbol{\theta}|\boldsymbol{y})$ is proper on $\Omega = \left(\lambda_g^{-1}, \lambda_1^{-1}\right) \times (0, \infty) \times \mathbb{R}^k$ when $g > k$, $\left(X^T X\right)^{-1}$ exists, and $\left(\boldsymbol{y}^T M W \boldsymbol{y}\right)^2 \neq \boldsymbol{y}^T W^T M W \boldsymbol{y} \boldsymbol{y}^T M \boldsymbol{y}$, where $M := I_g - X\left(X^T X\right)^{-1} X^T$.*

Proof. See Appendix.

Thus, given the two mild regularity conditions in Corollary 1 (ii) hold, the flat prior yields a proper posterior when the number of actors in a network is larger than the number of external covariates. While the first regularity condition can be easily controlled for by avoiding perfect collinearity, the second one is of technical nature and needs to be checked for each data set.

4.2. Jeffreys rule prior

Flat priors are only one possible way to state prior ignorance; they are driven mainly by what intuitively seems to represent non-informativeness, rather than being based on a set of formal rules that defines non-informativeness mathematically. The first formal rule for specifying non-informative prior distributions was introduced by Sir Harold Jeffreys, and much of subsequent, related work is based on modifications of Jeffreys' scheme (Jeffreys, 1961; Kass and Wasserman, 1996). The main motivation for the Jeffreys rule prior is that statistical inference should not depend on any specific parametrization of the model which could often be rather arbitrary. For instance, if the network autocorrelation model is rewritten in terms of a precision parameter $\omega := 1/\sigma^2$, rather than σ^2, applying Jeffreys rule prior to the model formulated with respect to ω or σ^2 results in the same posterior conclusions about the network effect. Hence, when using Jeffreys rule prior, there is no need to determine a privileged parametrization as the prior is parametrization-invariant. Formally, Jeffreys rule prior is defined as

$$p_{\mathrm{J}}(\boldsymbol{\theta}) \propto \sqrt{\det\left(I(\boldsymbol{\theta})\right)},$$

where $I(\boldsymbol{\theta})$ denotes the model's Fisher information matrix (Doreian, 1981). The exact analytical form of the prior is given in Theorem 1. Since the Jeffreys rule prior for the network autocorrelation model is improper, the propriety of the resulting posterior needs to be checked and is verified in Corollary 2.

Theorem 1. *Consider the network autocorrelation model given in* (1) *and assume that* $\left(X^T X\right)^{-1}$ *exists. Then, the Jeffreys rule prior for* $\boldsymbol{\theta} = \left(\rho, \sigma^2, \boldsymbol{\beta}\right)$*, denoted by* $p_J(\boldsymbol{\theta})$*, is*

$$p_J(\boldsymbol{\theta}) \propto \left(\sigma^2\right)^{-\frac{k+2}{2}} \left\{ \operatorname{tr}\left(B_\rho^T B_\rho\right) + \operatorname{tr}\left(B_\rho^2\right) + \frac{1}{\sigma^2} \boldsymbol{\beta}^T X^T B_\rho^T M B_\rho X \boldsymbol{\beta} - \frac{2}{g} \operatorname{tr}^2\left(B_\rho\right) \right\}^{\frac{1}{2}}, \qquad (5)$$

where $B_\rho := W A_\rho^{-1}$*.*

Proof. See Appendix.

Corollary 2. *Consider the network autocorrelation model given in* (1) *and assume that* $\left(X^T X\right)^{-1}$ *exists. Then,*

(i) *The Jeffreys rule prior* $p_J(\boldsymbol{\theta})$ *is unbounded and not integrable on* $\Omega = \left(\lambda_g^{-1}, \lambda_1^{-1}\right) \times (0, \infty) \times \mathbb{R}^k$*.*

(ii) *The Jeffreys rule posterior* $p_J(\boldsymbol{\theta}|\boldsymbol{y})$ *is proper on* $\Omega = \left(\lambda_g^{-1}, \lambda_1^{-1}\right) \times (0, \infty) \times \mathbb{R}^k$ *when* $\left(\boldsymbol{y}^T M W \boldsymbol{y}\right)^2 \neq \boldsymbol{y}^T W^T M W \boldsymbol{y} \boldsymbol{y}^T M \boldsymbol{y}$*.*

Proof. See Appendix.

4.3. Independence Jeffreys prior

Jeffreys rule prior has the desirable property to be invariant under one-to-one parameter transformations and most often results in reasonable priors for one-dimensional problems. However, applying the rule in multiparameter models may result in poor frequentist properties of Bayesian inferences (Berger et al., 2001; De Oliveira, 2010; De Oliveira and Song, 2008) or even improper posteriors (Berger et al., 2001; Bolstad, 2009; Rubio and Steel, 2014). Thus, already Jeffreys himself argued that it is often better to consider certain blocks of parameters as a priori "independent" from each other and to compute the marginal prior for each parameter block using Jeffreys rule, assuming the other parameters to be known (De Oliveira and Song, 2008). The resulting product of the marginal priors is then called Independence Jeffreys prior. Following Bayesian analyses of related autoregressive and spatial models (Berger et al., 2001; De Oliveira, 2010, 2012; De Oliveira and Song, 2008), we split the network autocorrelation model's parameters into the two blocks $\left(\rho, \sigma^2\right)$ and $\boldsymbol{\beta}$ and we derive the Independence Jeffreys prior based on this partitioning of the model parameters. We give the prior's analytical form in Theorem 2 and provide its main theoretical properties in Corollary 3.

Theorem 2. *Consider the network autocorrelation model given in* (1)*. Then, the Independence Jeffreys prior for* $\boldsymbol{\theta} = \left(\rho, \sigma^2, \boldsymbol{\beta}\right)$*, denoted by* $p_{IJ}(\boldsymbol{\theta})$*, is*

$$p_{IJ}(\boldsymbol{\theta}) \propto \frac{1}{\sigma^2} \left\{ \operatorname{tr}\left(B_\rho^T B_\rho\right) + \operatorname{tr}\left(B_\rho^2\right) + \frac{1}{\sigma^2} \boldsymbol{\beta}^T X^T B_\rho^T B_\rho X \boldsymbol{\beta} - \frac{2}{g} \operatorname{tr}^2\left(B_\rho\right) \right\}^{\frac{1}{2}}.$$

Proof. See Appendix.

Corollary 3. *Consider the network autocorrelation model given in* (1)*. Then,*

(i) *The Independence Jeffreys prior* $p_{IJ}(\boldsymbol{\theta})$ *is unbounded and not integrable on* $\Omega = \left(\lambda_g^{-1}, \lambda_1^{-1}\right) \times (0, \infty) \times \mathbb{R}^k$*.*

(ii) *The Independence Jeffreys posterior* $p_{IJ}(\boldsymbol{\theta}|\boldsymbol{y})$ *is proper on* $\Omega = \left(\lambda_g^{-1}, \lambda_1^{-1}\right) \times (0, \infty) \times \mathbb{R}^k$ *when* $g > k$*,* $\left(X^T X\right)^{-1}$ *exists, and* $\left(\boldsymbol{y}^T M W \boldsymbol{y}\right)^2 \neq \boldsymbol{y}^T W^T M W \boldsymbol{y} \boldsymbol{y}^T M \boldsymbol{y}$*.*

Proof. See Appendix.

The analytical expression of the Independence Jeffreys prior is similar, but slightly simpler, to the one of Jeffreys rule prior in (5). The major difference between the two is that for the Jeffreys rule prior the exponent of the error variance depends on the number of covariates, k, while it does not for the Independence Jeffreys prior. For related models (Berger et al., 2001; De Oliveira, 2012; De Oliveira and Song, 2008), it has been shown that having k in the exponent of σ^2, as in the Jeffreys rule prior, could result in an underestimation of the error variance. We will therefore investigate whether this is also the case in the network autocorrelation model, and if this underestimation occurs whether it can be circumvented by using the Independence Jeffreys prior. Hence, while the Jeffreys rule prior is based on an invariance principle, the Independence Jeffreys prior is a heuristic modification of Jeffreys rule prior that can result in better inferences.

Table 1: Characteristics of the studies used for the specification of the empirical informative prior for ρ.

	Study	Field	g	Type of W	Method	ρ
1	Andersson et al. (2010)	Property prices	1,034	Inverse distance	ML	.52
2	Anselin (1984)	House values	49	First-order contiguity	ML	.28
3	Anselin (1990)	Wage rates	25	First-order contiguity	ML	-.62
4	Anselin and Le Gallo (2006)	House prices	115,732	First-order contiguity	ML	.44
5	Anselin and Lozano-Gracia (2008)	House prices	103,867	First-order contiguity	2SLS	.33
6	Anselin et al. (2010)	House rents	1,671	First-order contiguity	HAC	.24
7	Anselin et al. (2000)	Innovation transfer	89	Distance-based contiguity	2SLS	.23
8.1	Arbia and Basile (2005)	GDP growth rates	92	First-order contiguity	ML	.33
8.2						.18
8.3						.34
9	Armstrong and Rodriguez (2006)	Property values	1,860	Inverse distance	ML	.36
10.1	Baller et al. (2001)	Homicide rates	1,412	10 nearest neighbors	IV	.71
10.2						.65
10.3						.18
10.4						.23
11.1	Bernat (1996)	Economic growth	49	Squared inverse distance	ML	.35
11.2						.42
11.3						.70
12.1	Bivand and Szymanski (2000)	Garbage collection costs	324	First-order contiguity	ML	.15
12.2						.10
13	Bordignon et al. (2003)	Tax rates	143	First-order contiguity	ML	.16
14.1	Brueckner and Saavedra (2001)	Tax rates	70	Population weights	ML	.16
14.2						.04
14.3						.26
15.1	Buonanno et al. (2009)	Crime patterns	103	Inverse traveling distance	2SLS	-.54
15.2						.19
15.3						.21
16.1	Burt and Doreian (1982)	Scientific publishing	52	Structural equivalence	ML	.26
16.2						.21
16.3						.25
16.4						.45
16.5						.29
16.6						.31
16.7						.26

Note: g = Network size. 2SLS = Two-stage least squares. 2SLS robust = Heteroskedastic two-stage least squares. 3SLS = Three-stage least squares. HAC = Kelejian-Prucha heteroskedasticity and autocorrelation consistent estimator. IV = Instrumental variables. ML = Maximum likelihood.

Table 1: Characteristics of the studies used for the specification of the empirical informative prior for ρ.

	Study	Field	g	Type of W	Method	ρ
16.8						.54
17	Can (1992)	House prices	563	Squared inverse distance	ML	.41
18	Carruthers and Clark (2010)	House prices	28,165	4 nearest neighbors	2SLS	.17
19.1	Chang (2008)	Water quality	94	First-order contiguity	ML	.19
19.2						.14
19.3						.49
19.4						.48
19.5						.56
19.6						.15
19.7						.42
19.8						.43
19.9						.37
19.10						.56
19.11						.44
19.12						.41
19.13						.55
19.14						.47
19.15						.36
19.16						.24
19.17						.35
19.18						.29
19.19						.25
19.20						.28
19.21						.24
19.22						.50
19.23						.42
19.24						.51
19.25						.47
20	Cohen and Coughlin (2008)	House prices	508	Inverse distance	ML	.26
21	Conway et al. (2010)	House prices	260	First-order contiguity	ML	.11
22	Dallerba (2005)	GDP growth rates	48	5 most accessible neighbors	ML	.40
23	Doreian (1980)	Huk rebellion	57	First-order contiguity	ML	.47
24.1	Doreian (1980)	Louisiana voting behavior	64	First-order contiguity	ML	.61
24.2						.26

Note: g = Network size. 2SLS = Two-stage least squares. 2SLS robust = Heteroskedastic two-stage least squares. 3SLS = Three-stage least squares. HAC = Kelejian-Prucha heteroskedasticity and autocorrelation consistent estimator. IV = Instrumental variables. ML = Maximum likelihood.

Table 1: Characteristics of the studies used for the specification of the empirical informative prior for ρ.

	Study	Field	g	Type of W	Method	ρ
24.3	Doreian (1981)					.12
24.4						.29
24.5	Leenders (2002)					.31
25	Dow (2007)	Subsistence contributions	158	Lexical distance	2SLS	.76
26	Easterly and Levine (1998)	GDP growth rates	234	Neighbor's total GDP	2SLS	.55
27	Elhorst (2014)	Crime rates	49	First-order contiguity	ML	.43
28	Ertur et al. (2007)	GDP growth rates	138	10 nearest neighbors	ML	.75
29.1	Fingleton (2001)	Productivity growth rates	178	Economic size and distance	3SLS	-.19
29.2						.56
29.3						.73
30	Fingleton et al. (2005)	Change in employment	408	Squared inverse distance	2SLS	.41
31	Fingleton and Le Gallo (2008)	House prices	353	Economic distance	ML	.72
32	Florax et al. (2002)	Agricultural yields	100	First-order contiguity	ML	.50
33	Ford and Rork (2010)	Patent rates	186	First-order contiguity	ML	.08
34	Fornango (2010)	Homicide rates	110	First-order contiguity	ML	.30
35	Gimpel and Schuknecht (2003)	Voting turnout	363	Distance-based	ML	.67
36.1	Gould (1991)	Networks in the Paris commune	20	Crossdistrict enlistment	ML	.29
36.2						.49
36.3						.49
37	Greenbaum (2002)	Teacher's salaries	483	Inverse difference in income	ML	.66
38	Heikkila and Kantiotou (1992)	Police expenditures	57	First-order contiguity	ML	.43
39.1	Heyndels and Vuchelen (1998)	Tax rates	589	First-order contiguity	3SLS	.67
39.2						.70
40	Holloway et al. (2002)	Crop adoption	406	First-order contiguity	Bayes Probit	.54
41	Hunt et al. (2005)	Fishing trip prices	770	Inverse distance-based	ML	.80
42.1	Joines et al. (2003)	Hospitalization rates	100	First-order contiguity	ML	.53
42.2						.51
43	Kalenkoski and Lacombe (2008)	Youth employment	3,065	First-order contiguity	ML	.49
44.1	Kalnins (2003)	Fast food prices	1,385	Distance & contiguity-based	ML	.11
44.2						.21
45.1	Kim and Goldsmith (2009)	Property values	262	3 nearest neighbors	2SLS robust	.22
45.2			523			.19
45.3			730			.14
46	Kim and Zhang (2005)	Land values	731	Nearest neighbor	ML	.39

Note: g = Network size. 2SLS = Two-stage least squares. 2SLS robust = Heteroskedastic two-stage least squares. 3SLS = Three-stage least squares. HAC = Kelejian-Prucha heteroskedasticity and autocorrelation consistent estimator. IV = Instrumental variables. ML = Maximum likelihood.

Table 1: Characteristics of the studies used for the specification of the empirical informative prior for ρ.

	Study	Field	g	Type of W	Method	ρ
47.1	Kirk and Papachristos (2011)	Homicide rates	342	First-order contiguity	ML	.43
47.2						.33
48.1	Land et al. (1991)	Church adherence rates	731	Inverse distance	2SLS	.33
48.2			697			.29
48.3			663			.28
49	Lauridsen et al. (2010)	Pharmaceutical expenditures	400	Inverse distance	ML	.87
50	LeSage (1997b)	House values	88	First-order contiguity	ML	.45
51	Levine et al. (1995)	Road accidents	362	Squared inverse distance	ML	.22
52.1	Lin (2010)	GPA scores	68,131	Friendship	ML	.30
52.2			49,559			.29
52.3			79,067			.30
53	Lu and Zhang (2011)	Tree heights	3,982	Variogram	ML	.59
54.1	McPherson and Nieswiadomy (2005)	Species threat	113	Shared border length	ML	.23
54.2						.16
55	McMillen (2010)	Land ratios	1,322	First-order contiguity	ML	.71
56.1	McMillen et al. (2007)	Tuition fees	929	Distance & contiguity-based	ML	.22
56.2						.34
57	Moreno and Trehan (1997)	Worker output growth	89	Inverse distance	ML	.51
58.1	Morenoff (2003)	Birth weights	342	First-order contiguity	2SLS	.53
58.2						.69
59.1	Mur et al. (2008)	Purchasing power parity	1,274	Nearest neighbor & distance-based	ML	.60
59.2						.61
60	Niebuhr (2010)	R&D spillovers	95	First-order contiguity	ML	.16
61.1	Osland (2010)	Voting patterns	1,691	Nearest neighbor	ML	.07
61.2			766			.06
62	Patton and McErlean (2003)	Land prices	197	Squared inverse distance	IV	.66
63	Plümper and Neumayer (2010)	Tax rates	581	First-order contiguity	ML	.12
64.1	Pons-Novell and Elisabet (1999)	GDP growth rates	74	First-order contiguity	ML	.23
64.2						.20
64.3						.17
65	Revelli (2003)	Expenditure levels	238	Contiguity-based	ML	.11
66	Ruggles (2007)	Intergenerational coresidence	276	Shared border	ML	.15
67	Rupasingha et al. (2002)	Income growth	2,995	First-order contiguity	ML	.49
68.1	Saavedra (2000)	Welfare competition	47	First-order contiguity	ML	.28

Note: g = Network size. 2SLS = Two-stage least squares. 2SLS robust = Heteroskedastic two-stage least squares. 3SLS = Three-stage least squares. HAC = Kelejian-Prucha heteroskedasticity and autocorrelation consistent estimator. IV = Instrumental variables. ML = Maximum likelihood.

Table 1: Characteristics of the studies used for the specification of the empirical informative prior for ρ.

	Study	Field	g	Type of W	Method	ρ
68.2						.30
68.3						.32
69	Seldadyo et al. (2010)	Governance patterns	188	10 nearest neighbors	ML	.28
70	Shin and Ward (1999)	Military spending	95	Distance & contiguity-based	ML	.08
71.1	Tam Cho (2003)	Campaign donations	671	Inverse distance	2SLS	.06
71.2			455		ML	.04
71.3			657		ML	.03
71.4			1,183		ML	.03
71.5			1,420		2SLS	.03
71.6			2,072		2SLS	.03
71.7			1,821		2SLS	.03
71.8			2,288		2SLS	.02
71.9			2,206		2SLS	.03
71.10			291		ML	.07
71.11			229		ML	.06
71.12			249		ML	.06
71.13			273		ML	.05
71.14			458		2SLS	.05
71.15			502		2SLS	.05
71.16			698		2SLS	.05
71.17			606		2SLS	.04
71.18			660		2SLS	.05
71.19			752		2SLS	.03
71.20			401		2SLS	.00
71.21			613		2SLS	.02
71.22			581		2SLS	.02
71.23			324		ML	.05
71.24			918		ML	.01
71.25			760		2SLS	.03
71.26			701		ML	.06
71.27			980		2SLS	.05
71.28			874		ML	.07
72	Tita and Greenbaum (2009)	Gun violence	244	Gang rivalry	ML	.22
73	Varga (2000)	High technology innovations	125	Distance-based contiguity	IV	.14

Note: g = Network size. 2SLS = Two-stage least squares. 2SLS robust = Heteroskedastic two-stage least squares. 3SLS = Three-stage least squares. HAC = Kelejian-Prucha heteroskedasticity and autocorrelation consistent estimator. IV = Instrumental variables. ML = Maximum likelihood.

Table 1: Characteristics of the studies used for the specification of the empirical informative prior for ρ.

	Study	Field	g	Type of W	Method	ρ
74	Vega and Elhorst (2015)	Cigarette sales	1,380	First-order contiguity	ML	.20
75	Vitale et al. (2016)	Student performance	66	Personal advice	ML	.31
76	Voss and Chi (2006)	Population change	1,837	7 nearest neighbors	ML	.27
77.1	Voss et al. (2006)	Child poverty	3,136	First-order contiguity	ML	.31
77.2						.27
78	Wilhelmsson (2002)	House prices	1,377	Inverse distance	ML	.95
79.1	Whitt (2010)	Crime rates	85	First-order contiguity	ML	.37
79.2						.58
79.3						.50
79.4						.54
80	Won Kim et al. (2003)	House prices	609	Distance & contiguity-based	2SLS robust	.55
81.1	Yang et al. (2012)	Wine prices	79	3 nearest neighbors	ML	.33
81.2			876	35 nearest neighbors		.34

Note: g = Network size. 2SLS = Two-stage least squares. 2SLS robust = Heteroskedastic two-stage least squares. 3SLS = Three-stage least squares. HAC = Kelejian-Prucha heteroskedasticity and autocorrelation consistent estimator. IV = Instrumental variables. ML = Maximum likelihood.

4.4. An informative prior for ρ

Having discussed three prominent non-informative priors above, in this section, we derive a population distribution for ρ based on an extensive literature review of empirical applications of the network autocorrelation model. Subsequently, this population distribution is used as an informative prior for ρ.

In our literature search, we considered 81 peer-reviewed papers and a total of 183 estimated network autocorrelations ρ. The most important characteristics of the sample are summarized in Table 1.[3] As network effects from one paper are usually from closely related fields, this suggests that these network autocorrelations are more similar than network effects coming from different studies. To take this into account, we rely on a hierarchical approach and use the following multilevel model (Gelman et al., 2003) to estimate the population distribution of the network effects:

$$\begin{aligned} \text{Level 1: } & \rho_{ij} \sim N\left(\rho_j, \sigma_\rho^2\right), \\ \text{Level 2: } & \rho_j \sim N\left(\mu_\rho, \tau_\rho^2\right), \end{aligned} \tag{6}$$

where $N(\cdot, \cdot)$ denotes the normal distribution, $i \in \{1, ..., n_j\}$, $j \in \{1, ..., J\}$, ρ_{ij} is the observed i-th network effect from field j, and $\{\rho_j\}_j, \mu_\rho, \sigma_\rho^2$, and τ_ρ^2 are model parameters which have to be estimated. The distribution in Level 1 corresponds to the empirical distribution of a network effect in a specific field. The distribution in Level 2 denotes the overall population distribution in which we are ultimately interested in. We fitted the model in a Bayesian framework in R using standard non-informative uniform priors for μ_ρ, τ_ρ, and $\log\left(\sigma_\rho^2\right)$ (Gelman, 2006). This resulted in posterior mean estimates of $\mu_\rho = .36$ and $\tau_\rho = .19$.[4] The resulting informative prior for ρ, $p(\rho) = N\left(\mu_\rho, \tau_\rho^2\right)$, along with the histogram of the average network effects from each field, is plotted in Figure 1. As can be seen, the multilevel model in (6) provides a reasonably good fit to the empirical data.

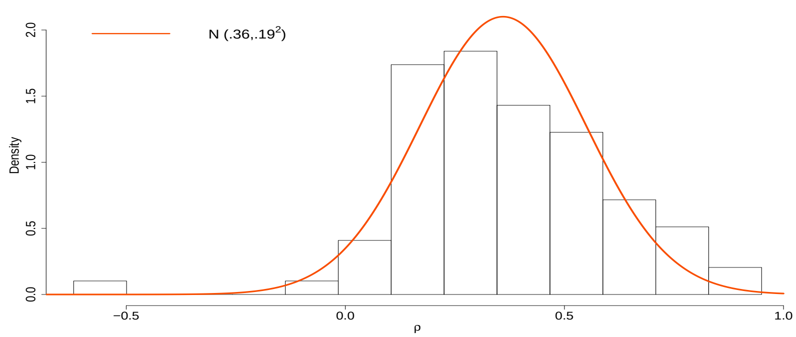

Figure 1: Histogram of the average network effects in each field $\{\bar{\rho}_j\}_j$, $\bar{\rho}_j := \sum_{i=1}^{n_j} \rho_{ij}/n_j$, and probability density function of the fitted normal population distribution for ρ.

[3]We did not attempt to be fully comprehensive here and do not claim to have included all available literature on empirical applications of the network autocorrelation model. Our selection features work that (i) uses row-standardized connectivity matrices, (ii) specifies the network size and the type of connectivity matrix, and (iii) employs appropriate estimation techniques for the given type of data.

[4]The associated 95%-credible intervals for μ_ρ and τ_ρ were (.33, .39) and (.16, .22), respectively.

Figure 1 shows that there are substantially more reported positive network effects than negative ones in the literature. This finding conflicts with a flat prior for ρ on $\left(\lambda_g^{-1}, \lambda_1^{-1}\right)$ which typically implies that negative network effects are a priori more likely than positive network effects and is clearly unrealistic.[5]

We combine this empirical informative prior for ρ with the standard non-informative priors for $\left(\boldsymbol{\beta}, \sigma^2\right)$ from Section 4.1, assuming all parameters to be a priori independent.[6] Hence, the resulting empirical informative joint prior for $\boldsymbol{\theta}$ is

$$p_{\text{EI}}\left(\boldsymbol{\theta}\right) = p_{\text{EI}}\left(\rho\right) \times p_{\text{F}}\left(\sigma^2\right) \times p_{\text{F}}\left(\boldsymbol{\beta}\right) \propto N_\rho\left(.36, .19^2\right) \times 1/\sigma^2.$$

4.5. A weakly informative prior for ρ

There may be cases where a researcher does not expect a network effect to be present in the data, or it may be that the researcher does not (want to) entertain the prior belief that the level of autocorrelation in a dataset is likely to fit with the empirical literature at large. In these cases, a researcher might purposely prefer to use less prior knowledge than actually available in the literature and rely on a so-called weakly informative prior distribution (Gelman et al., 2003). We construct such a weakly informative prior for ρ by imposing a normal prior that is centered around .36, as is the empirical informative prior, but with a deliberately much larger standard deviation, accounting for the uncertainty in one's prior beliefs. We set the weakly informative prior's standard deviation to .7, compared to .19 for the empirical informative prior, yielding a broad and fairly flat prior that still results in at least 62% of prior probability mass being contained in the unit interval $(0, 1)$. As for the empirical informative prior, we impose standard non-informative priors for $\left(\boldsymbol{\beta}, \sigma^2\right)$, assuming all parameters to be a priori independent. Thus,

$$p_{\text{WI}}\left(\boldsymbol{\theta}\right) = p_{\text{WI}}\left(\rho\right) \times p_{\text{F}}\left(\sigma^2\right) \times p_{\text{F}}\left(\boldsymbol{\beta}\right) \propto N_\rho\left(.36, .7^2\right) \times 1/\sigma^2.$$

4.6. Graphical prior comparison

In order to get more insight into the differences between the discussed priors, we inspect them graphically. We base our visualization on a randomly generated network of 20 actors with four covariates, including an intercept term. The shape of these priors is essentially the same for other data sets that are generated under different specifications of W and X (not shown).

Figure 2 shows the flat, the conditional Jeffreys rule, the conditional Independence Jeffreys, the empirical informative prior, and the weakly informative prior for ρ for the simulated data set. We fixed $\boldsymbol{\beta}$ and σ^2 to (1,1,1,1) and 1 for both versions of Jeffreys prior as the respective marginal priors for ρ are analytically not available. The graphs of the two versions of Jeffreys prior are "bathtub-shaped", contrary to the flat and the informative priors for ρ. In particular, $p_{\text{IJ}}\left(\rho|\sigma^2, \boldsymbol{\beta}\right)$ assigns substantial weight to values of ρ close to the boundaries of the admissible interval for ρ,

[5] For row-standardized connectivity matrices it holds that $\lambda_g^{-1} \leq -1$ (Stewart (2009), Property 10.1.2), and for most of the simulated data sets we considered, we observed that $\lambda_g^{-1} < -1$. Thus, as $\lambda_1^{-1} = 1$, in these cases the flat prior assigns more probability mass to negative network effects than to positive ones.

[6] Propriety of the resulting posterior distribution, under the conditions given in Corollary 1, follows immediately from the corollary's proof. Our informative prior for ρ can be easily combined with informative priors for $\boldsymbol{\beta}$ and σ^2 as well. We use non-informative improper priors for the latter parameters because our main focus lies on estimating the network effect ρ.

while $p_{\mathrm{J}}\left(\rho|\sigma^2,\boldsymbol{\beta}\right)$ does essentially the same but with slightly more weight for values of ρ close to the left boundary and less prior mass for values of ρ close to the right boundary.[7]

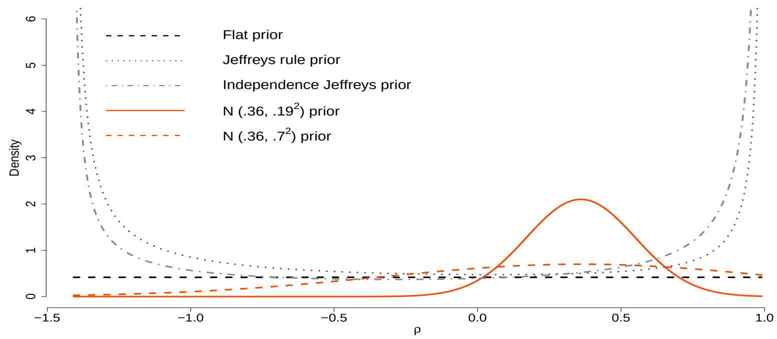

Figure 2: Conditional prior distributions for $\left(\rho|\boldsymbol{\beta}=(1,1,1,1),\sigma^2=1\right)$ for simulated data.

As the main analytical difference between the Jeffreys rule and the Independence Jeffreys prior is that for the latter the exponent of the error variance does not depend on the number of covariates, we also considered the bivariate conditional density for $\left(\rho,\sigma^2|\boldsymbol{\beta}=(1,1,1,1)\right)$. In contrast to the conditional prior for ρ, $p_{\mathrm{J}}\left(\rho,\sigma^2|\boldsymbol{\beta}\right)$ places more prior mass at boundary values of the two-dimensional parameter space, $\left(\lambda_g^{-1},\lambda_1^{-1}\right)\times(0,\infty)$, compared to $p_{\mathrm{IJ}}\left(\rho,\sigma^2|\boldsymbol{\beta}\right)$ (not shown). Thus, we expect the Bayesian posterior estimates for ρ and σ^2 based on Jeffreys rule prior to tend more towards their respective boundary values than the ones based on the Independence Jeffreys prior.

5. Bayesian computation

In this section, we present an efficient algorithm to perform a Bayesian estimation of the network autocorrelation model. The methodology can be used to sample from the various arising posterior distributions based on the priors discussed in Section 4. As is common in Bayesian computation, the goal is to obtain a sample from the joint posterior of the unknown model parameters by sequentially drawing from the conditional posterior distributions, i.e. given the remaining parameters and the data (Gelfand and Smith, 1990; Geman and Geman, 1984). This is repeated until a sufficiently large sample is obtained.[8] We propose to sample the parameters according to the following blocks: (ρ,β_1), $\widetilde{\boldsymbol{\beta}}$, and σ^2, where β_1 denotes the model's intercept and $\widetilde{\boldsymbol{\beta}}=(\beta_2,...,\beta_k)$ contains the regression coefficients of the covariates. The reason for simultaneously sampling ρ and β_1 in one block

[7]The (inverse of the) eigenvalues of the simulated network yield $(-1.42,1)$ as the admissible interval for ρ as defined in Section 2. As referred to above, the shape of these priors does not rely on this specific interval and is similar under data sets generated under different specications of W and X.

[8]This approach is needed as for none of the priors previously discussed the corresponding posterior belongs to a family of known probability distributions. Gelfand and Smith (1990) showed that sampling from the sequence of conditional posteriors for all parameters indeed produces estimates that converge in the limit to the true (joint) posterior distribution of the parameters.

is the high posterior correlation between these parameters.[9] Sampling these parameters separately would result in *slow mixing*, i.e. more draws would be needed to get both a good approximation of the posterior distribution and small estimation errors (Brooks, 1998; Gelman et al., 2003; Raftery and Lewis, 1996).

We illustrate the sampling algorithm when relying on the flat and the informative priors first, before discussing sampling from the more complex posteriors based on Jeffreys rule and Independence Jeffreys prior. For the former, the conditional posteriors are given by (LeSage, 1997a)

$$p\left((\rho, \beta_1)\,|\sigma^2, \widetilde{\boldsymbol{\beta}}, \boldsymbol{y}\right) \propto |A_\rho| \exp\left(-\frac{1}{2\sigma^2}\left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)^T \left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)\right) p\left(\rho\right), \tag{7}$$

$$p\left(\sigma^2|\,(\rho, \beta_1)\,, \widetilde{\boldsymbol{\beta}}, \boldsymbol{y}\right) \sim IG\left(\frac{g}{2}, \frac{\left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)^T \left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)}{2}\right), \tag{8}$$

$$p\left(\widetilde{\boldsymbol{\beta}}|\,(\rho, \beta_1)\,, \sigma^2, \boldsymbol{y}\right) \sim N\left(\boldsymbol{\mu}_{\widetilde{\boldsymbol{\beta}}}, \Sigma_{\widetilde{\boldsymbol{\beta}}}\right), \tag{9}$$

where $IG\left(\cdot, \cdot\right)$ denotes the inverse gamma distribution, and $\boldsymbol{\mu}_{\widetilde{\boldsymbol{\beta}}}$ and $\Sigma_{\widetilde{\boldsymbol{\beta}}}$ are given in the Appendix.

Sampling from the inverse gamma and the normal distribution in (8) and (9) is straightforward, whereas due to the appearance of the determinant in (7), the conditional posterior of (ρ, β_1) does not have a well-known form. In order to efficiently sample from this distribution, we rely on the *Metropolis-Hastings algorithm* (Hastings, 1970; Metropolis et al., 1953). In this algorithm, a candidate-generating distribution is chosen from which candidate values for the target distribution - here the conditional posterior - are drawn. The specification of the candidate-generating density is crucial for the algorithm's efficiency, where we aim at constructing a density that closely approximates the actual conditional posterior target distribution and typically results in efficient solutions (Chib and Greenberg, 1995, 1998).

In (7), we approximate $\log\left(|A_\rho|\right)$ at $\rho = 0$ by a second-order Taylor polynomial which results in a normal approximation of the first factor, $|A_\rho|$. The second factor, $\exp\left(\cdot\right)$, if considered as a function of (ρ, β_1), has a bivariate normal kernel. The third factor, i.e. the marginal prior for ρ, is ignored in the candidate-generating density when using the flat prior and is normal for the informative priors. Hence, the overall product of these normal distributions results in a bivariate normal candidate-generating distribution for (ρ, β_1) which incorporates the dependence between the two parameters and is tailored to the conditional posterior of (ρ, β_1).

Due to the complex prior expressions for both Jeffreys rule and Independence Jeffreys prior, a Metropolis-Hastings step is needed to sample from all three conditional posteriors when employing these priors. For the first parameter block, (ρ, β_1), we use the same candidate-generating distribution as for the flat prior, as the prior information for (ρ, β_1) is quite vague compared to the likelihood. For the conditional posterior of σ^2, we propose inverse gamma distributions as candidate-generating distributions but with different shape parameters than those used in (8), accounting for the different exponents for σ^2 in the two priors. Finally, we rely on the normal distribution in (9) as the candidate-generating distribution for the conditional posterior of $\widetilde{\boldsymbol{\beta}}$. All details and the full sampling schemes for all of the discussed priors can be found in the Appendix.

We implemented our approach and compared its performance to existing sampling schemes

[9]This correlation is particularly pronounced for high network densities, and we have not found this issue being discussed in the literature before. Only Hepple (1995b) provides a plot of the bivariate marginal posterior density $p_{\mathrm{F}}\left((\rho, \beta_1)\,|\boldsymbol{y}\right)$ for a real data set that clearly shows this dependence.

which do not block (ρ, β_1) but rely on a one-dimensional random walk algorithm to generate draws for ρ instead (Holloway et al., 2002; LeSage, 2000; LeSage and Pace, 2009). We found that our proposed method produces well-mixed Markov chains with very low autocorrelations. Figure 3 displays sample trace plots of posterior draws for ρ based on our algorithm (left panel) and based on existing schemes (right panel) when using the flat prior from Section 4.1. As can be seen, our algorithm generates Markov chains that are moving quicker and explore the parameter space much faster compared to traditional methods.[10]

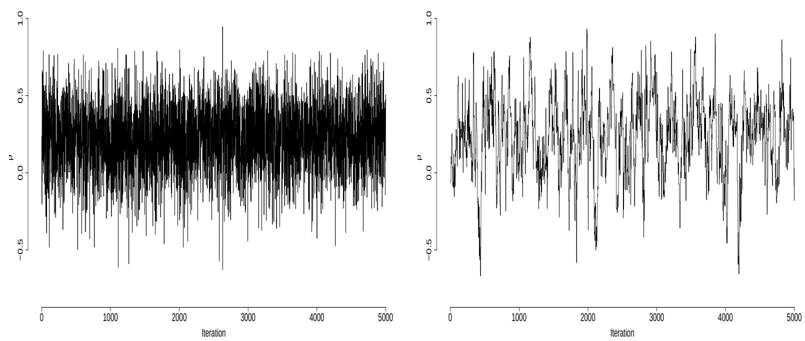

Figure 3: Trace plots of posterior draws of ρ for novel (left) and random walk scheme (right) for simulated data.

6. Simulation Study

We performed a thorough simulation study to examine properties of the Bayesian estimators based on the flat, the Jeffreys rule, the Independence Jeffreys, the empirical informative, and the weakly informative prior, and compared them to those based on maximum likelihood estimation. The main focus in this study is to evaluate the bias of ρ and the frequentist coverage of credible and confidence intervals for ρ for the different estimators, i.e. the extent to which the true network effect is contained in the credible and confidence intervals. Furthermore, as the most likely outcome of the negative bias in the estimation of ρ is a Type II error, we report the average of Type I and Type II errors as well.[11] Such average error rates are increasingly used as optimal decision criteria instead of the prevailing paradigm which is fixing Type I error probability and then minimizing Type II error probability (Chance and Rossman, 2006; DeGroot and Schervish, 2010; Pericchi and Pereira, 2016). Lastly, we also investigated the estimation of σ^2 as it is known that Jeffreys rule prior can result in poor estimates of the error variance in multi-parameter models (De Oliveira, 2012; De Oliveira and Song, 2008).

6.1. Study design

In our study design, we largely followed setups from previous simulation studies of the network autocorrelation model (Mizruchi and Neuman, 2008; Neuman and Mizruchi, 2010; Wang et al.,

[10]Also note that there are no parameters to be tuned in the Metropolis-Hastings steps in our approach, such as the variance of a candidate-generating distribution. This stands in stark contrast to existing schemes where this is commonly done in order to achieve specific acceptance rates.

[11]We thank an anonymous reviewer for this suggestion.

2014). Hence, we generated data $\boldsymbol{y}$ by using random networks and varying the density of the network, the size of the network, the magnitude of ρ, and the number of covariates. We did so by $\boldsymbol{y} = (I_g - \rho W)^{-1} (X\boldsymbol{\beta} + \boldsymbol{\varepsilon})$.[12] We considered six levels of network densities ($d = .1, .2, .3, .4, .5, .6$), three different network sizes ($g = 10, 20, 50$), three fixed levels of network effect sizes ($\rho = 0, .2, .5$), and two sets of covariates ($k = 3, 6$) plus an intercept term.[13] We obtained random, binary connectivity matrices with zeros in the diagonal entries relying on the "rgraph" function from the sna package in R (Butts, 2008) and subsequently row-normalized the raw connectivity matrices. Moreover, we drew independently values from a standard normal distribution for the elements of X (excluding the first column of X, which is a vector of ones), $\boldsymbol{\beta}$, and $\boldsymbol{\varepsilon}$, so $\sigma^2 = 1$. In addition to simulating data using a fixed network autocorrelation ρ, we also allowed for fluctuations in the underlying network effects by sampling them from the estimated population distribution from Section 4.4. As the true network autocorrelation is unknown in practice, this appears to be a much more realistic simulation setup compared to setting ρ to a specific value a priori.[14] In total, we considered 120 scenarios and ran 500 replications for each data set and estimator.

For the Bayesian estimators, we drew 10,000 samples from the corresponding joint posteriors, applying the sampling schemes described in Section 5. We used the marginal posterior median as point estimator and 95% equal-tailed credible intervals by discarding the 2.5% smallest and largest draws, respectively, for coverage analyses of ρ and σ^2. In contrast to that, we employed asymptotic standard errors based on the model's observed information matrix to obtain maximum likelihood-based confidence intervals for ρ and σ^2.[15]

[12] For all the simulated data sets we looked at, none of the regularity conditions needed for posterior propriety was violated, and it seems highly unlikely to encounter such a situation in a real-life empirical situation.

[13] Simulation results for negative values for ρ are available from the authors upon request. We do not present them here because (i) our literature review showed that such scenarios are almost never observed in practice, and (ii) the analyses provide no additional, i.e. different, insights.

[14] In fact, we sampled ρ from the estimated population distribution truncated to $(-1, 1)$ to ensure that the generated network effects always lie in the chosen admissible interval $\left(\lambda_g^{-1}, 1\right)$. Note that less than 0.1% of probability mass of the estimated population distribution actually falls outside $(-1, 1)$. For each draw for ρ from this estimated population distribution, we recorded the drawn value for ρ (which is the true value for ρ for that particular draw) amd base our simulation results, i.e. bias and coverage rates, on those recorded, true underlying network effects.

[15] All computation was performed in R using self-written routines for all estimators. We used maximum likelihood estimates as starting values for the MCMC procedures and discarded the first 1,000 iterations as so-called *burn-in* values (Gelman et al., 2003). We opted for the posterior median as point estimator as most of the marginal posterior densities of ρ and σ^2 are skewed, so the posterior median is a less extreme estimator than the posterior mean or posterior mode.

Table 2: Average bias of ρ based on using the flat prior (F), Jeffreys rule prior (J), Independence Jeffreys prior (IJ), the empirical informative prior (EI), the weakly informative prior (WI), and the maximum likelihood estimator (ML). The average bias was computed by taking the average of the difference between the estimated network effect and the true ρ values used for data-generation over 500 replications. For each scenario, the best estimate is printed in bold face.

$g=10, k=3$			$\rho=0$						$\rho=.2$						$\rho=.5$						$\rho \sim N\left(.36, .19^2\right)(-1,1)$			
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	-.028	-.046	-.034	.240	**.027**	-.041	-.103	-.074	-.085	.076	**-.064**	-.079	-.151	**-.068**	-.104	-.109	-.136	-.082	-.124	-.075	-.095	**-.020**	-.099	-.083
.2	-.032	-.040	-.032	.237	**.024**	-.038	-.120	-.095	-.105	**.075**	-.077	-.097	-.193	**-.105**	-.150	-.128	-.174	-.118	-.160	-.105	-.133	**-.038**	-.131	-.112
.3	-.088	-.115	-.093	.247	**-.008**	-.110	-.158	-.137	-.141	**.087**	-.095	-.140	-.242	-.145	-.197	**-.139**	-.210	-.161	-.223	-.173	-.195	**-.056**	-.175	-.180
.4	-.135	-.172	-.135	.268	**-.017**	-.168	-.205	-.177	-.178	**.099**	-.111	-.186	-.330	-.233	-.281	**-.153**	-.267	-.254	-.295	-.241	-.261	**-.054**	-.220	-.253
.5	-.172	-.203	-.158	.282	**-.015**	-.204	-.253	-.220	-.216	**.106**	-.125	-.235	-.453	-.367	-.405	**-.169**	-.343	-.391	-.364	-.296	-.319	**-.051**	-.248	-.318
.6	-.233	-.256	-.204	.297	**-.018**	-.266	-.371	-.362	-.335	**.108**	-.171	-.380	-.530	-.446	-.473	**-.167**	-.370	-.477	-.476	-.429	-.432	**-.059**	-.292	-.454
$g=20, k=3$			$\rho=0$						$\rho=.2$						$\rho=.5$						$\rho \sim N\left(.36, .19^2\right)(-1,1)$			
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	-.023	-.026	-.024	.143	**-.001**	-.025	-.055	**-.039**	-.047	.046	-.040	-.042	-.085	**-.047**	-.067	-.088	-.085	-.054	-.059	-.032	-.046	**-.014**	-.052	-.038
.2	-.041	-.043	-.039	.197	**.001**	-.043	-.086	-.067	-.075	.065	**-.057**	-.072	-.144	**-.094**	-.120	-.117	-.136	-.107	-.120	-.083	-.101	**-.040**	-.103	-.093
.3	-.082	-.080	-.072	.234	**-.009**	-.085	-.151	-.125	-.131	**.076**	-.096	-.136	-.228	-.156	-.192	**-.141**	-.202	-.181	-.198	-.144	-.173	**-.052**	-.159	-.164
.4	-.150	-.143	-.134	.255	**-.032**	-.155	-.203	-.157	-.169	**.097**	-.110	-.181	-.323	-.232	-.281	**-.157**	-.267	-.269	-.288	-.217	-.252	**-.052**	-.216	-.248
.5	-.227	-.205	-.196	.268	**-.054**	-.229	-.307	-.254	-.265	**.093**	-.163	-.287	-.404	-.288	-.353	**-.162**	-.301	-.340	-.370	-.274	-.321	**-.066**	-.257	-.320
.6	-.277	-.229	-.231	.288	**-.044**	-.268	-.380	-.303	-.325	**.103**	-.175	-.350	-.502	-.360	-.435	**-.164**	-.349	-.428	-.490	-.383	-.437	**-.064**	-.303	-.439
$g=20, k=6$			$\rho=0$						$\rho=.2$						$\rho=.5$						$\rho \sim N\left(.36, .19^2\right)(-1,1)$			
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	-.018	-.021	-.018	.132	**.001**	-.020	-.048	**-.031**	-.041	.034	-.037	-.033	-.077	**-.038**	-.063	-.081	-.078	-.042	-.062	-.032	-.051	**-.026**	-.057	-.035
.2	-.028	-.031	-.028	.169	**.003**	-.030	-.056	**-.036**	-.048	.061	**-.036**	-.039	-.112	**-.054**	-.092	-.103	-.108	-.064	-.098	-.060	-.084	**-.033**	-.085	-.065
.3	-.057	-.057	-.052	.213	**-.002**	-.059	-.094	-.067	-.080	.078	**-.056**	-.074	-.172	**-.099**	-.143	-.126	-.158	-.116	-.152	-.101	-.129	**-.033**	-.122	-.113
.4	-.102	-.095	-.089	.244	**-.013**	-.102	-.141	-.097	-.112	.093	**-.077**	-.112	-.254	-.161	-.215	**-.147**	-.220	-.187	-.215	-.140	-.181	**-.043**	-.165	-.163
.5	-.121	-.098	-.096	.272	**.003**	-.115	-.227	-.181	-.191	**.094**	-.121	-.202	-.321	-.203	-.269	**-.155**	-.261	-.242	-.274	-.185	-.232	**-.044**	-.195	-.216
.6	-.220	-.198	-.184	.283	**-.030**	-.221	-.293	-.229	-.246	**.104**	-.138	-.263	-.425	-.290	-.368	**-.159**	-.312	-.341	-.373	-.274	-.322	**-.048**	-.242	-.314
$g=50, k=3$			$\rho=0$						$\rho=.2$						$\rho=.5$						$\rho \sim N\left(.36, .19^2\right)(-1,1)$			
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	-.011	-.011	-.010	.125	**.006**	-.011	-.043	-.035	-.039	.034	**-.033**	-.037	-.051	**-.030**	-.043	-.068	-.052	-.036	-.047	-.032	-.041	**-.019**	-.043	-.037
.2	-.045	-.039	-.040	.194	**-.002**	-.043	-.089	-.073	-.081	**.061**	-.062	-.081	-.124	**-.079**	-.106	-.113	-.120	-.100	-.109	-.078	-.095	**-.030**	-.094	-.093
.3	-.093	-.078	-.082	.225	**-.020**	-.089	-.140	-.106	-.121	**.077**	-.088	-.125	-.215	-.145	-.191	**-.140**	-.193	-.181	-.175	-.117	-.152	**-.043**	-.140	-.147
.4	-.156	-.122	-.134	.248	**-.039**	-.146	-.189	-.130	-.161	**.090**	-.106	-.165	-.298	-.201	-.263	**-.152**	-.248	-.252	-.254	-.175	-.222	**-.051**	-.189	-.218
.5	-.274	-.224	-.246	.262	**-.080**	-.261	-.297	-.216	-.262	**.095**	-.154	-.265	-.437	-.319	-.398	**-.168**	-.332	-.383	-.343	-.236	-.303	**-.051**	-.233	-.294
.6	-.368	-.284	-.330	.278	**-.089**	-.342	-.406	-.291	-.360	**.098**	-.190	-.361	-.515	-.360	-.467	**-.169**	-.361	-.444	-.479	-.349	-.432	**-.035**	-.286	-.424
$g=50, k=6$			$\rho=0$						$\rho=.2$						$\rho=.5$						$\rho \sim N\left(.36, .19^2\right)(-1,1)$			
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	-.010	-.011	-.033	.097	**.001**	-.011	-.023	**-.016**	-.020	.032	-.017	-.017	-.047	**-.029**	-.042	-.063	-.049	-.033	-.043	-.031	-.039	**-.024**	-.041	-.033
.2	-.029	-.027	-.027	.162	**.000**	-.029	-.063	-.051	-.057	.050	**-.045**	-.055	-.085	**-.052**	-.072	-.095	-.085	-.063	-.077	-.052	-.067	**-.033**	-.069	-.061
.3	-.067	-.060	-.061	.197	**-.015**	-.065	-.084	-.060	-.072	.074	**-.052**	-.071	-.144	**-.093**	-.124	-.122	-.138	-.115	-.117	-.075	-.100	**-.033**	-.099	-.094
.4	-.080	-.059	-.067	.238	**-.002**	-.073	-.113	-.074	-.095	.087	**-.065**	-.094	-.202	**-.130**	-.175	-.136	-.182	-.163	-.170	-.107	-.145	**-.042**	-.136	-.137
.5	-.165	-.137	-.147	.248	**-.044**	-.157	-.195	-.137	-.164	**.089**	-.110	-.169	-.260	-.157	-.225	**-.145**	-.220	-.206	-.270	-.188	-.240	**-.059**	-.204	-.229
.6	-.203	-.146	-.172	.272	**-.032**	-.183	-.296	-.214	-.257	**.093**	-.155	-.261	-.396	-.272	-.357	**-.164**	-.306	-.334	-.355	-.253	-.316	**-.051**	-.240	-.307

Table 3: 95% coverage rates of credible and confidence intervals for ρ based on using the flat prior (F), Jeffreys rule prior (J), Independence Jeffreys prior (IJ), the empirical informative prior (EI), the weakly informative prior (WI), and the maximum likelihood estimator (ML). Coverage was computed by counting the proportion of the 500 replications in which the true ρ was contained in the credible or confidence interval. For each scenario, the most accurate coverage rate is printed in bold face.

$g=10, k=3$			$\rho=0$						$\rho=.2$						$\rho=.5$						$\rho \sim N\left(.36, .19^2\right)(-1,1)$			
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	.976	.860	**.966**	.842	.984	.800	.978	.878	**.964**	.992	.988	.800	.942	.836	.940	.966	**.946**	.776	.946	.868	.946	.922	**.950**	.820
.2	.984	.892	**.972**	.864	.990	.818	.972	.864	**.960**	.998	.984	.794	.942	.858	.938	.980	**.946**	.810	.936	.852	**.942**	.914	**.942**	.804
.3	.976	.838	**.956**	.880	.980	.784	**.956**	.832	.942	.994	.978	.750	**.948**	.870	**.948**	.988	.960	.802	.936	.834	.930	.934	**.952**	.776
.4	.976	.840	**.960**	.876	.988	.770	.972	.862	**.960**	1	.992	.814	.930	.854	.932	.986	**.960**	.788	.914	.826	.918	.890	**.940**	.762
.5	.976	.854	**.954**	.902	.992	.800	.974	.878	**.952**	.996	.988	.816	.926	.856	.928	.990	**.956**	.806	.936	.872	.940	.914	**.958**	.796
.6	.974	.876	**.964**	.920	.990	.804	.964	.862	**.944**	.996	.992	.792	.918	.854	**.928**	.998	.986	.802	.900	.860	.920	.930	**.944**	.788
$g=20, k=3$			$\rho=0$						$\rho=.2$						$\rho=.5$						$\rho \sim N\left(.36, .19^2\right)(-1,1)$			
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	.962	.918	**.952**	.876	.970	.900	**.944**	.906	.934	.984	**.956**	.896	.920	.906	.926	**.944**	.924	.896	.946	.904	.938	.932	**.950**	.864
.2	.964	.914	**.950**	.864	.978	.880	.958	.922	**.950**	.990	.966	.892	.940	.916	.938	.970	**.950**	.894	.938	.912	.942	.928	**.948**	.886
.3	.962	.914	**.950**	.862	.976	.892	.964	.918	**.956**	1	.980	.894	.926	.906	.930	.972	**.942**	.876	.934	.918	.940	.904	**.952**	.892
.4	.964	.938	**.958**	.892	.990	.912	.970	.914	**.964**	1	.990	.888	.916	.918	.936	.992	**.954**	.890	.934	.916	.944	.930	**.952**	.894
.5	.970	.920	**.964**	.930	.994	.886	**.942**	.916	**.942**	.998	.986	.874	.922	.926	.932	1	**.964**	.888	.926	.928	.934	.934	**.956**	.896
.6	.960	.926	**.958**	.914	.998	.880	**.950**	.930	**.950**	1	.988	.892	.914	**.938**	.936	1	.974	.894	.896	.902	.916	.922	**.950**	.868
$g=20, k=6$			$\rho=0$						$\rho=.2$						$\rho=.5$						$\rho \sim N\left(.36, .19^2\right)(-1,1)$			
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	.964	.874	**.954**	.900	.978	.852	**.952**	.874	.946	.986	.968	.844	.932	.876	**.942**	.932	.932	.850	.960	.872	**.954**	**.954**	.960	.850
.2	.970	.882	**.958**	.882	.970	.848	**.952**	.870	.946	.978	.962	.848	**.950**	.888	.946	.968	.956	.848	.940	.868	.940	.934	**.948**	.838
.3	.962	.874	**.954**	.870	.980	.842	.978	.914	**.970**	.994	.984	.890	.946	.876	**.948**	.986	.954	.890	.928	.880	.932	.924	**.946**	.850
.4	.964	.880	**.944**	.890	.990	.854	.978	.900	**.968**	.998	.998	.878	.936	.896	.934	.990	**.950**	.878	.938	.896	.938	.924	**.950**	.868
.5	.978	.886	**.972**	.870	.998	.856	.954	.882	**.948**	.998	.978	.848	.932	.906	**.942**	.992	.962	.848	**.950**	.904	.948	.932	.964	.858
.6	.974	.868	**.964**	.882	.994	.824	.952	.902	**.950**	1	.990	.866	**.946**	.914	.958	.998	.984	.866	.926	.872	**.930**	**.930**	**.970**	.826
$g=50, k=3$			$\rho=0$						$\rho=.2$						$\rho=.5$						$\rho \sim N\left(.36, .19^2\right)(-1,1)$			
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	.960	**.948**	.958	.896	.968	.944	.940	.930	.934	.986	**.946**	.928	.956	**.946**	.956	.956	.956	.936	.944	.932	.928	.944	**.946**	.924
.2	.964	**.948**	.960	.874	.972	.936	**.954**	.936	**.946**	.986	.962	.926	.934	.936	.938	.962	**.940**	.924	**.948**	.944	**.948**	.936	.946	.938
.3	**.948**	.930	.942	.856	.970	.924	**.948**	.942	.944	.998	.984	.932	.926	.942	.936	.982	**.946**	.928	.944	**.946**	.938	.940	.966	.924
.4	**.954**	**.946**	**.954**	.864	.978	.934	.960	**.948**	.954	1	.998	.934	.934	.946	.938	.990	**.952**	.932	.924	**.942**	.938	.922	.940	.930
.5	**.930**	.916	.928	.916	.988	.898	.946	.940	**.950**	1	.978	.922	.892	.922	.904	.998	**.932**	.904	.916	.942	.924	.922	**.946**	.932
.6	.938	.934	**.946**	.922	.996	.914	.914	.920	**.926**	.998	.990	.898	.884	.936	.906	.998	**.950**	.916	.900	.930	.910	.936	**.954**	.916
$g=50, k=6$			$\rho=0$						$\rho=.2$						$\rho=.5$						$\rho \sim N\left(.36, .19^2\right)(-1,1)$			
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	.934	.910	.928	.882	**.946**	.902	.962	.932	**.956**	.972	.962	.918	.954	.930	**.948**	.944	.954	.926	.946	.938	**.948**	.940	.942	.924
.2	.960	.922	**.956**	.864	.966	.914	.940	.904	.938	.976	**.950**	.902	.948	.930	.952	.956	**.950**	.924	**.954**	.942	.958	.934	.964	.936
.3	.958	.922	**.948**	.856	.976	.906	.954	.936	**.952**	.996	.974	.922	.948	.938	**.950**	.982	.954	.926	**.950**	.930	**.950**	.954	.958	.916
.4	.958	.916	**.950**	.848	.976	.914	.968	.960	.966	.994	.978	**.946**	.956	**.946**	.958	.992	.962	.934	.944	.936	.942	.918	**.948**	.920
.5	.954	.932	**.952**	.886	.990	.918	**.964**	.904	.966	1	.984	.928	**.954**	.962	.958	.988	.972	.942	.918	.928	.928	.916	**.938**	.916
.6	.972	**.952**	.970	.914	.992	.938	**.940**	.932	**.940**	.998	.978	.912	.928	.936	.932	.996	**.948**	.926	.918	.936	.928	**.940**	.956	.908

Table 4: Average of Type I and Type II error rates resulting from the 95% credible and confidence intervals for ρ based on using the flat prior (F), Jeffreys rule prior (J), Independence Jeffreys prior (IJ), the empirical informative prior (EI), the weakly informative prior (WI), and the maximum likelihood estimator (ML). For the scenarios where the true ρ was equal to zero, Type I error rates were computed by counting the proportion of the 500 replications in which 0 was not contained in the credible or confidence interval. For the scenarios where the true ρ was not zero, Type II error rates were computed by counting the proportion of the 500 replications in which 0 was contained in the credible or confidence interval. For each scenario, the smallest average error rate is printed in bold face.

$g=10, k=3$			$\rho=.2$						$\rho=.5$						$\rho \sim N\left(.36, .19^2\right)(-1,1)$			
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	.488	.482	.482	**.413**	.478	.486	.364	.273	.323	**.204**	.342	.269	.400	.344	.379	**.284**	.389	.343
.2	.491	.470	.485	**.405**	.488	.470	.412	.317	.382	**.207**	.398	.299	.428	.374	.413	**.295**	.414	.367
.3	.494	.497	.497	**.422**	.497	.477	.454	.392	.427	**.252**	.442	.374	.461	.430	.449	**.320**	.450	.404
.4	.502	.519	.502	**.444**	.497	.524	.480	.446	.467	**.322**	.465	.451	.489	.476	.482	**.392**	.479	.472
.5	.505	.522	.510	**.452**	.500	.522	.502	.505	.505	**.396**	.496	.495	.499	.502	.500	**.402**	.491	.474
.6	.504	.511	.504	**.471**	.501	.504	.507	.507	.508	**.405**	.499	.501	.508	.513	.507	**.434**	.501	.511
$g=20, k=3$			$\rho=.2$						$\rho=.5$						$\rho \sim N\left(.36, .19^2\right)(-1,1)$			
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	.452	.431	.446	**.321**	.440	.417	.164	.147	.152	**.092**	.149	.141	.290	.263	.278	**.200**	.280	.258
.2	.470	.453	.469	**.374**	.457	.457	.327	.278	.305	**.170**	.290	.272	.373	.342	.360	**.258**	.357	.343
.3	.501	.506	.504	**.421**	.498	.501	.426	.384	.412	**.252**	.411	.374	.490	.426	.450	**.318**	.443	.419
.4	.504	.494	.499	**.420**	.495	.489	.482	.437	.469	**.317**	.462	.423	.494	.472	.487	**.383**	.480	.467
.5	.509	.503	.506	**.452**	.499	.507	.498	.481	.486	**.346**	.482	.470	.501	.500	.495	**.387**	.491	.494
.6	.509	.509	.505	**.472**	.501	.516	.514	.507	.509	**.398**	.497	.504	.508	.500	.504	**.424**	.499	.491
$g=20, k=6$			$\rho=.2$						$\rho=.5$						$\rho \sim N\left(.36, .19^2\right)(-1,1)$			
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	.427	.390	.420	**.293**	.410	.385	.141	.115	.124	**.073**	.120	.122	.271	.223	.262	**.173**	.256	.221
.2	.452	.421	.449	**.335**	.444	.426	.263	.203	.241	**.126**	.245	.205	.338	.298	.323	**.219**	.333	.302
.3	.497	.487	.492	**.380**	.481	.489	.383	.316	.358	**.185**	.354	.302	.445	.391	.431	**.306**	.430	.385
.4	.509	.498	.513	**.402**	.494	.486	.456	.397	.445	**.251**	.434	.389	.477	.440	.469	**.329**	.461	.425
.5	.504	.501	.501	**.448**	.499	.499	.486	.447	.478	**.323**	.472	.434	.493	.470	.483	**.372**	.479	.459
.6	.499	.514	.499	**.462**	.497	.522	.499	.493	.492	**.373**	.483	.491	.500	.492	.497	**.399**	.489	.489
$g=50, k=3$			$\rho=.2$						$\rho=.5$						$\rho \sim N\left(.36, .19^2\right)(-1,1)$			
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	.412	.393	.403	**.295**	.396	.388	.098	.093	.095	**.065**	.088	.087	.237	.223	.236	**.172**	.229	.219
.2	.482	.475	.482	**.366**	.482	.475	.296	.274	.292	**.149**	.282	.275	.381	.361	.372	**.267**	.365	.359
.3	.506	.497	.498	**.423**	.494	.496	.416	.406	.413	**.248**	.401	.400	.452	.436	.447	**.299**	.437	.421
.4	.513	.506	.510	**.429**	.504	.504	.482	.441	.467	**.309**	.464	.436	.489	.462	.477	**.349**	.472	.452
.5	.516	.516	.515	**.451**	.496	.516	.512	.498	.506	**.353**	.482	.494	.515	.496	.510	**.381**	.485	.484
.6	.511	.507	.506	**.470**	.499	.507	.519	.505	.509	**.385**	.493	.497	.514	.504	.510	**.423**	.499	.505
$g=50, k=6$			$\rho=.2$						$\rho=.5$						$\rho \sim N\left(.36, .19^2\right)(-1,1)$			
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	.351	.335	.351	**.242**	.338	.326	.073	.066	.071	.064	**.058**	.069	.193	.187	.190	**.160**	.180	.188
.2	.461	.456	.460	**.345**	.448	.451	.205	.190	.204	**.103**	.196	.188	.304	.297	.301	**.217**	.300	.290
.3	.486	.472	.479	**.360**	.471	.470	.350	.312	.338	**.167**	.323	.307	.404	.380	.398	**.265**	.384	.381
.4	.502	.503	.504	**.404**	.492	.494	.430	.384	.420	**.239**	.413	.371	.440	.427	.434	**.308**	.427	.412
.5	.509	.501	.502	**.431**	.490	.504	.481	.435	.464	**.269**	.451	.425	.490	.469	.490	**.361**	.475	.468
.6	.501	.504	.503	**.450**	.503	.505	.493	.480	.489	**.336**	.483	.477	.489	.478	.486	**.309**	.486	.478

Table 5: Average bias of σ^2 based on using the flat prior (F), Jeffreys rule prior (J), Independence Jeffreys prior (IJ), the empirical informative prior (EI), the weakly informative prior (WI), and the maximum likelihood estimator (ML). The average bias was computed by taking the average of the difference between the estimated network effect and the true $\sigma^2 = 1$ value used for data-generation over 500 replications. For each scenario, the best estimate is printed in bold face.

$g = 10, k = 3$	$\rho = 0$						$\rho = .2$						$\rho = .5$						$\rho \sim N\left(.36, .19^2\right)(-1, 1)$					
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	.072	-.440	**-.008**	.220	.078	-.492	.122	-.416	**.034**	.139	.112	-.470	.192	-.413	**.054**	.187	.183	-.467	.099	-.452	**-.018**	.139	.095	-.501
.2	.119	-.414	**.027**	.263	.124	-.470	.059	-.452	**-.038**	.095	.053	-.503	.144	-.443	**.006**	.126	.133	-.485	.102	-.448	**-.023**	.143	.098	-.499
.3	**.021**	-.466	-.064	.211	.036	-.516	.067	-.446	**-.032**	.126	.065	-.497	.147	-.423	**.010**	.132	.134	-.476	.083	-.453	**-.044**	.130	.081	-.504
.4	**.014**	-.469	-.078	.205	.038	-.519	.060	-.441	**-.036**	.127	.062	-.493	.104	-.436	**-.024**	.105	.094	-.486	.079	-.443	**-.039**	.133	.078	-.495
.5	**.007**	-.469	-.079	.189	.034	-.518	**.033**	-.453	-.062	.107	.040	-.503	.098	-.423	**-.028**	.111	.092	-.475	.081	-.436	**-.037**	.126	.080	-.488
.6	**-.007**	-.474	-.095	.182	.034	-.522	**.009**	-.464	-.088	.120	.031	-.513	**.040**	-.452	-.080	.070	.041	-.500	**.038**	-.452	-.074	.109	.053	-.500
$g = 20, k = 3$	$\rho = 0$						$\rho = .2$						$\rho = .5$						$\rho \sim N\left(.36, .19^2\right)(-1, 1)$					
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	-.006	-.238	-.043	.041	**-.005**	-.274	.017	-.223	-.023	.016	**.014**	-.259	.040	-.218	**-.011**	.044	.039	-.253	.039	-.214	**-.008**	.049	.038	-.250
.2	**.012**	-.222	-.027	.071	.014	-.260	.023	-.215	**-.019**	.031	.021	-.252	.056	-.197	**.007**	.055	.054	-.234	.007	-.234	-.041	.019	**.005**	-.270
.3	.021	-.213	**-.018**	.089	.026	-.250	.021	-.213	**-.019**	.041	.020	-.251	.037	-.208	**-.012**	.038	.034	-.244	.033	-.209	**-.014**	.047	.032	-.245
.4	**.009**	-.221	-.030	.081	.018	-.257	**.021**	-.211	**-.021**	.053	.023	-.248	.043	-.200	**-.009**	.046	.040	-.236	.067	-.179	**-.019**	.085	.066	-.217
.5	-.018	-.240	-.055	.064	**-.004**	-.276	**.014**	-.216	-.028	.055	.020	-.253	.047	-.194	**-.006**	.051	.044	-.231	.016	-.216	-.031	.034	**.015**	-.252
.6	-.011	-.234	-.050	.069	**.005**	-.271	-.014	-.235	-.053	.036	**-.003**	-.271	.036	-.200	**-.014**	.047	.035	-.236	**.003**	-.226	-.045	.039	.010	-.261
$g = 20, k = 6$	$\rho = 0$						$\rho = .2$						$\rho = .5$						$\rho \sim N\left(.36, .19^2\right)(-1, 1)$					
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	.037	-.372	**-.016**	.094	.037	-.400	.043	-.371	**-.012**	.037	.040	-.399	.049	-.380	**-.018**	.050	.049	-.407	.088	-.354	**.022**	.105	.088	-.382
.2	**.009**	-.388	-.044	.067	**.009**	-.416	.006	-.391	-.050	.007	**.003**	-.419	.081	-.355	**.013**	.078	.078	-.384	.068	-.360	**.004**	.082	.066	-.388
.3	**.021**	-.379	-.032	.097	.023	-.408	.056	-.356	**.001**	.063	.052	-.386	.029	-.382	-.033	.025	**.024**	-.409	.041	-.371	**-.019**	.049	.038	-.399
.4	**.008**	-.385	-.044	.088	.012	-.413	.023	-.376	-.029	.039	**.021**	-.404	.042	-.370	**-.021**	.039	.037	-.398	.065	-.354	**.005**	.078	.063	-.383
.5	**.005**	-.385	-.047	.075	.011	-.413	**.014**	-.379	-.039	.049	.015	-.407	.054	-.360	**-.009**	.047	.047	-.388	.047	-.362	**-.014**	.064	.043	-.391
.6	**.019**	-.375	-.031	.108	.034	-.403	**.000**	-.385	-.053	.041	.004	-.414	.084	-.337	**.018**	.085	.079	-.366	**.000**	-.390	-.057	.026	.001	-.417
$g = 50, k = 3$	$\rho = 0$						$\rho = .2$						$\rho = .5$						$\rho \sim N\left(.36, .19^2\right)(-1, 1)$					
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	**.002**	-.090	-.013	.016	**.002**	-.107	.002	-.090	-.014	**.000**	-.001	-.107	-.008	-.103	-.026	**-.004**	-.008	-.118	**.003**	-.092	-.014	.009	**.003**	-.108
.2	**.005**	-.085	-.010	.025	.006	-.102	**.007**	-.084	-.009	.011	.013	-.101	**.001**	-.093	-.018	.003	**.001**	-.109	**.000**	-.092	-.017	.005	**.000**	-.108
.3	-.004	-.093	-.019	.021	**-.002**	-.110	.011	-.080	-.006	.019	**-.002**	-.097	.022	-.072	**.003**	.021	.021	-.088	**.009**	-.084	**-.009**	.015	**.009**	-.100
.4	**.001**	-.088	-.014	.029	.004	-.105	**.008**	-.081	**-.008**	.018	.015	-.099	.026	-.068	**.007**	.026	.025	-.084	.019	-.074	**.001**	.024	.019	-.090
.5	-.012	-.099	-.028	.022	**-.006**	-.117	**-.005**	-.093	-.022	.009	-.007	-.110	.015	-.076	**-.004**	.020	.015	-.093	**.002**	-.088	-.015	.010	.003	-.104
.6	**.004**	-.083	-.012	**.004**	.013	-.101	**.002**	-.085	-.014	.021	.015	-.102	-.006	-.094	-.024	**.000**	-.005	-.111	**.006**	-.083	-.011	.021	.009	-.100
$g = 50, k = 6$	$\rho = 0$						$\rho = .2$						$\rho = .5$						$\rho \sim N\left(.36, .19^2\right)(-1, 1)$					
d	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML	F	J	IJ	EI	WI	ML
.1	.017	-.141	**-.010**	.032	.018	-.156	**.000**	-.156	-.018	-.003	-.001	-.170	.016	-.145	**-.004**	.019	.016	-.159	**.007**	-.151	-.012	.011	**.007**	-.166
.2	.035	-.125	**.017**	.054	.035	-.140	.014	-.143	**-.004**	.015	.013	-.158	.028	-.133	**.007**	.028	.027	-.148	.012	-.146	**-.008**	.016	.012	-.161
.3	**.004**	-.150	-.014	.028	.005	-.165	**-.001**	-.155	-.020	.004	-.002	-.170	.028	-.132	**.008**	.027	.027	-.147	.012	-.144	**-.007**	.015	.011	-.159
.4	**.005**	-.148	-.012	.027	.006	-.163	.016	-.139	**-.003**	.020	.015	-.155	.008	-.147	-.012	**.007**	**.007**	-.162	.013	-.143	**-.007**	.017	.012	-.158
.5	-.007	-.158	-.024	.021	**-.005**	-.173	-.006	-.157	-.024	**.003**	-.007	-.172	.022	-.135	**.001**	.020	.020	-.150	.024	-.132	**.005**	.028	.023	-.148
.6	.007	-.145	-.012	**.003**	.009	-.161	.013	-.139	**-.005**	.028	.015	-.155	.008	-.145	-.012	.009	**.006**	-.160	.015	-.139	**-.004**	.023	.015	-.154

6.2. Simulation results

Table 2 shows the average bias of ρ for the different estimators. Overall, the Bayesian estimators based on the non-informative priors yield similar results to those based on maximum likelihood estimation. In particular, there is still some negative bias present which is a well-known issue in the network autocorrelation model. On the other hand, if the true underlying ρ equals zero, the Bayesian estimator based on the weakly informative prior eliminates virtually all the negative bias in the estimation of ρ. Furthermore, when the data-generating network effect is positive, using the empirical informative prior generally results in the least absolute bias of ρ. Given our review of empirically observed network autocorrelations in Section 4.4, this is clearly the most common situation to be encountered in practice. Lastly, we also observe a much smaller increase in bias for higher network densities for this estimator, compared to the non-informative Bayesian ones and the maximum likelihood estimator.

Table 3 shows the empirical frequentist coverages of the Bayesian equal-tailed 95% credible intervals for ρ for the Bayesian estimators and coverages of the asymptotic 95% confidence intervals for ρ. The coverages of the credible intervals based on the Independence Jeffreys prior and the flat prior are similar to each other and very close to the nominal .95. In contrast to that, the coverages of the credible intervals corresponding to the Jeffreys rule prior and the coverages of the maximum likelihood-based confidence intervals are below nominal for all considered scenarios. The problem of subpar coverages of the maximum likelihood-based confidence intervals for ρ is completely resolved when using Bayesian estimators based on the flat prior or the Independence Jeffreys prior.

Table 4 reports the average of the empirical Type I and Type II error rates of ρ for the different estimators. In general, the average error rates increase with the network density due to the negative bias in the estimation of ρ, and they decrease for higher network autocorrelations as a result of higher power. For all considered scenarios, the Bayesian estimator based on the empirical informative prior clearly yields the smallest average Type I and Type II error rates across the board. The other estimators perform relatively similar to each other, with the maximum likelihood estimator having slightly smaller average error rates than the remaining Bayesian ones but still considerably higher than the estimator based on the empirical informative prior. The greater power of the maximum likelihood estimator, compared to the Bayesian estimators based on the non-informative priors, comes at the price of underestimating the standard error of ρ. In turn, this results in narrower confidence intervals for ρ, leading to lower coverage rates but slightly higher power. Regardless, estimating ρ using the empirical informative prior yields the lowest average Type I and Type II error rates.

Table 5 displays the average bias of σ^2 for the Bayesian estimators and the maximum likelihood one. The estimates for σ^2 corresponding to the use of the flat, the Independence Jeffreys, and the informative priors are nearly unbiased, while the results based on the Jeffreys rule prior and maximum likelihood estimation exhibit a large negative bias. This bias is particularly pronounced for a higher number of covariates. We also investigated the associated coverages of the Bayesian equal-tailed 95% credible intervals and the asymptotic 95% confidence intervals for σ^2. In line with the results for the average bias of σ^2, we found that the coverages of the credible intervals based on the flat, the Independence Jeffreys, and the informative prior are very close to the nominal .95. On the other hand, the coverages of the credible intervals corresponding to the Jeffreys rule prior and the coverages of the maximum likelihood-based confidence intervals are well below the nominal rate for all considered scenarios. These results are not shown here but are available from the authors on request.

Based on our simulation output, we suggest the following: first, if a researcher is willing to expect that his, or her, study might have a network effect along the lines of the overall distribution of autocorrelation effects across the literature at large, using the empirical informative prior is highly recommended as it leads to a dramatic decrease of the bias in the estimation of ρ. Furthermore, the corresponding estimator exhibits by far the smallest average Type I and Type II error rates and accurately estimates σ^2. At the same time, applying the empirical informative prior can result in a mild overestimation of ρ for small positive network effects. However, we believe this to be less of a concern than falsely dismissing positive network effects and stress that overall this estimator performs clearly the best.

Second, if a researcher does not expect a network effect to be present in the data or if the researcher does not (want to) entertain the prior belief that the level of autocorrelation in a dataset is likely to fit with the extant empirical empirical literature, relying on the weakly informative prior yields nearly adequate point estimates of the network effect in these cases. This does, however, require the researcher to sacrifice the Type I and Type II reducing benefits of the empirical informative prior.

Third, if a researcher prefers to refrain from employing any empirical-based prior information, we recommend using the non-informative Independence Jeffreys prior. While this does not attenuate the negative bias in the estimation of ρ, the issue of poor coverage of the confidence intervals, associated with maximum likelihood estimation of the model, can be completely eluded at least. We wish to emphasize that there is never a case where maximum likelihood estimation can be recommended.

Lastly, when analyzing a real data set, we advise researchers to estimate the model using all three recommended priors. If the resulting estimates for ρ are close to each other, this implies that the data contain sufficient information, and the estimates are most likely highly reliable; else, this strongly points at (negative) bias in the estimation of the network effect.

7. Conclusions

In this work, we derived two versions of Jeffreys prior for the network autocorrelation model which provide default Bayesian analyses for this model. Moreover, we specified an empirical informative prior and a weakly informative prior for the network effect ρ based on reported network effects from the literature. We evaluated the Bayesian estimators by means of a simulation study and compared their performance to the maximum likelihood estimator. We found that the Bayesian estimator based on the empirical informative prior performs superior and that the estimator based on the weakly informative prior can be a useful alternative. Concomitantly, we also provided a very efficient MCMC implementation of the Bayesian approach which is preferable to existing sampling schemes and ensures a fast and accurate Bayesian estimation of the network autocorrelation model.

In order to allow researchers and practitioners to easily use the newly developed methods in this paper, it is essential to make them accessible in a statistical software package which will be made available by the time of publication of this paper. In addition, as we primarily focused on Bayesian point estimation in this work, further work needs to be done in studying Bayesian model selection procedures for the discussed priors. Finally, despite the improved numerical properties of the Bayesian estimators, the negative bias of ρ in the model is not entirely resolved. We did resolve much of the bias for data sets that are typical in the empirical literature at large, but more research is needed to untangle it completely. It remains a major challenge to investigate what causes this negative bias and why it becomes increasingly salient at high levels of network density.

Appendices

Auxiliary facts

(i) Let $A_\rho = I_g - \rho W$. If $\left(X^T X\right)^{-1}$ exists, then

$$(A_\rho \boldsymbol{y} - X\boldsymbol{\beta})^T (A_\rho \boldsymbol{y} - X\boldsymbol{\beta}) = \left(A_\rho \boldsymbol{y} - X\hat{\boldsymbol{\beta}}\right)^T \left(A_\rho \boldsymbol{y} - X\hat{\boldsymbol{\beta}}\right) + \left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)^T \left(X^T X\right) \left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)$$
$$= \boldsymbol{y}^T A_\rho^T M A_\rho \boldsymbol{y} + \left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)^T \left(X^T X\right) \left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right),$$

where $\hat{\boldsymbol{\beta}} := \left(X^T X\right)^{-1} X^T A_\rho \boldsymbol{y}$ and $M = I_g - X\left(X^T X\right)^{-1} X^T$. Note that M is symmetric and idempotent.

(ii) Let A be an invertible matrix, and let $\boldsymbol{u}$ and $\boldsymbol{v}$ two vectors. Then, $A + \boldsymbol{u}\boldsymbol{v}^T$ is invertible and

$$\det\left(A + \boldsymbol{u}\boldsymbol{v}^T\right) = \left(1 + \boldsymbol{v}^T A^{-1} \boldsymbol{u}\right) \det(A).$$

(iii) Let $\boldsymbol{Z} \sim N(\boldsymbol{\mu}, \Sigma)$, and let A be a symmetric matrix. Then,

$$\mathbb{E}\left[\boldsymbol{Z}^T A \boldsymbol{Z}\right] = \operatorname{tr}(A\Sigma) + \boldsymbol{\mu}^T A \boldsymbol{\mu}.$$

(iv) Let A and B be symmetric and positive semidefinite matrices. Then (Yang (2000), Lemma 1),

$$0 \leq \operatorname{tr}(AB) \leq \operatorname{tr}(A)\operatorname{tr}(B).$$

(v) Let A and B be matrices. Then,

$$\operatorname{tr}(AB) \leq \frac{1}{2}\left(\operatorname{tr}\left(A^2\right) + \operatorname{tr}\left(B^2\right)\right).$$

(vi) Let A be an idempotent matrix. Then, the eigenvalues of A are either 0 or 1.

Proof of Corollary 1

(i) Follows immediately from the prior's definition and $\Omega = \left(\lambda_g^{-1}, \lambda_1^{-1}\right) \times (0, \infty) \times \mathbb{R}^k$.

(ii) Using auxiliary fact (i), Hepple (1995a) showed that if $\left(X^T X\right)^{-1}$ exists and $g > k$, the corresponding marginal posterior for ρ is

$$p_{\mathrm{F}}(\rho|\boldsymbol{y}) \propto |A_\rho| \left(\boldsymbol{y}^T A_\rho^T M A_\rho \boldsymbol{y}\right)^{-\frac{g-k}{2}}.$$

As $|A_\rho| \leq 1$ and the assumption that $\left(\boldsymbol{y}^T M W \boldsymbol{y}\right)^2 \neq \boldsymbol{y}^T W^T M W \boldsymbol{y}\boldsymbol{y}^T M \boldsymbol{y}$ ensures that $\boldsymbol{y}^T A_\rho^T M A_\rho \boldsymbol{y} > 0$, $p_{\mathrm{F}}(\rho|\boldsymbol{y})$ is bounded on $\left(\lambda_g^{-1}, \lambda_1^{-1}\right)$ which proves the statement.

Remark: As $|A_\rho| = \mathcal{O}\left(\rho^{g-m_0}\right)$ for $\rho \to \infty$ (where $m_0 \geq 0$ denotes the algebraic multiplicity of an eventual zero eigenvalue of W) and $\left(\boldsymbol{y}^T A_\rho^T M A_\rho \boldsymbol{y}\right)^{-(g-k)/2} = \mathcal{O}\left(\rho^{k-g}\right)$, it follows that $p_{\mathrm{F}}(\rho|\boldsymbol{y}) = \mathcal{O}\left(\rho^{k-m_0}\right)$. Hence, the marginal posterior for ρ is integrable over $\mathbb{R} \setminus \left\{{\lambda_1}^{-1}, {\lambda_2}^{-1}, ..., {\lambda_g}^{-1}\right\}$ only if $k < m_0 - 1$ which is typically not the case.

Proof of Theorem 1

The model's Fisher Information Matrix of $\boldsymbol{\theta} = \left(\rho, \sigma^2, \boldsymbol{\beta}\right)$ is (Doreian, 1981)

$$
\begin{aligned}
I(\boldsymbol{\theta}) &= \frac{1}{\sigma^2} I^*(\boldsymbol{\theta}) \\
&= \frac{1}{\sigma^2} \begin{bmatrix} \sigma^2 \left(\operatorname{tr}\left(B_\rho^T B_\rho\right) + \operatorname{tr}\left(B_\rho^2\right)\right) + \boldsymbol{\beta}^T X^T B_\rho^T B_\rho X \boldsymbol{\beta} & \operatorname{tr}\left(B_\rho\right) & \left(X^T B_\rho X \boldsymbol{\beta}\right)^T \\ \operatorname{tr}\left(B_\rho\right) & \frac{g}{2\sigma^2} & \mathbf{0}^T \\ X^T B_\rho X \boldsymbol{\beta} & \mathbf{0} & X^T X \end{bmatrix} \qquad (10) \\
&= \frac{1}{\sigma^2} \begin{bmatrix} I^*_{\rho,\rho} & I^*_{\rho,\sigma^2} & I^*_{\rho,\boldsymbol{\beta}} \\ I^*_{\sigma^2,\rho} & I^*_{\sigma^2,\sigma^2} & I^*_{\sigma^2,\boldsymbol{\beta}} \\ I^*_{\boldsymbol{\beta},\rho} & I^*_{\boldsymbol{\beta},\sigma^2} & I^*_{\boldsymbol{\beta},\boldsymbol{\beta}} \end{bmatrix}.
\end{aligned}
$$

Using cofactor expansion and determinant properties of block matrices, we can write

$$
\begin{aligned}
\det\left(I^*(\boldsymbol{\theta})\right) &= -I^*_{\sigma^2,\rho} \det \begin{pmatrix} I^*_{\rho,\sigma^2} & I^*_{\rho,\boldsymbol{\beta}} \\ I^*_{\boldsymbol{\beta},\sigma^2} & I^*_{\boldsymbol{\beta},\boldsymbol{\beta}} \end{pmatrix} + I^*_{\sigma^2,\sigma^2} \det \begin{pmatrix} I^*_{\rho,\rho} & I^*_{\rho,\boldsymbol{\beta}} \\ I^*_{\boldsymbol{\beta},\rho} & I^*_{\boldsymbol{\beta},\boldsymbol{\beta}} \end{pmatrix} \\
&= -I^*_{\sigma^2,\rho} \det\left(I^*_{\rho,\sigma^2}\right) \det\left(I^*_{\boldsymbol{\beta},\boldsymbol{\beta}}\right) + I^*_{\sigma^2,\sigma^2} \det \begin{pmatrix} I^*_{\rho,\rho} & I^*_{\rho,\boldsymbol{\beta}} \\ I^*_{\boldsymbol{\beta},\rho} & I^*_{\boldsymbol{\beta},\boldsymbol{\beta}} \end{pmatrix} \\
&= -{I^*_{\sigma^2,\rho}}^2 \det\left(I^*_{\boldsymbol{\beta},\boldsymbol{\beta}}\right) + I^*_{\sigma^2,\sigma^2} \det\left(I^*_{\rho,\rho}\right) \det\left(I^*_{\boldsymbol{\beta},\boldsymbol{\beta}} - I^*_{\boldsymbol{\beta},\rho} {I^*_{\rho,\rho}}^{-1} I^*_{\rho,\boldsymbol{\beta}}\right) \\
&= -{I^*_{\sigma^2,\rho}}^2 \det\left(I^*_{\boldsymbol{\beta},\boldsymbol{\beta}}\right) + I^*_{\sigma^2,\sigma^2} I^*_{\rho,\rho} \det\left(I^*_{\boldsymbol{\beta},\boldsymbol{\beta}} - I^*_{\boldsymbol{\beta},\rho} {I^*_{\rho,\rho}}^{-1} I^*_{\rho,\boldsymbol{\beta}}\right). \qquad (11)
\end{aligned}
$$

By auxiliary fact (ii), we can further write (11) as

$$
\begin{aligned}
\det\left(I^*(\boldsymbol{\theta})\right) &= -{I^*_{\sigma^2,\rho}}^2 \det\left(I^*_{\boldsymbol{\beta},\boldsymbol{\beta}}\right) + I^*_{\sigma^2,\sigma^2} I^*_{\rho,\rho} \left(1 - I^*_{\rho,\boldsymbol{\beta}} {I^*_{\boldsymbol{\beta},\boldsymbol{\beta}}}^{-1} I^*_{\boldsymbol{\beta},\rho} {I^*_{\rho,\rho}}^{-1}\right) \det\left(I^*_{\boldsymbol{\beta},\boldsymbol{\beta}}\right) \\
&= \det\left(I^*_{\boldsymbol{\beta},\boldsymbol{\beta}}\right) \left(I^*_{\sigma^2,\sigma^2} I_{\rho,\rho} - I^*_{\sigma^2,\sigma^2} I^*_{\rho,\boldsymbol{\beta}} {I^*_{\boldsymbol{\beta},\boldsymbol{\beta}}}^{-1} I^*_{\boldsymbol{\beta},\rho} - {I^*_{\sigma^2,\rho}}^2\right).
\end{aligned}
$$

Plugging in the actual entries for the respective blocks yields

$$
\begin{aligned}
& I^*_{\sigma^2,\sigma^2} I^*_{\rho,\rho} - I^*_{\sigma^2,\sigma^2} I^*_{\rho,\boldsymbol{\beta}} {I^*_{\boldsymbol{\beta},\boldsymbol{\beta}}}^{-1} I^*_{\boldsymbol{\beta},\rho} - {I^*_{\sigma^2,\rho}}^2 \\
&= \frac{g}{2\sigma^2} \left(\sigma^2 \left(\operatorname{tr}\left(B_\rho^T B_\rho\right) + \operatorname{tr}\left(B_\rho^2\right)\right) + \boldsymbol{\beta}^T X^T B_\rho^T B_\rho X \boldsymbol{\beta}\right) - \frac{g}{2\sigma^2} \left(X^T B_\rho X \boldsymbol{\beta}\right)^T \left(X^T X\right)^{-1} X^T B_\rho X \boldsymbol{\beta} - \operatorname{tr}^2\left(B_\rho\right) \\
&= \frac{g}{2} \left(\operatorname{tr}\left(B_\rho^T B_\rho\right) + \operatorname{tr}\left(B_\rho^2\right) + \frac{1}{\sigma^2} \boldsymbol{\beta}^T X^T B_\rho^T \left(I_g - X \left(X^T X\right)^{-1} X^T\right) B_\rho X \boldsymbol{\beta} - \frac{2}{g} \operatorname{tr}^2\left(B_\rho\right)\right) \\
&= \frac{g}{2} \left(\operatorname{tr}\left(B_\rho^T B_\rho\right) + \operatorname{tr}\left(B_\rho^2\right) + \frac{1}{\sigma^2} \boldsymbol{\beta}^T X^T B_\rho^T M B_\rho X \boldsymbol{\beta} - \frac{2}{g} \operatorname{tr}^2\left(B_\rho\right)\right).
\end{aligned}
$$

Thus,

$$\begin{aligned}
\det\left(I\left(\boldsymbol{\theta}\right)\right) &= \det\left(\frac{1}{\sigma^2} I^*\left(\boldsymbol{\theta}\right)\right) = \left(\sigma^2\right)^{-k-2} \det\left(I^*\left(\boldsymbol{\theta}\right)\right) \\
&= \left(\sigma^2\right)^{-k-2} \det\left(X^T X\right) \frac{g}{2} \left(\operatorname{tr}\left(B_\rho^T B_\rho\right) + \operatorname{tr}\left(B_\rho^2\right) + \frac{1}{\sigma^2} \boldsymbol{\beta}^T X^T B_\rho^T M B_\rho X \boldsymbol{\beta} - \frac{2}{g} \operatorname{tr}^2\left(B_\rho\right)\right) \\
&\propto \left(\sigma^2\right)^{-k-2} \left(\operatorname{tr}\left(B_\rho^T B_\rho\right) + \operatorname{tr}\left(B_\rho^2\right) + \frac{1}{\sigma^2} \boldsymbol{\beta}^T X^T B_\rho^T M B_\rho X \boldsymbol{\beta} - \frac{2}{g} \operatorname{tr}^2\left(B_\rho\right)\right),
\end{aligned}$$

from which, together with the definition of Jeffreys rule prior, the result follows.

Proof of Corollary 2

(i) From the definition of Jeffreys rule prior it follows that

$$\begin{aligned}
&\int_0^\infty p_{\mathrm{J}}\left(\boldsymbol{\theta}\right) \mathrm{d}\sigma^2 > \int_0^1 p_{\mathrm{J}}\left(\boldsymbol{\theta}\right) \mathrm{d}\sigma^2 \\
&\propto \int_0^1 \left(\sigma^2\right)^{-\frac{k+2}{2}} \left\{\operatorname{tr}\left(B_\rho^T B_\rho\right) + \operatorname{tr}\left(B_\rho^2\right) + \frac{1}{\sigma^2} \boldsymbol{\beta}^T X^T B_\rho^T M B_\rho X \boldsymbol{\beta} - \frac{2}{g} \operatorname{tr}^2\left(B_\rho\right)\right\}^{1/2} \mathrm{d}\sigma^2 \\
&> \int_0^1 \left(\sigma^2\right)^{-\frac{k+2}{2}} \left\{\operatorname{tr}\left(B_\rho^T B_\rho\right) + \operatorname{tr}\left(B_\rho^2\right) + \boldsymbol{\beta}^T X^T B_\rho^T M B_\rho X \boldsymbol{\beta} - \frac{2}{g} \operatorname{tr}^2\left(B_\rho\right)\right\}^{1/2} \mathrm{d}\sigma^2 \\
&= \left\{\operatorname{tr}\left(B_\rho^T B_\rho\right) + \operatorname{tr}\left(B_\rho^2\right) + \boldsymbol{\beta}^T X^T B_\rho^T M B_\rho X \boldsymbol{\beta} - \frac{2}{g} \operatorname{tr}^2\left(B_\rho\right)\right\}^{1/2} \int_0^1 \left(\sigma^2\right)^{-\frac{k+2}{2}} \mathrm{d}\sigma^2 \\
&= \infty.
\end{aligned}$$

(ii) Defining $h_1\left(\rho\right) := \operatorname{tr}\left(B_\rho^T B_\rho\right) + \operatorname{tr}\left(B_\rho^2\right) \geq 0$ and $h_2\left(\rho, \boldsymbol{\beta}\right) := \boldsymbol{\beta}^T X^T B_\rho^T M B_\rho X \boldsymbol{\beta} \geq 0$, we can write for the Jeffreys rule posterior

$$\begin{aligned}
p_{\mathrm{J}}\left(\boldsymbol{\theta}|\boldsymbol{y}\right) &\propto \left(\sigma^2\right)^{-\frac{k+2}{2}} \sqrt{\operatorname{tr}\left(B_\rho^T B_\rho\right) + \operatorname{tr}\left(B_\rho^2\right) + \frac{1}{\sigma^2} \boldsymbol{\beta}^T X^T B_\rho^T M B_\rho X \boldsymbol{\beta} - \frac{2}{g} \operatorname{tr}^2\left(B_\rho\right)} \\
&\quad \times \left|A_\rho\right| \left(\sigma^2\right)^{-\frac{g}{2}} \exp\left(-\frac{1}{2\sigma^2} \left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)^T \left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)\right) \\
&\leq \left(\sigma^2\right)^{-\frac{g+k}{2}-1} \sqrt{h_1\left(\rho\right) + \frac{1}{\sigma^2} h_2\left(\rho, \boldsymbol{\beta}\right)} \left|A_\rho\right| \exp\left(-\frac{1}{2\sigma^2} \left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)^T \left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)\right).
\end{aligned} \tag{12}$$

Using auxilliary fact (i) and integrating (12) over $\boldsymbol{\beta}$ yields

$$\int_{\mathbb{R}^k} \left(\sigma^2\right)^{-\frac{g+k}{2}-1} \sqrt{h_1(\rho) + \frac{1}{\sigma^2} h_2(\rho, \boldsymbol{\beta})} |A_\rho| \exp\left(-\frac{1}{2\sigma^2}\left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)^T \left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)\right) \mathrm{d}\boldsymbol{\beta}$$

$$\leq \left(\sigma^2\right)^{-\frac{g+k}{2}-1} |A_\rho| \int_{\mathbb{R}^k} \left(\sqrt{h_1(\rho)} + \sqrt{\frac{1}{\sigma^2} h_2(\rho, \boldsymbol{\beta})}\right) \exp\left(-\frac{1}{2\sigma^2}\left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)^T \left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)\right) \mathrm{d}\boldsymbol{\beta}$$

$$= \left(\sigma^2\right)^{-\frac{g+k}{2}-1} |A_\rho| \exp\left(-\frac{1}{2\sigma^2} \boldsymbol{y}^T A_\rho^T M A_\rho \boldsymbol{y}\right)$$

$$\int_{\mathbb{R}^k} \left(\sqrt{h_1(\rho)} + \sqrt{\frac{1}{\sigma^2} h_2(\rho, \boldsymbol{\beta})}\right) \exp\left(-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)^T \left(X^T X\right)\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right) \mathrm{d}\boldsymbol{\beta}$$

$$= \left(\sigma^2\right)^{-\frac{g+k}{2}-1} |A_\rho| \exp\left(-\frac{1}{2\sigma^2} \boldsymbol{y}^T A_\rho^T M A_\rho \boldsymbol{y}\right) \sqrt{h_1(\rho)} \int_{\mathbb{R}^k} \exp\left(-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)^T \left(X^T X\right)\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right) \mathrm{d}\boldsymbol{\beta} \tag{13}$$

$$+ \left(\sigma^2\right)^{-\frac{g+k}{2}-1} |A_\rho| \exp\left(-\frac{1}{2\sigma^2} \boldsymbol{y}^T A_\rho^T M A_\rho \boldsymbol{y}\right) \left(\sigma^2\right)^{-\frac{1}{2}} \int_{\mathbb{R}^k} \sqrt{h_2(\rho, \boldsymbol{\beta})} \exp\left(-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)^T \left(X^T X\right)\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right) \mathrm{d}\boldsymbol{\beta}. \tag{14}$$

The integrand in (13) is the kernel of a random variable $\boldsymbol{Z} \sim N\left(\hat{\boldsymbol{\beta}}, \sigma^2 \left(X^T X\right)^{-1}\right)$, so when $\left(X^T X\right)^{-1}$ exists, it follows that

$$\int_{\mathbb{R}^k} \exp\left(-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)^T \left(X^T X\right)\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right) \mathrm{d}\boldsymbol{\beta} \propto \left(\sigma^2\right)^{\frac{k}{2}}.$$

Similarly, the integrand in (14) can be expressed as the expectation of the square root of a quadratic form involving the random variable $\boldsymbol{Z}$ from above. By using auxiliary fact (iii) and Jensen's inequality, we can write

$$\int_{\mathbb{R}^k} \sqrt{h_2(\rho, \boldsymbol{\beta})} \exp\left(-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)^T \left(X^T X\right)\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right) \mathrm{d}\boldsymbol{\beta}$$

$$\propto \left(\sigma^2\right)^{\frac{k}{2}} \int_{\mathbb{R}^k} \sqrt{\boldsymbol{z}^T X^T B_\rho^T M B_\rho X \boldsymbol{z}} \left((2\pi)^k \left|\sigma^2 \left(X^T X\right)^{-1}\right|\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(\boldsymbol{z} - \hat{\boldsymbol{\beta}}\right)^T \left(X^T X\right)\left(\boldsymbol{z} - \hat{\boldsymbol{\beta}}\right)\right) \mathrm{d}\boldsymbol{z}$$

$$= \left(\sigma^2\right)^{\frac{k}{2}} \mathbb{E}_{\boldsymbol{Z}}\left[\sqrt{\boldsymbol{Z}^T X^T B_\rho^T M B_\rho X \boldsymbol{Z}}\right] \leq \left(\sigma^2\right)^{k/2} \sqrt{\mathbb{E}_{\boldsymbol{Z}}\left[\boldsymbol{Z}^T X^T B_\rho^T M B_\rho X \boldsymbol{Z}\right]}$$

$$= \left(\sigma^2\right)^{\frac{k}{2}} \sqrt{\operatorname{tr}\left(X^T B_\rho^T M B_\rho X \sigma^2 \left(X^T X\right)^{-1}\right) + \hat{\boldsymbol{\beta}}^T X^T B_\rho^T M B_\rho X \hat{\boldsymbol{\beta}}}$$

$$= \left(\sigma^2\right)^{\frac{k}{2}} \sqrt{\sigma^2 \operatorname{tr}\left(B_\rho^T M B_\rho P\right) + \hat{\boldsymbol{\beta}}^T X^T B_\rho^T M B_\rho X \hat{\boldsymbol{\beta}}}$$

$$\leq \left(\sigma^2\right)^{\frac{k}{2}} \left(\left(\sigma^2\right)^{\frac{1}{2}} \sqrt{\operatorname{tr}\left(B_\rho^T M B_\rho P\right)} + \sqrt{\hat{\boldsymbol{\beta}}^T X^T B_\rho^T M B_\rho X \hat{\boldsymbol{\beta}}}\right),$$

where $P := X\left(X^T X\right)^{-1} X^T$, which is by definition symmetric, idempotent, and positive semi-definite, so $\operatorname{tr}\left(B_\rho^T M B_\rho P\right) \geq 0$ by auxiliary fact (iv). Combining these observations with

(13) and (14), it follows that

$$\left(\sigma^2\right)^{-\frac{g+k}{2}-1} |A_\rho| \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{y}^T A_\rho^T M A_\rho \boldsymbol{y}\right)$$
$$\int_{\mathbb{R}^k} \left(\sqrt{h_1(\rho)} + \sqrt{\frac{1}{\sigma^2} h_2(\rho, \boldsymbol{\beta})}\right) \exp\left(-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)^T \left(X^T X\right)\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right) \mathrm{d}\boldsymbol{\beta}$$
$$\leq \left(\sigma^2\right)^{-\frac{g+k}{2}-1} |A_\rho| \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{y}^T A_\rho^T M A_\rho \boldsymbol{y}\right)$$
$$\left(\sqrt{h_1(\rho)} \left(\sigma^2\right)^{\frac{k}{2}} + \left(\sigma^2\right)^{-\frac{1}{2}} \left(\sigma^2\right)^{\frac{k}{2}} \left(\left(\sigma^2\right)^{\frac{1}{2}} \sqrt{\mathrm{tr}\left(B_\rho^T M B_\rho P\right)} + \sqrt{\hat{\boldsymbol{\beta}}^T X^T B_\rho^T M B_\rho X \hat{\boldsymbol{\beta}}}\right)\right)$$
$$= \left(\sigma^2\right)^{-\frac{g}{2}-1} |A_\rho| \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{y}^T A_\rho^T M A_\rho \boldsymbol{y}\right) \left(\sqrt{h_1(\rho)} + \sqrt{\mathrm{tr}\left(B_\rho^T M B_\rho P\right)}\right) \tag{15}$$
$$+ \left(\sigma^2\right)^{-\frac{g+1}{2}-1} |A_\rho| \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{y}^T A_\rho^T M A_\rho \boldsymbol{y}\right) \sqrt{\hat{\boldsymbol{\beta}}^T X^T B_\rho^T M B_\rho X \hat{\boldsymbol{\beta}}}. \tag{16}$$

Next, observe that the terms involving σ^2 in (15) and (16) correspond to kernels of an inverse gamma distributed random variable, so integrating over σ^2 yields

$$\int_0^\infty \left(\sigma^2\right)^{-\frac{g}{2}-1} \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{y}^T A_\rho^T M A_\rho \boldsymbol{y}\right) \mathrm{d}\sigma^2 \propto \left(\boldsymbol{y}^T A_\rho^T M A_\rho \boldsymbol{y}\right)^{-\frac{g}{2}}, \tag{17}$$

$$\int_0^\infty \left(\sigma^2\right)^{-\frac{g+1}{2}-1} \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{y}^T A_\rho^T M A_\rho \boldsymbol{y}\right) \mathrm{d}\sigma^2 \propto \left(\boldsymbol{y}^T A_\rho^T M A_\rho \boldsymbol{y}\right)^{-\frac{g+1}{2}}, \tag{18}$$

where the terms on the right hand side in (17) and (18) are bounded on $\left(\lambda_g^{-1}, \lambda_1^{-1}\right)$ by assumption. Hence, it only remains to show that

$$|A_\rho| \left(\sqrt{h_1(\rho)} + \sqrt{\mathrm{tr}\left(B_\rho^T M B_\rho P\right)} + \sqrt{\hat{\boldsymbol{\beta}}^T X^T B_\rho^T M B_\rho X \hat{\boldsymbol{\beta}}}\right)$$

is bounded for $\rho \in \left(\lambda_g^{-1}, \lambda_1^{-1}\right)$. First, applying auxiliary fact (v) yields

$$\begin{aligned}\sqrt{h_1(\rho)} &= \sqrt{\mathrm{tr}\left(B_\rho^T B_\rho\right) + \mathrm{tr}\left(B_\rho^2\right)} \leq \sqrt{\mathrm{tr}\left(B_\rho^T B_\rho\right)} + \sqrt{\mathrm{tr}\left(B_\rho^2\right)} \\ &\leq \sqrt{\frac{1}{2}\left(\mathrm{tr}\left(B_\rho^T B_\rho^T\right) + \mathrm{tr}\left(B_\rho^2\right)\right)} + \sqrt{\mathrm{tr}\left(B_\rho^2\right)} \propto \sqrt{\mathrm{tr}\left(B_\rho^2\right)} \\ &= \sqrt{\sum_{i=1}^g \frac{\lambda_i^2}{\left(1-\rho\lambda_i\right)^2}} \leq \sum_{i=1}^g \sqrt{\frac{\lambda_i^2}{\left(1-\rho\lambda_i\right)^2}} \leq \sum_{i=1}^g \frac{|\lambda_i|}{1-\rho\lambda_i}.\end{aligned}$$

Second, using auxiliary fact (iv) and (vi), it holds that

$$\begin{aligned}\sqrt{\mathrm{tr}\left(B_\rho^T M B_\rho P\right)} &\leq \sqrt{\mathrm{tr}\left(B_\rho^T M B_\rho\right)\mathrm{tr}\left(P\right)} \propto \sqrt{\mathrm{tr}\left(B_\rho^T M B_\rho\right)} = \sqrt{\mathrm{tr}\left(B_\rho B_\rho^T M\right)} \\ &\leq \sqrt{\mathrm{tr}\left(B_\rho B_\rho^T\right)\mathrm{tr}\left(M\right)} \propto \sqrt{\mathrm{tr}\left(B_\rho^T B_\rho\right)} \leq \sum_{i=1}^g \frac{|\lambda_i|}{1-\rho\lambda_i},\end{aligned} \tag{19}$$

where (19) follows from the considerations above and the idempotence of M and P. Finally, with auxiliary facts (iv), (v), (vi), and after some algebraic manipulations we can write

$$\begin{aligned}
&\sqrt{\hat{\boldsymbol{\beta}}^T X^T B_\rho^T M B_\rho X \hat{\boldsymbol{\beta}}} = \sqrt{\left(\left(X^T X\right)^{-1} X^T A_\rho \boldsymbol{y}\right)^T X^T B_\rho^T M B_\rho X \left(X^T X\right)^{-1} X^T A_\rho \boldsymbol{y}} \\
&= \sqrt{\operatorname{tr}\left(B_\rho^T M B_\rho P A_\rho \boldsymbol{y}\boldsymbol{y}^T A_\rho^T P\right)} = \sqrt{\operatorname{tr}\left(B_\rho^T M B_\rho P A_\rho \boldsymbol{y}\boldsymbol{y}^T A_\rho^T P^T\right)} \\
&\leq \sqrt{\operatorname{tr}\left(B_\rho^T M B_\rho\right) \operatorname{tr}\left(\left(\boldsymbol{y}^T A_\rho^T P^T\right)^T \boldsymbol{y}^T A_\rho^T P^T\right)} = \sqrt{\operatorname{tr}\left(B_\rho B_\rho^T M\right) \boldsymbol{y}^T A_\rho^T P A_\rho \boldsymbol{y}} \\
&\leq \sqrt{\operatorname{tr}\left(B_\rho B_\rho^T\right) \operatorname{tr}\left(M\right) \boldsymbol{y}^T A_\rho^T P A_\rho \boldsymbol{y}} \propto \sqrt{\operatorname{tr}\left(B_\rho^T B_\rho\right) \boldsymbol{y}^T A_\rho^T P A_\rho \boldsymbol{y}} \\
&= \sqrt{\operatorname{tr}\left(B_\rho^T B_\rho\right)} \sqrt{\left(\boldsymbol{y}^T A_\rho^T P A_\rho \boldsymbol{y}\right)}, \qquad (20)
\end{aligned}$$

As $\boldsymbol{y}^T A_\rho^T P A_\rho \boldsymbol{y}$ is bounded for $\rho \in \left(\lambda_g^{-1}, \lambda_1^{-1}\right)$, the expression in (20) can be bounded again by a multiple of the sum term in (19). Furthermore, if m_1 and m_g denote the algebraic multiplicity of λ_1 and λ_g, respectively, then

$$\begin{aligned}
|A_\rho| \sum_{i=1}^{g} \frac{|\lambda_i|}{1-\rho\lambda_i} &= \left(\prod_{i=1}^{g} (1-\rho\lambda_i)\right) \left(\sum_{i=1}^{g} \frac{|\lambda_i|}{1-\rho\lambda_i}\right) \\
&= \left(\prod_{i=1}^{g} (1-\rho\lambda_i)\right) \frac{|\lambda_1|}{1-\rho\lambda_1} + \dots + \left(\prod_{i=1}^{g} (1-\rho\lambda_i)\right) \frac{|\lambda_g|}{1-\rho\lambda_g} \\
&= |\lambda_1| \left(\prod_{i=m_1+1}^{g} (1-\rho\lambda_i)\right) + \dots + |\lambda_g| \left(\prod_{i=g-m_g}^{g} (1-\rho\lambda_i)\right) \qquad (21) \\
&< \infty,
\end{aligned}$$

as every summand in (21) is bounded for $\rho \in \left(\lambda_g^{-1}, \lambda_1^{-1}\right)$. This completes the proof.

Proof of Theorem 2

The model's Fisher Information Matrix in (10) gives

$$\begin{aligned}
&\det\left(I_{(\rho,\sigma^2),(\rho,\sigma^2)}\left(\boldsymbol{\theta}\right)\right) = \left(\sigma^2\right)^{-2} \left(I^*_{\rho,\rho} I^*_{\sigma^2,\sigma^2} - {I^*_{\rho,\sigma^2}}^2\right) \\
&= \left(\sigma^2\right)^{-2} \left(\frac{g}{2\sigma^2}\left(\sigma^2\left(\operatorname{tr}\left(B_\rho^T B_\rho\right) + \operatorname{tr}\left(B_\rho^2\right)\right) + \boldsymbol{\beta}^T X^T B_\rho^T B_\rho X \boldsymbol{\beta}\right) - \operatorname{tr}^2\left(B_\rho\right)\right) \\
&\propto \left(\sigma^2\right)^{-2} \left(\operatorname{tr}\left(B_\rho^T B_\rho\right) + \operatorname{tr}\left(B_\rho^2\right) + \frac{1}{\sigma^2}\boldsymbol{\beta}^T X^T B_\rho^T B_\rho X \boldsymbol{\beta} - \frac{2}{g}\operatorname{tr}^2\left(B_\rho\right)\right),
\end{aligned}$$

and $\det\left(I_{(\boldsymbol{\beta},\boldsymbol{\beta})}\left(\boldsymbol{\theta}\right)\right) \propto 1$. The result follows from these observations and by the definition of the Independence Jeffreys prior.

Proof of Corollary 3

These results are proved in an identical way as the ones in Corollary 2 and follow almost immediately.

Posterior sampling

We outlined the sampling procedure and gave the conditional posteriors based on the flat and the informative priors in Section 5. However, it remains to specify the exact forms of the candidate-generating distributions for the conditional posteriors of the parameter blocks (ρ, β_1) and $\widetilde{\boldsymbol{\beta}}$. As to the conditional posterior of (ρ, β_1), we first approximated $\log(|A_\rho|)$ by a second-order Taylor polynomial at $\rho = 0$, so $|A_\rho| \approx \exp\left(-\rho^2 \sum_{i=1}^{g} \lambda_i^2/2\right)$. Using this approximation, we can write

$$
\begin{aligned}
&p_{\mathrm{F}}\left((\rho, \beta_1)\,|\sigma^2, \widetilde{\boldsymbol{\beta}}, \boldsymbol{y}\right) \propto |A_\rho| \exp\left(-\frac{1}{2\sigma^2}\left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)^T \left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)\right) \\
&\approx \exp\left(-\frac{\rho^2}{2}\sum_{i=1}^{g}\lambda_i^2\right) \exp\left(-\frac{1}{2\sigma^2}\left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)^T \left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)\right) \\
&\propto \exp\left(-\frac{\rho^2}{2}\sum_{i=1}^{g}\lambda_i^2 - \frac{1}{2\sigma^2}\left(\rho^2 \boldsymbol{y}^T W^T W \boldsymbol{y} - 2\rho \boldsymbol{y}^T W^T \left(\boldsymbol{y} - \widetilde{X}\widetilde{\boldsymbol{\beta}}\right) + 2\rho\beta_1 \boldsymbol{y}^T W^T \mathbf{1}_g - 2\beta_1 \mathbf{1}_g^T \left(\boldsymbol{y} - \widetilde{X}\widetilde{\boldsymbol{\beta}}\right) + \beta_1^2 g\right)\right),
\end{aligned}
\tag{22}
$$

where the proportionality holds with respect to (ρ, β_1), $\widetilde{X}$ denotes the matrix X with its first column removed, $\widetilde{\boldsymbol{\beta}} = (\beta_2, ..., \beta_g)$, and $\mathbf{1}_g$ is the vector of ones of length g. The expression in (22) corresponds to the kernel of a bivariate normal distribution $q(\rho, \boldsymbol{\beta}_1) \sim N(\boldsymbol{\mu}, \Sigma)$, with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. By equating coefficients and after some algebraic manipulation, it follows that

$$
\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \frac{\mathbf{1}_g^T\left(\left(\boldsymbol{y} - \widetilde{X}\widetilde{\boldsymbol{\beta}}\right)\boldsymbol{y}^T W^T \mathbf{1}_g - \mathbf{1}_g \boldsymbol{y}^T W^T \left(\boldsymbol{y} - \widetilde{X}\widetilde{\boldsymbol{\beta}}\right)\right)}{\left(\boldsymbol{y}^T W^T \mathbf{1}_g\right)^2 - g\left(\sigma^2 \sum_{i=1}^{g} \lambda_i^2 + \boldsymbol{y}^T W^T W \boldsymbol{y}\right)} \\ \frac{\boldsymbol{y}^T W^T \left(\boldsymbol{y} - \widetilde{X}\widetilde{\boldsymbol{\beta}}\right) - \mu_1\left(\sigma^2 \sum_{i=1}^{g} \lambda_i^2 + \boldsymbol{y}^T W^T W \boldsymbol{y}\right)}{\boldsymbol{y}^T W^T \mathbf{1}_g} \end{pmatrix}, \tag{23}
$$

$$
\Sigma = \sigma^2 \begin{pmatrix} \sigma^2 \sum_{i=1}^{g} \lambda_i^2 + \boldsymbol{y}^T W^T W \boldsymbol{y} & \boldsymbol{y}^T W^T \mathbf{1}_g \\ \boldsymbol{y}^T W^T \mathbf{1}_g & g \end{pmatrix}^{-1}. \tag{24}
$$

We used this candidate-generating distribution also for the conditional posterior of (ρ, β_1) based on Jeffreys rule and Independence Jeffreys prior, as for these priors the prior information for (ρ, β_1) is quite vague compared to the likelihood. Note that due to the chosen parameter space for ρ, $q(\rho, \beta_1)$ is in fact truncated to $\left(\lambda_g^{-1}, \lambda_1^{-1}\right) \times \mathbb{R}$. In the simulation study, we relied on the "rtmvnorm" function from the tmvtnorm package in R to sample from this distribution (Wilhelm and Manjunath, 2015). Similarly, we can obtain the corresponding mean vector and covariance matrix of the candidate-generating bivariate normal density for (ρ, β_1) when using a normal prior for ρ.

The conditional posterior of $\widetilde{\boldsymbol{\beta}}$ based on the flat and the informative priors is a multivariate normal distribution and can be directly sampled from. We used the "rmvnorm" function from the mvtnorm package in R therefor (Genz et al., 2014). Its mean vector and covariance matrix are given by

$$
\begin{aligned}
\boldsymbol{\mu}_{\widetilde{\boldsymbol{\beta}}} &= \boldsymbol{\mu}_{\boldsymbol{\beta}_2} + \Sigma_{\boldsymbol{\beta}_{21}} \Sigma_{\boldsymbol{\beta}_{11}}^{-1} \left(\beta_1 - \boldsymbol{\mu}_{\boldsymbol{\beta}_1}\right), \\
\Sigma_{\widetilde{\boldsymbol{\beta}}} &= \Sigma_{\boldsymbol{\beta}_{22}} - \Sigma_{\boldsymbol{\beta}_{21}} \Sigma_{\boldsymbol{\beta}_{11}}^{-1} \Sigma_{\boldsymbol{\beta}_{12}},
\end{aligned}
$$

where

$$\boldsymbol{\mu}_{\boldsymbol{\beta}} = \left(X^T X\right)^{-1} X^T A_\rho \boldsymbol{y} = \begin{pmatrix} \boldsymbol{\mu}_{\boldsymbol{\beta}_1} \\ \boldsymbol{\mu}_{\boldsymbol{\beta}_2} \end{pmatrix} \text{ with sizes } \begin{pmatrix} 1 \times 1 \\ (k-1) \times 1 \end{pmatrix}, \tag{25}$$

$$\Sigma_{\boldsymbol{\beta}} = \sigma^2 \left(X^T X\right)^{-1} = \begin{pmatrix} \Sigma_{\boldsymbol{\beta}_{11}} & \Sigma_{\boldsymbol{\beta}_{12}} \\ \Sigma_{\boldsymbol{\beta}_{21}} & \Sigma_{\boldsymbol{\beta}_{22}} \end{pmatrix} \text{ with sizes } \begin{pmatrix} 1 \times 1 & 1 \times (k-1) \\ (k-1) \times 1 & (k-1) \times (k-1) \end{pmatrix}. \tag{26}$$

For the same reasons as before, we used this candidate-generating distribution also for the conditional posterior of $\widetilde{\boldsymbol{\beta}}$ based on Jeffreys rule and Independence Jeffreys prior.

Combining (2) and (5), the full conditionals based on Jeffreys rule prior can be written as

$$p_{\text{J}}\left((\rho, \beta_1) \,|\sigma^2, \widetilde{\boldsymbol{\beta}}, \boldsymbol{y}\right) \propto |A_\rho| \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}\right)\left(\text{tr}\left(B_\rho^T B_\rho\right) + \text{tr}\left(B_\rho^2\right) + \frac{1}{\sigma^2}\boldsymbol{\beta}^T X^T B_\rho^T M B_\rho X \boldsymbol{\beta} - \frac{2}{g}\text{tr}^2\left(B_\rho\right)\right)^{\frac{1}{2}},$$

$$p_{\text{J}}\left(\sigma^2 | \,(\rho, \beta_1)\,, \widetilde{\boldsymbol{\beta}}, \boldsymbol{y}\right) \propto \left(\sigma^2\right)^{-\frac{g+k}{2}-1} \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}\right)\left(\text{tr}\left(B_\rho^T B_\rho\right) + \text{tr}\left(B_\rho^2\right) + \frac{1}{\sigma^2}\boldsymbol{\beta}^T X^T B_\rho^T M B_\rho X \boldsymbol{\beta} - \frac{2}{g}\text{tr}^2\left(B_\rho\right)\right)^{\frac{1}{2}},$$

$$p_{\text{J}}\left(\widetilde{\boldsymbol{\beta}} | \,(\rho, \beta_1)\,, \sigma^2, \boldsymbol{y}\right) \propto \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}\right)\left(\text{tr}\left(B_\rho^T B_\rho\right) + \text{tr}\left(B_\rho^2\right) + \frac{1}{\sigma^2}\boldsymbol{\beta}^T X^T B_\rho^T M B_\rho X \boldsymbol{\beta} - \frac{2}{g}\text{tr}^2\left(B_\rho\right)\right)^{\frac{1}{2}},$$

where $\boldsymbol{\varepsilon} = A_\rho \boldsymbol{y} - X\boldsymbol{\beta}$. As none of the full conditionals is of closed form, an additional Metropolis-Hastings step for each parameter (block) is needed. The candidate-generating densities for the conditional posteriors of (ρ, β_1) and $\widetilde{\boldsymbol{\beta}}$ have already been given above, while we propose

$$q_{\text{J}}\left(\sigma^2 | \,(\rho, \beta_1)\,, \widetilde{\boldsymbol{\beta}}, \boldsymbol{y}\right) \sim IG\left(\frac{g+k+1}{2}, \frac{\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}}{2}\right)$$

as candidate-generating density for $p_{\text{J}}\left(\sigma^2 | \,(\rho, \beta_1)\,, \widetilde{\boldsymbol{\beta}}, \boldsymbol{y}\right)$ which resulted in well-mixed Markov chains. Using the same steps, we can also easily derive the conditional posteriors based on the Independence Jeffreys prior, where we employed

$$q_{\text{IJ}}\left(\sigma^2 | \,(\rho, \beta_1)\,, \widetilde{\boldsymbol{\beta}}, \boldsymbol{y}\right) \sim IG\left(\frac{g+1}{2}, \frac{\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}}{2}\right)$$

as corresponding candidate-generating density for $p_{\text{IJ}}\left(\sigma^2 | \,(\rho, \beta_1)\,, \widetilde{\boldsymbol{\beta}}, \boldsymbol{y}\right)$.

Subsequently, when drawing from the candidate-generating distributions in the Metropolis-Hastings algorithm only a subset of the proposals is accepted. This subset is based on a probability rule that ensures that the sequence of accepted draws can be, after some appropriate number of steps, regarded as a sample from the target distribution (Chib and Greenberg, 1995). We outline the full sampling algorithm based on using the flat prior in the following:

1. Set starting values $(\rho, \beta_1)^{(0)}, {\sigma^2}^{(0)}$, and $\widetilde{\boldsymbol{\beta}}^{(0)}$, e.g. to their maximum likelihood estimates.
2. Use a Metropolis-Hastings step to draw $(\rho, \beta_1)^{(1)}$ from the (truncated) normal distribution based on (23), (24), given the values of ${\sigma^2}^{(0)}$ and $\widetilde{\boldsymbol{\beta}}^{(0)}$.

3. Sample ${\sigma^2}^{(1)}$ from the inverse gamma distribution in (8), given the values of $(\rho, \beta_1)^{(1)}$ and $\widetilde{\boldsymbol{\beta}}^{(0)}$.
4. Sample $\widetilde{\boldsymbol{\beta}}^{(1)}$ from the multivariate normal distribution based on (25), (26), given the values of $(\rho, \beta_1)^{(1)}$ and ${\sigma^2}^{(1)}$.
5. Repeat Steps (1) through (4) for $s = 1, ..., S$.

Note that when using Jeffreys rule prior or Independence Jeffreys prior, the direct sampling procedures in (3) and (4) above are replaced by Metropolis-Hastings steps based on the corresponding candidate-generating densities.

References

Andersson, H., Jonsson, L., and Ögren, M. (2010). Property prices and exposure to multiple noise sources: Hedonic regression with road and railway noise. *Environmental and Resource Economics*, 45(1):73–89.

Anselin, L. (1982). A note on small sample properties of estimators in a first-order spatial autoregressive model. *Environment and Planning A*, 14(8):1023–1030.

Anselin, L. (1984). Specification tests on the structure of interaction in spatial econometric models. *Papers of the Regional Science Association*, 54(1):165–182.

Anselin, L. (1990). Some robust approaches to testing and estimation in spatial econometrics. *Regional Science and Urban Economics*, 20(2):141–163.

Anselin, L. (2002). Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27(3):247–267.

Anselin, L. and Le Gallo, J. (2006). Interpolation of air quality measures in hedonic house price models: Spatial aspects. *Spatial Economic Analysis*, 1(1):31–52.

Anselin, L. and Lozano-Gracia, N. (2008). Errors in variables and spatial effects in hedonic house price models of ambient air quality. *Empirical Economics*, 34(1):5–34.

Anselin, L., Lozano-Gracia, N., Deichmann, U., and Lall, S. (2010). Valuing access to water - a spatial hedonic approach, with an application to bangalore, india. *Spatial Economic Analysis*, 5(2):161–179.

Anselin, L., Varga, A., and Acs, Z. (2000). Geographical and sectoral characteristics of academic knowledge externalities. *Papers in Regional Science*, 79(4):435–443.

Arbia, G. and Basile, R. (2005). Spatial dependence and non-linearities in regional growth behaviour in italy. *Statistica*, 65(2):145–167.

Armstrong, R. J. and Rodriguez, D. A. (2006). An evaluation of the accessibility benefits of commuter rail in eastern massachusetts using spatial hedonic price functions. *Transportation*, 33(1):21–43.

Baller, R. D., Anselin, L., Messner, S. F., Deane, G., and Hawkins, D. F. (2001). Structural covariates of u.s. county homicide rates: Incorporating spatial effects. *Criminology*, 39(3):561–588.

Beck, N., Gleditsch, K. S., and Beardsley, K. (2006). Space is more than geography: Using spatial econometrics in the study of political economy. *International Studies Quarterly*, 50(1):27–44.

Berger, J. (2006). The case for objective bayesian analysis. *Bayesian Analysis*, 1(3):385–402.

Berger, J. O., de Oliveira, V., and Sanso, B. (2001). Objective bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, 96(456):1361–1374.

Bernardo, J. M. (1979). Reference posterior distributions for bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(2):113–147.

Bernat, G. A. (1996). Does manufacturing matter? a spatial econometric view of kaldor's laws. *Journal of Regional Science*, 36(3):463–477.

Bivand, R. and Piras, G. (2015). Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software*, 63(1):1–36.

Bivand, R. and Szymanski, S. (2000). Modelling the spatial impact of the introduction of compulsory competitive tendering. *Regional Science and Urban Economics*, 30(2):203–219.

Bolstad, W. M. (2009). *Understanding Computational Bayesian Statistics*. John Wiley & Sons.

Bordignon, M., Cerniglia, F., and Revelli, F. (2003). In search of yardstick competition: a spatial analysis of italian municipality property tax setting. *Journal of Urban Economics*, 54(2):199–217.

Box, G. E. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. John Wiley & Sons.

Brooks, S. P. (1998). Markov chain monte carlo method and its application. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 47(1):69–100.

Brueckner, J. K. and Saavedra, L. A. (2001). Do local governments engage in strategic property-tax competition? *National Tax Journal*, 54(2):203–229.

Buonanno, P., Montolio, D., and Vanin, P. (2009). Does social capital reduce crime? *Journal of Law and Economics*, 52(1):145–170.

Burt, R. S. and Doreian, P. (1982). Testing a structural model of perception: Conformity and deviance with respect to journal norms in elite sociological methodology. *Quality and Quantity*, 16(2):109–150.

Butts, C. T. (2008). Social network analysis with sna. *Journal of Statistical Software*, 24(6):1–51.

Can, A. (1992). Specification and estimation of hedonic housing price models. *Regional Science and Urban Economics*, 22(3):453–474.

Carruthers, J. I. and Clark, D. E. (2010). Valuing environmental quality: A space-based strategy. *Journal of Regional Science*, 50(4):801–832.

Chance, B. L. and Rossman, A. J. (2006). *Investigating Statistical Concepts, Applications, and Methods*. Duxbury.

Chang, H. (2008). Spatial analysis of water quality trends in the han river basin, south korea. *Water Research*, 42(13):3285–3304.

Chib, S. and Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327–335.

Chib, S. and Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, 85(2):347–361.

Cohen, J. P. and Coughlin, C. C. (2008). Spatial hedonic models of airport noise, proximity, and housing prices. *Journal of Regional Science*, 48(5):859–878.

Conway, D., Li, C. Q., Wolch, J., Kahle, C., and Jerrett, M. (2010). A spatial autocorrelation approach for examining the effects of urban greenspace on residential property values. *The Journal of Real Estate Finance and Economics*, 41(2):150–169.

Dallerba, S. (2005). Productivity convergence and spatial dependence among spanish regions. *Journal of Geographical Systems*, 7(2):207–227.

De Oliveira, V. (2010). Objective bayesian analysis for gaussian random fields. In Chen, M.-H., Müller, P., Sun, D., Ye, K., and Dey, D. K., editors, *Frontiers of Statistical Decision Making and Bayesian Analysis: In Honor of James O. Berger*, pages 497–511. Springer.

De Oliveira, V. (2012). Bayesian analysis of conditional autoregressive models. *Annals of the Institute of Statistical Mathematics*, 64(1):107–133.

De Oliveira, V. and Song, J. J. (2008). Bayesian analysis of simultaneous autoregressive models. *Sankhya: The Indian Journal of Statistics, Series B (2008-)*, 70(2):323–350.

DeGroot, M. H. and Schervish, M. J. (2010). *Probability and Statistics*. Addison-Wesley.

Doreian, P. (1980). Linear models with spatially distributed data: Spatial disturbances or spatial effects? *Sociological Methods & Research*, 9(1):29–60.

Doreian, P. (1981). Estimating linear models with spatially distributed data. *Sociological Methodology*, 12:359–388.

Dow, M. M. (2007). Galton's problem as multiple network autocorrelation effects: Cultural trait transmission and ecological constraint. *Cross-Cultural Research*, 41(4):336–363.

Dow, M. M., Burton, M. L., and White, D. R. (1982a). Network autocorrelation: A simulation study of a foundational problem in regression and survey research. *Social Networks*, 4(2):169–200.

Dow, M. M., White, D. R., and Burton, M. L. (1982b). Multivariate modeling with interdependent network data. *Cross-Cultural Research*, 17(3 & 4):216–245.

Easterly, W. and Levine, R. (1998). Troubles with the neighbours: Africa's problem, africa's opportunity. *Journal of African Economies*, 7(1):120–142.

Elhorst, J. P. (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer.

Ertur, C., Le Gallo, J., and LeSage, J. P. (2007). Local versus global convergence in europe: A bayesian spatial econometric approach. *The Review of Regional Studies*, 37(1):82–108.

Fernandez, C., Ley, E., and Steel, M. F. (2001). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics*, 16(5):563–576.

Fingleton, B. (2001). Theoretical economic geography and spatial econometrics: dynamic perspectives. *Journal of Economic Geography*, 1(2):201–225.

Fingleton, B., Igliori, D., and Moore, B. (2005). Cluster dynamics: New evidence and projections for computing services in great britain. *Journal of Regional Science*, 45(2):283–311.

Fingleton, B. and Le Gallo, J. (2008). Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: Finite sample properties. *Papers in Regional Science*, 87(3):319–339.

Florax, R. J., Voortman, R. L., and Brouwer, J. (2002). Spatial dimensions of precision agriculture: a spatial econometric analysis of millet yield on sahelian coversands. *Agricultural Economics*, 27(3):425–443.

Ford, T. C. and Rork, J. C. (2010). Why buy what you can get for free? the effect of foreign direct investment on state patent rates. *Journal of Urban Economics*, 68(1):72–81.

Fornango, R. J. (2010). When space matters: Spatial dependence, diagnostics, and regression models. *Journal of Criminal Justice Education*, 21(2):117–135.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–533.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis, Second Edition*. Chapman & Hall/CRC Texts in Statistical Science.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2014). *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.0-2.

Gimpel, J. and Schuknecht, J. (2003). Political participation and the accessibility of the ballot box. *Political Geography*, 22(5):471–488.

Gould, R. V. (1991). Multiple networks and mobilization in the paris commune, 1871. *American Sociological Review*, 56(6):716–729.

Greenbaum, R. T. (2002). A spatial study of teachers' salaries in pennsylvania school districts. *Journal of Labor Research*, 23(1):69–86.

Griffith, D. A. (1979). Urban dominance, spatial structure, and spatial dynamics: Some theoretical conjectures and empirical implications. *Economic Geography*, 55(2):95–113.

Han, X. and Lee, L.-f. (2013). Bayesian estimation and model selection for spatial durbin error model with finite distributed lags. *Regional Science and Urban Economics*, 43(5):816–837.

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

Heikkila, E. J. and Kantiotou, C. (1992). Calculating fiscal impacts where spatial effects are present. *Regional Science and Urban Economics*, 22(3):475–490.

Hepple, L. W. (1979). Bayesian analysis of the linear model with spatial dependence. In Bartels, C. P. and Ketellapper, R., editors, *Exploratory and explanatory statistical analysis of spatial data*, pages 179–199. Springer Netherlands.

Hepple, L. W. (1995a). Bayesian techniques in spatial and network econometrics: 1. model comparison and posterior odds. *Environment and Planning A*, 27(3):447–469.

Hepple, L. W. (1995b). Bayesian techniques in spatial and network econometrics: 2. computational methods and algorithms. *Environment and Planning A*, 27(4):615–644.

Heyndels, B. and Vuchelen, J. (1998). Tax mimicking among belgian municipalities. *National Tax Journal*, 51(1):89–101.

Holloway, G., Shankar, B., and Rahman, S. (2002). Bayesian spatial probit estimation: a primer and an application to hyv rice adoption. *Agricultural Economics*, 27(3):383–402.

Hunt, L. M., Boxall, P., Englin, J., and Haider, W. (2005). Remote tourism and forest management: a spatial hedonic analysis. *Ecological Economics*, 53(1):101–113.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 186(1007):453–461.

Jeffreys, H. (1961). *Theory of Probability*. Oxford, third edition.

Joines, J. D., Hertz-Picciotto, I., Carey, T. S., Gesler, W., and Suchindran, C. (2003). A spatial analysis of county-level variation in hospitalization rates for low back problems in north carolina. *Social Science & Medicine*, 56(12):2541–2553.

Kalenkoski, C. M. and Lacombe, D. J. (2008). Effects of minimum wages on youth employment: the importance of accounting for spatial correlation. *Journal of Labor Research*, 29(4):303–317.

Kalnins, A. (2003). Hamburger prices and spatial econometrics. *Journal of Economics & Management Strategy*, 12(4):591–616.

Kass, R. E. and Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(435):1343–1370.

Kim, J. and Goldsmith, P. (2009). A spatial hedonic approach to assess the impact of swine production on residential property values. *Environmental and Resource Economics*, 42(4):509–534.

Kim, J. and Zhang, M. (2005). Determining transits impact on seoul commercial land values: An application of spatial econometrics. *International Real Estate Review*, 8(1):1–26.

Kirk, D. S. and Papachristos, A. V. (2011). Cultural mechanisms and the persistence of neighborhood violence. *American Journal of Sociology*, 116(4):1190–1233.

Land, K. C., Deane, G., and Blau, J. R. (1991). Religious pluralism and church membership: A spatial diffusion model. *American Sociological Review*, 56(2):237–249.

Lauridsen, J., Maté Sánchez, M., and Bech, M. (2010). Public pharmaceutical expenditure: identification of spatial effects. *Journal of Geographical Systems*, 12(2):175–188.

Leenders, R. T. (1995). *Structure and Influence: Statistical models for the dynamics of actor attributes, network structure and their interdependence*. PhD thesis, Thesis Publishers Amsterdam.

Leenders, R. T. (2002). Modeling social influence through network autocorrelation: constructing the weight matrix. *Social Networks*, 24(1):21–47.

Leifeld, P., Cranmer, S. J., and Desmarais, B. A. (2015). *xergm: Extensions for Exponential Random Graph Models*. R package version 1.6.

LeSage, J. P. (1997a). Bayesian estimation of spatial autoregressive models. *International Regional Science Review*, 20(1-2):113–129.

LeSage, J. P. (1997b). Regression analysis of spatial data. *Journal of Regional Analysis and Policy*, 27(2):83–94.

LeSage, J. P. (1999). Spatial econometrics. `http://www.rri.wvu.edu/WebBook/LeSage/spatial/wbook.pdf`.

LeSage, J. P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis*, 32(1):19–35.

LeSage, J. P. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Statistics: A Series of Textbooks and Monographs. CRC Press.

LeSage, J. P. and Parent, O. (2007). Bayesian model averaging for spatial econometric models. *Geographical Analysis*, 39(3):241–267.

Levine, N., Kim, K. E., and Nitz, L. H. (1995). Spatial analysis of honolulu motor vehicle crashes: Ii. zonal generators. *Accident Analysis and Prevention*, 27(5):675–685.

Lin, X. (2010). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics*, 28(4):825–860.

Lu, J. and Zhang, L. (2011). Modeling and prediction of tree heightdiameter relationships using spatial autoregressive models. *Forest Science*, 57(3):252–264.

Marsden, P. V. and Friedkin, N. E. (1993). Network studies of social influence. *Sociological Methods & Research*, 22(1):127–151.

McMillen, D. (2013). *McSpatial: Nonparametric spatial data analysis*. R package version 2.0.

McMillen, D. P. (2010). Issues in spatial data analysis. *Journal of Regional Science*, 50(1):119–141.

McMillen, D. P., Singell Jr., L. D., and Waddell, G. R. (2007). Spatial competition and the price of college. *Economic Inquiry*, 45(4):817–833.

McPherson, M. A. and Nieswiadomy, M. L. (2005). Environmental kuznets curve: threatened species and spatial effects. *Ecological Economics*, 55(3):395–407.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092.

Mizruchi, M. S. and Neuman, E. J. (2008). The effect of density on the level of bias in the network autocorrelation model. *Social Networks*, 30(3):190–200.

Moreno, R. and Trehan, B. (1997). Location and the growth of nations. *Journal of Economic Growth*, 2(4):399–418.

Morenoff, J. D. (2003). Neighborhood mechanisms and the spatial dynamics of birth weight. *American Journal of Sociology*, 108(5):976–1017.

Mur, J., López, F., and Angulo, A. (2008). Symptoms of instability in models of spatial dependence. *Geographical Analysis*, 40(2):189–211.

Neuman, E. J. and Mizruchi, M. S. (2010). Structure and bias in the network autocorrelation model. *Social Networks*, 32(4):290–300.

Niebuhr, A. (2010). Migration and innovation: Does cultural diversity matter for regional r&d activity? *Papers in Regional Science*, 89(3):563–585.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70(349):120–126.

Osland, L. (2010). An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Research*, 32(3):289–320.

Patton, M. and McErlean, S. (2003). Spatial effects within the agricultural land market in northern ireland. *Journal of Agricultural Economics*, 54(1):35–54.

Pericchi, L. and Pereira, C. (2016). Adaptative significance levels using optimal decision rules: Balancing by weighting the error probabilities. *Brazilian Journal of Probability and Statistics*, 30(1):70–90.

Pisati, M. (2001). Tools for spatial data analysis. *Stata Technical Bulletin*, 60:21–48.

Plümper, T. and Neumayer, E. (2010). Model specification in the analysis of spatial dependence. *European Journal of Political Research*, 49(3):418–442.

Pons-Novell, J. and Elisabet, V.-M. (1999). Kaldors laws and spatial dependence: Evidence for the european regions. *Regional Studies*, 33(5):443–451.

Raftery, A. E. and Lewis, S. M. (1996). Implementing mcmc. In Gilks, W., Richardson, S., and Spiegelhalter, D., editors, *Markov Chain Monte Carlo in Practice*, pages 115–130. Chapman & Hall.

Revelli, F. (2003). Reaction or interaction? spatial process identification in multi-tired government structures. *Journal of Urban Economics*, 53(1):29–53.

Rey, S. J. and Anselin, L. (2007). Pysal: A python library of spatial analytical methods. *The Review of Regional Studies*, 37(1):5–27.

Rubio, F. J. and Steel, M. F. (2014). Inference in two-piece location-scale models with jeffreys priors. *Bayesian Analysis*, 9(1):1–22.

Ruggles, S. (2007). The decline of intergenerational coresidence in the united states, 1850 to 2000. *American Sociological Review*, 72(6):964–989.

Rupasingha, A., Goetz, S. J., and Freshwater, D. (2002). Social and institutional factors as determinants of economic growth: Evidence from the united states counties. *Papers in Regional Science*, 81(2):139–155.

Saavedra, L. A. (2000). A model of welfare competition with evidence from afdc. *Journal of Urban Economics*, 47(2):248–279.

Seldadyo, H., Elhorst, J. P., and De Haan, J. (2010). Geography and governance: Does space matter? *Papers in Regional Science*, 89(3):625–640.

Shin, M. and Ward, M. D. (1999). Lost in space: Political geography and the defense-growth trade-off. *The Journal of Conflict Resolution*, 43(6):793–817.

Smith, T. E. (2009). Estimation bias in spatial models with strongly connected weight matrices. *Geographical Analysis*, 41(3):307–332.

Stewart, W. J. (2009). *Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling*. Princeton University Press.

Tam Cho, W. K. (2003). Contagion effects and ethnic contribution networks. *American Journal of Political Science*, 47(2):368–387.

Tita, G. E. and Greenbaum, R. T. (2009). Crime, neighborhoods, and units of analysis: Putting space in its place. In Weisburd, D., Bernasco, W., and Bruinsma, G. J., editors, *Putting Crime in its Place*, pages 145–170. Springer New York.

Tita, G. E. and Radil, S. M. (2011). Spatializing the social networks of gangs to explore patterns of violence. *Journal of Quantitative Criminology*, 27(4):521–545.

Varga, A. (2000). Local academic knowledge transfers and the concentration of economic activity. *Journal of Regional Science*, 40(2):289–309.

Vega, S. H. and Elhorst, J. P. (2015). The slx model. *Journal of Regional Science*, 55(3):339–363.

Vitale, M. P., Porzio, G. C., and Doreian, P. (2016). Examining the effect of social influence on student performance through network autocorrelation models. *Journal of Applied Statistics*, 43(1):115–127.

Voss, P. R. and Chi, G. (2006). Highways and population change. *Rural Sociology*, 71(1):33–58.

Voss, P. R., Long, D. D., Hammer, R. B., and Friedman, S. (2006). County child poverty rates in the us: a spatial regression approach. *Population Research and Policy Review*, 25(4):369–391.

Wang, W., Neuman, E. J., and Newman, D. A. (2014). Statistical power of the social network autocorrelation model. *Social Networks*, 38:88–99.

Whitt, H. P. (2010). The civilizing process and its discontents: Suicide and crimes against persons in france, 18251830. *American Journal of Sociology*, 116(1):130–186.

Wilhelm, S. and Godinho de Matos, M. (2015). *spatialprobit: Spatial Probit Models*. R package version 0.9-11.

Wilhelm, S. and Manjunath, B. (2015). *tmvtnorm: Truncated Multivariate Normal and Student t Distribution*. R package version 1.4-10.

Wilhelmsson, M. (2002). Spatial models in real estate economics. *Housing, Theory and Society*, 19(2):92–101.

Won Kim, C., Phipps, T. T., and Anselin, L. (2003). Measuring the benefits of air quality improvement: a spatial hedonic approach. *Journal of Environmental Economics and Management*, 45(1):24–39.

Yang, N., McCluskey, J. J., and Brady, M. P. (2012). The value of good neighbors: A spatial analysis of the california and washington state wine industries. *Land Economics*, 88(4):674–684.

Yang, X. (2000). A matrix trace inequality. *Journal of Mathematical Analysis and Applications*, 250(1):372–374.