ESTIMATION OF STATISTICAL NETWORK AND REGION-WISE VARIABLE SELECTION

By

Sayan Chakraborty

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Statistics - Doctor of Philosophy

2016

ABSTRACT

ESTIMATION OF STATISTICAL NETWORK AND REGION-WISE VARIABLE SELECTION

By

Sayan Chakraborty

Network models are widely used to represent relations between actors or nodes. Recent studies of the network literature and graph model revealed various characteristics of the actors and how they influenced the characteristics of neighboring actors.

The first methodology is motivated by formulating a large network through the Exponential Random Graph Model and applying a Bayesian approach through the reference prior technique to control the sensitivity of the inference and to get the maximum information from the model. We consider a large Amazon product co-purchasing network (customers who bought this item also bought other products), and the purpose is to show how the blending of the Exponential Random Graph Model and Bayesian Computation efficiently handles the estimation procedure and calculates the probability of certain graph structures.

The second methodology we discuss is an approach to a network problem where the network adjacency structure remains unobserved, and instead we have a nodal variable that inherits a hidden network structure. The key assumption in this method is that the nodes are assumed to have a specific position in an Euclidean social space.

The main analysis is based on three big U.S. auto manufacturers and their suppliers, and recent research has explored the differences of the financial markets and an emphasis has been given to reveal the strategic interactions among companies and their industry rivals and suppliers, all of which have important implications for some fundamental questions in the financial economics. Economic shocks are transmitted through the customer supplier

network and the whole industry could be affected by these shocks as they can move through the links of the actors in an industry. We developed an algorithm that captures the latent linkages between firms based on sales and cost data that influence various financial decision-making issues and financial strategies.

Finally, we extend the problem of network estimation to Bayesian variable selection whereby an observed adjacency structure between different regions has been considered. The main idea is to select relevant variables region-wise. We investigate this problem using a Bayesian approach by introducing the Bayesian Group LASSO technique with a bi-level selection that not only selects the relevant variable groups but also selects the relevant variables within that group. We use spike and slab priors, along with the Conditional Autoregressive structure among the model coefficients, which validates the spatial interaction among the covariates. Median thresholding is used instead of posterior mean to have exact zeros for the variables that are not relevant. We finally implement the problem in the auto industry data and incorporate more variables to see whether the estimated adjacency structure helps us to indicate the relevant variables over different manufacturers and suppliers.

ACKNOWLEDGMENTS

As a PhD student of the Department of Statistics and Probability at Michigan State University, I feel very fortunate in having the opportunity to work with some amazing professors in the past few years. First, my advisor Professor Tapabrata Maiti, who has helped me a lot, and without his endless support and encouragement in my research, it would have been really tough to reach my research goals. I'm thankful to him that he had faith in me.

I'm also thankful to my research committee members, Professor Chae Young Lim, Professor Srinivas Talluri and Professor Jongeun Choi who gave me an opportunity to work with them and to prosper in my learning process, which also helped me a lot in looking towards my research career. I also had opportunity to work with Prof. Arnab Bhattacharjee from Hariot-Watt University and I'm thankful to him for his significant help in my research.

I'm also thankful to the Department of Statistics and Probability at Michigan State University for providing me with funding support so that I could continue my research. I'm thankful to the other professors in our department who contributed to my wonderful years here through their courses and helpful discussion sessions.

I'm thankful to Dell Inc., and specifically to Dr. Thomas Hill, for providing me with a Research Assistantship for spring, 2016, which made it easier for me focus on my research as I was relieved from TA duties.

I also have to mention the help received from some of my friends who are also graduate students in the department as I benefited from nice discussions and brain-storming. Personally, I feel proud to be a part of such a prestigious department, and I'm thankful to everyone in the department for supporting me in reaching my objectives.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Social Networks

Social nertwork modeling has become increasingly popular in past few years due it's ability to find causal links between nodes and for explaining those links in probabilistic terms. It is very important to model irregular social behavior that lies beyond the regular variability and brings stochasticity in the model. Moreover in a complex social environment, it is very important to not only have a probabilistic explanation of the edges but also to explain some specific structure to explore some interesting social interactions.

A statistical network is a representation of relational data in the form of a mathematical graph where each node represents an individual and a relation between a pair of nodes is represented by an edge between those two nodes. Network data typically consist of a set of N nodes and a relational tie y_{ij} measured on each ordered pair of nodes. This framework has many applications in social network literature. The simplest situation is when y_{ij} is a dichotomous variable that indicates the presense or absense of some relation of interest. The data are often represented by an $N \times N$ socio matrix or the adjacency matrix $\mathbf{Y}$. Various probabilistic models of network relations have been developed within past few years.

A statistical network is a graph consists of a set of N nodes (or Vertices)= $\{n_1, n_2$,-$..., n_N\}$. and a set of L edges (or connections) = $\{l_1, l_2, ..., l_L\}$ that denotes the links between nodes. An Adjacency or Sociomatrix $\mathbf{Y}$ of dimension $N \times N$ can be used to represent the network by,

$$y_{ij} = \begin{cases} 1 & \text{if edge exists from node } n_i \text{ to} \\ & \text{node } n_j, \\ 0 & \text{Otherwise.} \end{cases} \tag{1.1}$$

Holland and Leinhardt (1981) includes the parameters for the propensity of ties to be reciprocal, as well as parameters for the number of ties and individual tendencies to give or receive ties. Although this model assumes the $\binom{n}{2}$ dyads to be independent and known as p_1 model. Frank and Strauss (1986), Pattison and Wasserman (1999) and Wasserman and Pattison (1996) have generalized the idea of p_1 model to p^* model through dyad dependency assumption.

Wang and Wong (1987) developed a stochastic block model where nodes (firms) belong to some prespecified groups. Nowicki and Snijders (2001) present a model where group membership is unobserved and the dyads in a social network are conditionally independent given the latent class membership of each actor. In the spatial context, Castro et al. (2015) developed a model where latent group membership is inferred using spatial clustering with an unknown number of clusters. Likewise, in the classical spatial econometrics literature, Bhattacharjee and Holly (2013) develop GMM methods to infer on a latent network of members in a committee; for related classical inferences on latent spatial networks, see also Bhattacharjee and Jensen-Butler (2013), Bailey et al. (2015) and Bhattacharjee et al. (2015).

1.1.1 Exponential Random Graph Model

Frank and Strauss (1986) characterized the exponential random graph model (ERGM) that allows an estimation of various graphical structures through an assumption of dyad dependence. The typical form of exponential random graph model (ERGM) is given by,

$$\mathbb{P}_\theta(Y = y) = \frac{e^{\theta^t S(y)}}{c(\theta)} \tag{1.2}$$

where, $S(y)$ is a known vector of graph configuration, θ is the parameter corresponding to the configuration $S(y)$, $c(\theta)$ is the normalizing constant.

ERGM is very important in the sense that it goes beyond the idea of discovering the link probability between a pair of nodes by considering some graphical characteristic among a set of nodes. That is, $S(y)$ can represent different network configurations, if we observe $\{y_{34}, y_{43}\}$ and $\{y_{12}, y_{21}\}$, as means we can expect to have some reciprocating characteristic between the nodes, which means if node i is linked with node j, then we can expect j will also be linked with i. A typical example of such links can be a friendship network. But this configuration might not be true in all instances. For example, we can think an electricity power supply network that is unusual to be reciprocated. The corresponding parameter θ estimates the frequency of appearance of the specific configuration present in the network.

The main issue that the Maximum Likelihood Estimation of the ERGM model faces is to calculate $c(\theta)$. Suppose G denotes all possible graphs of $\mathbf{Y}$. Hence, $c(\theta) = \sum_{G} e^{\theta^t S(y)}$. Now G consists of $2^{\binom{n}{2}}$ possible undirected graphs and it is extremely difficult to evaluate the normalizing constant even for moderately small graphs.

To deal with the complex issues of computation intensity with ERGM for even moderate sized network, Besag (2000), Handcock (2000), Snijders (2002) have developed a likelihood-based inference based on MCMC algorithms. Although, Monte Carlo maximum likelihood estimation suffers from the problem of model degeneracy as we get a very poor estimate of the normlizing constant if the initiatial value of θ lies in the degenerate region. Approximate maximum likelihood approaches have been developed by Frank and Strauss (1986). A pseudolikehood approach is proposed by Strauss and Ikeda (1990) and Wassarman and Patterson (1996). But the statistical properties of pseudolikelihood estimators in this context have been criticized by Besag (2000) and Snijders (2002). Recent development on ERGM

has led to new specification that have been discussed by Hunter and Handcock (2006), called the curved ERGM.

1.1.2 Bayesian ERGM

A Bayesian extention to the Exponential Random Graph Model has been discussed in Caimo and Friel (2011) where they have considered $\pi(\theta \mid y) = \mathbb{P}_\theta(y)\pi(\theta)$, where a prior distribution $\pi(\theta)$ is placed on θ and interest is in the posterior $\pi(\theta \mid y)$. Such a Bayesian treatment easily solves the problem of evaluating the value of normalizing constant in the likelihood estimation case. A Bayesian treatment aslo solves the problem of model degeneracy for MCMC maximum likelood technique. Although, the posterior of this Bayesian problem becomes "doubly intractable" due to the intractibility of sampling directly from the posterior distribution but also due to the intractibility of the likelihood within the posterior. A simple implementation of Metropolis-Hastings algorithm proposing to move from θ to θ^* would require the calculation of the ratio,

$$\frac{e^{\theta^{*\prime} S(y)}\pi(\theta^*)}{e^{\theta^{\prime} S(y)}\pi(\theta)} \times \frac{c(\theta)}{c(\theta^*)}$$

which is unworkable due to the normalizing constant $c(\theta)$ and $c(\theta^*)$.

To handle the "doubly intractable" posterior, Murray et al. (2006) and Caimo and Friel (2011) proposed an exchange algorithm with samples from an augmented distribution.

$$\pi(\theta^*, y^*, \theta \mid y) \propto \mathbb{P}_\theta(y)\pi(\theta)h(\theta^* \mid \theta)\mathbb{P}_{\theta^*}(y^*) \tag{1.3}$$

where $\mathbb{P}_{\theta^*}(y^*)$ is the same distribution as the original distribution on which the data y is

defined. $h(\theta^* \mid \theta)$ is the proposal distribution. Clearly marginal distribution of θ is the posterior distribution of interest.

The steps for exchange algorithm are as follows:

1. Draw $\theta^* \sim h(* \mid \theta)$

2. Draw $y^* \sim \mathbb{P}_{\theta^*}(*)$

3. Propose the exchange move from θ to θ^* with probability

$$\alpha = min\left(1, \frac{e^{\theta' S(y^*)}\pi(\theta^*)h(\theta \mid \theta^*)e^{\theta^{*'} S(y)}}{e^{\theta' S(y)}\pi(\theta)h(\theta^* \mid \theta)e^{\theta^{*'} S(y^*)}}\right)$$

1.1.3 Latent Space Model

In a highly influential paper, Hoff et al. (2002) developed a latent variable model where node is assigned with a latent position z_i in the social space. The idea is that the probability of a relational tie between two individuals (or nodes) are higher if these individuals are similar in the unobserved characteristic space. In this context the social space refers to a space of unobserved latent characteristics that represent potential transitive tendencies in network relations. The resulting networks are probabilistically transitive since $i \to j$ and $j \to k$ suggests i and k are probably not far apart in the social space. Most recently, handcock and Raftery (2007) developed a model based clustering of social networks where they modeled the latent positions as a mixtures of multivariate normals.

The latent space model takes a conditional independence approach to modeling by assuming the presence or absence of a tie between two nodes that independent of all other ties,

given the unobserved positions in the latent space of the two nodes.

$$\mathbb{P}(\mathbf{Y} \mid Z, X, \theta) = \prod_{i,j} \mathbb{P}(y_{i,j} \mid z_i, z_j, x_{i,j}, \theta)$$

Here X and $\mathrm{x}_{i,j}$ are observed characteristic that are dyad specific and may be vector valued and θ and Z are respectively parameters and the unknown latent positions.

Consider a logistic regression model as below,

$$\begin{aligned} \eta_{i,j} \quad &= logodds(y_{i,j} = 1 \mid z_i, z_j, x_{i,j}, \alpha, \beta) \\ &= \alpha + \beta^{'} x_{i,j} - f(z_i, z_j) \end{aligned}$$

The function f is chosen to be simple which represents the forms of network dependence. Here we assume,

$$z_1,...,z_n \sim Normal(0, {\sigma_z}^2)$$

The latent space model is inherently reciprocal and transitive. If $i \to j$ ihen it means the distance between node i and node j is small, which makes $j \to i$ more probable. Again $i \to j$ and $j \to k$ impliy the distance between node j and node k is not two large, which makes the event $j \to k$ more probable.

$f(z_i,z_j)$ can be replaced by any arbitrary set of distances $d_{i,j}$ satisfying the triangle inequality. In general, we prefer to model the $d_{i,j}$'s as distances in some low-dimentional Euclidean space for reasons of parsimony and ease of model interpretablity.

We say a set of distances $d_{i,j}$ represents the network $\mathbf{Y}$ if

$$\{d_{i,j} > c \; \forall i, j : y_{i,j} = 0\}$$

and

$$\{d_{i,j} < c \; \forall i, j : y_{i,j} = 1\} \tag{1.4}$$

We say that a network is d_k representable if $\exists$ points $z_i \in \Re$ such that the distances $d_{i,j}$ satisfy (1.4). Hence, d_k representability is equivalent to being able to find a set of points for the actors such that $i \sim j$ iff i and j lie within k dimnetional unit balls centered around each other.

Given a network data $\mathbf{Y} = y_{i,j}$ and possible covariates of the model $\mathbf{X} = x_{i,j}$, the goal is to estimate the unknown parameters of the model, denoted as θ. The parameter θ includes the regressor coefficients α, β and the variance of the random positions of the nodes in the latent space.

We take a Bayesian approach for estimation using a prior probability distribution p(θ). Conditional distribution of the parameters given the information in the data is

$$p(\theta \mid \mathbf{Y}) = p(\mathbf{Y} \mid \theta) \times p(\theta)/p(\mathbf{Y})$$

The MCMC based inference constructs a dependent sequence of θ values as follows:

- Sample a parameter θ^* from a proposal distribution $h(\theta \mid \theta_k)$;

- Compute the acceptance probability

$$r = max\left(1, \frac{p(\mathbf{Y} \mid \theta^*)p(\theta^*)h(\theta_k \mid \theta^*)}{p(\mathbf{Y} \mid \theta)p(\theta)h(\theta^* \mid \theta_k)}\right);$$

- Set $\theta_{k+1} = \theta^*$ with probability r and $\theta_{k+1} = \theta_k$ with probability $1 - r$.

This algorithm produces a sequence of θ values having a distribution which is approximately equal to the target distribution $p(\theta \mid \mathbf{Y})$. A point estimate of θ is often taken to be the posterior mean, which is approximted by the average of the sampled θ values.

1.2 Bayesian Variable Selection

Recent developments in statistical literature put a huge emphasis on variable selection for the explosive sample space due to the increase in dimention, nice frameworks have been developed to handle the variable selection procedure in the Bayesian framework.

Liang, Song and Yu(2013) introduced the idea of Bayesian Subset Regression (BSR) starting with a subset model and taking Gaussian priors on the model coefficients $\beta_i's$ and they showed that if the true model becomes sparse, i.e., $\lim_{n\to\infty} \sum_{i=1}^{P_n} \mid \beta_i \mid < \infty$ where P_n is the model dimention, BSR reduces to EBIC. Under some mild conditions, they have also shown the posterior consistency of the model. They have also proposed a variable screening procedure based on the marginal inclusion probability of the predictors and they have shown that it has the same property of Sure Independence Screening (SIS) where we rank predictors according to their marginal utility and then selects a subset of the predictors of the marginal utility exceeding some predifined threshold.

Bondell and Reich (2012) proposed a variable selection criterion based on the posterior

credible region. Here they first fit the full model using all predictors and then used the highest posterior density region of β to have a sparse estimate. This sparse vector then determines the selected model.

Penalized regression is a method that not only selects the relevant variables but also estimates the regression coefficients simultaneously. LASSO regression (Tibshirani, 1996) provides a decent solution by putting upper bound on the $\mathbb{L}_1$-norm. Suitably selecting the penalty parameter can provide an exact zero estimate for the corresponding irrelavant variables. Tibshirani (1996) pointed out that the LASSO estimator can be interpreted as the maximum apostiriori (MAP) estimator when the regression parameters have independent and identical Laplace priors. Least Angle Regression (LARS) provides more attractive solution since it follows the full LASSO solution path with the cost of only one least square estimation (Efron et al., 2004).

Park and Casella (2008) introduced the LASSO in a similar Bayesian context where they introduced a laplace prior on the penalty parameter that boils down the whole problem in to a Bayesian context. They have also used the fact that the laplace distribution can be expressed as a gamma scaled mixture of Normal that facilitates the posterior computaion. A major advantage of using Bayesian LASSO over Frequetist LASSO is it provides reliable standard error over the non-Bayesian method (Knight and Fu, 2000; Chatterjee and Lahiri, 2001; Tibshirani, 1996). More specifically, the LASSO estimator is equivalent to the posterior mode with independent laplace priorfor the coefficients. Using the fact that laplace distribution can be represented as a scale mixture of normals, Park and Casella (2008) developed a fully Bayesian hierarchical model and efficient Gibbs sampler for the posterior compuations.

For large n small P regression, Liang, Truong, Wong (2001) established an explicit relationship between the Bayesian approach and the penalized likelihood approach for linear

regression. They showed empirically that Bayesian Subset Regression (BSR) that is choosing priors such that the resulting negative log-posterior probability of the subset model can be approximately reduced to frequentists subset model selection statistic upto a multiplicative constant.

Selecting relevant variables in a high-dimentional setup is a very common feature in various Bayesian and econometric applications. In an additive model, a set of continuous predictor may be represented as group of predictors. Huang et al. (2012) provides a nice insight for the application of group variable selection. Yuan and Lin (2006) proposed a group LASSO method that provides a group variable selection.

Bayesian group LASSO technique has been developed by Kyung et al. (2010) and Ramen et al. (2009) that handles the problem of selecting the variables at the group level only. If we consider a linear regression model of the following form,

$$\underset{\sim}{y_i} = \sum_{g=1}^{p} \mathbf{X}_g \underset{\sim}{\beta_g} + \underset{\sim}{\epsilon}$$

where $\underset{\sim}{\epsilon} \sim N(0, \sigma^2 \mathbf{I}_N)$, $\underset{\sim}{\beta_g}$ is a coefficient vector and $\mathbf{X}_g$ is the covariate matrix for the corresponding g^{th} group and consider the minimization problem as (Simon et al; 2012),

$$\min_{\underset{\sim}{\beta}} \left(||\underset{\sim}{y_i} - \sum_{g=1}^{p} \mathbf{X}_g \underset{\sim}{\beta_g}||_2^2 + \lambda_1 ||\underset{\sim}{\beta}||_1 + \lambda_2 \sum_{g=1}^{p} ||\underset{\sim}{\beta_g}||_2 \right)$$

then the second penalty term induces a variable selection in the group level and the first penalty term induces a variable selection within group level. It can easily shown that the laplace prior corresponding to above minimization problem can be expressed as a scale mixture of normals and hence a full Bayesian implementation would be easy.

A mixture prior with a point mass at zero is a very effective tool to have a controled amount of shrinkage in the variable selection technique. Spike and Slab priors (Ishwaran and Rao, 2005) and zero inflated mixture priors (Mitchell and Beauchamp, 1988) are very effective technique for Bayesian Variable Selection.

We set up the following hierarchical model as:

$$
\begin{aligned}
&y_i|\underset{\sim}{x}_i, \underset{\sim}{\beta}, \sigma^2 \sim N(\underset{\sim}{x}_i^{'}\underset{\sim}{\beta}, \sigma^2) \\
&\beta_k|\phi_k, \tau_k^2 \sim N(0, \phi_k\tau_k^2) \\
&\phi_k \mid \rho, v_0 \sim (1-\rho)\delta_{v_0}(.) + \rho\delta_1(.) \\
&\tau_k^{-2} \mid b_1, b_2 \sim Gamma((a_1, a_2) \\
&\rho \sim U[0,1] \\
&\sigma^2 \sim Gamma(b_1, b_2)
\end{aligned}
$$

Hence, the above hiearchical model implies:

$$\beta_k|\tau_k^2, \rho \sim (1-\rho)N(0, v_0\tau_k^2) + \rho N(0, \tau_k^2) \tag{1.5}$$

The selection of β_k can be controled by the two normals as the shrinkage effect v_0 pulls the first normal around 0 (spike) and the significance of β_k is controled through the second normal.

Zero inflated normal mixture priors in the hierarchical formulation for variable selection have been used in the linear regression model (George and McCulloch, 1997). Point mass mixture priors are also studied by Johnstone and Silverman (2004) and Xu and Ghosh (2015) for estimation of possibly sparse sequence of Gaussian observations with an emphasis

on using the posterior median instead of the Posterior mean, which has been proven to be more effective. A very strong result has been shown in Johnstone and Silverman (2004) and Yuan and Lin (2005) where they have combined the power of point mass mixture priors and double exponential distribution and the resulting empirical bayes estimator is closely related to LASSO estimator. Lykou and Ntzoufras (2013) proposed a similar approach by specifying a shrinkage parameter λ through Bayes factor and Zhang et al. (2014) generalizes this prior for group LASSO technique and proposed a hierarchical structured variable selection technique.

Chapter 2

Reference Prior Development in Exponential Random Graph Model

2.1 Main Idea

The key difference in Bayesian literature from the frequentist is that Bayesian uses prior information on the model parameters that, in some sense, makes the model more robust and protects it from being carried out by sampling errors or through a lack of samples. Some complex computational issues that arises in the frequesntist approach can also be overcomed through the implementation Bayesian techniques. It is therefore important to have a strong knowledge about the prior distribution of the model's parameters. In many situations, we may not have a strong hold on the priors and a stringent informative prior may drive the problem to produce some unrealistic estimates of the parameters. So it is sometimes necessary to be non-informative about the prior knowledge.

The key idea of a Bayesian problem is to choose a prior in such a way that it does not become very stringent and can produce the estimates in a data-driven fashion. The problem is then to figure out a non-informative prior that can extract the maximized information to build the posterior. (Bernardo 1979), Berger and Bernardo (1989, 1991a, 1991b) have developed the idea of reference prior that can maximize the information between the prior and the posterior for a given problem through Kullbeck-Libeler $(K-L)$ divergence.

Suppose we have data Y parameterized by Θ with sufficient statistic $T = T(Y)$.

Definition. *A reference prior is a prior that maximizes $K-L$ divergence from the posterior $\pi(\theta \mid t)$ averaged over the distribution of T i.e., we want to maximize*

$$I(\Theta, T) = \iint p(t)\pi(\theta \mid t) log \frac{\pi(\theta \mid t)}{\pi(\theta)} d\theta dt \tag{2.1}$$

over all $\pi(\theta)$.

The reference prior satisfying equation (2.1) maximizes the posterior information obtained from the class of all the default priors. In this case, we are looking for such a prior so that our data can have its maximum impact on the posterior estimates. Moreover, posterior distribution becomes highly sensitive with the choice of the informative prior. If we introduce more covariates in the model, the sensitivity increases. We would then like to use the virtue of the default priors that do not put any prior information on the parameter estimates and relies on the fact that it optimizes the posterior estimates in a data-driven information theoretic technique.

ERGM model faces a key theoritical issue of working with a single sample. In a typical ERGM problem, we work with one observed instance of the adjacency structure. Hence, technically, ERGM operates in a limited data atmosphere where the data driven information is very limited. Hence, it is very important to have a procedure that can optimize the extraction of the available information.

2.2 Reference Prior for One Parameter Erdos-Reyni Model

The main purpose of this chapter is to implement the reference prior technique for Bayesian ERGM. Note that the main purpose of introducing the reference prior for an ERGM model is that we can resolve the problem of the sensitivity of the posterior estimates caused through the informative priors as well as provide the posterior estimations that are optimized in a information theoretic sense.

The simplest example of one parameter ERGM model is the Erdos-Reyni network model. In this modeling scenario, we assume the edges are independent with fixed edge probability

θ (Erdos and Reyni, 1959). The model can be written as,

$$f(y \mid \theta) = \prod_{i \neq j} \theta^{y_{ij}} (1-\theta)^{1-y_{ij}} \quad , 0 < \theta < 1 \tag{2.2}$$

This model concentrates on the existence of the single possible edge configuration y_{ij}, which is parameterized by θ. Here θ is called the edge density parameter. Reparameterization of (2.2) gives,

$$f(y \mid \eta) \sim exp\{\eta \sum_{i \neq j} y_{ij}\} \tag{2.3}$$

So this is an Exponential Random Graph Model. Reference prior is defined in terms of mutual information, and since the mutual information is itself invariant, the reference priors become invariant under reparameterization. So we can calculate reference prior either for θ or for η.

Suppose $\underset{\sim}{x}$ be the vector of observations from the model and $\underset{\sim}{x}^{(k)} = (\underset{\sim}{x}_1, \underset{\sim}{x}_2, ..., \underset{\sim}{x}_k)$ be a vector of independent replicates of the vector observations from the model.

Let $t_k = t_k(\underset{\sim}{x}_1, \underset{\sim}{x}_2, ..., \underset{\sim}{x}_k) \in \tau_k$ be any sufficient statistics of the replicated observations. Let us define,

$$\pi^*(\theta \mid t_k) = \frac{p(t_k \mid \theta)\pi^*(\theta)}{\int_\Theta p(t_k \mid \theta)\pi^*(\theta) d\theta}$$

and,

$$f_k(\theta) = exp\Big(\int_{\tau_k} p(t_k \mid \theta) log[\pi^*(\theta \mid t_k)] dt_k \Big)$$

Theorem (Berger, Bernardo, and Sun; 2009). *Assume a standard model* $\mathcal{M} \equiv \{p(\underset{\sim}{x} \mid \theta), \underset{\sim}{x} \in \mathbb{X}, \theta \in \Theta \subset \mathbb{R}\}$ *and* $\mathcal{P}$ *is the standard class of candidate priors. Let* $\pi^*(\theta)$ *be a continuous strictly positive function such that the corresponding formal posterior* $\pi^*(\theta \mid t_k)$ *is proper and asymptotically consistent. Define for any interior point* θ_0 *of* Θ*,*

$$f(\theta) = \lim_{k \to \infty} \frac{f_k(\theta)}{f_k(\theta_0)}$$

1. *If each* $f_k(\theta)$ *is continuous and, for any fixed* θ *and sufficiently large* k*,*

   $\{f_k(\theta)/f_k(\theta_0)\}$ *is either monotonic in* k *or bounded above by some* $h(\theta)$ *which is integratable on any compact set, and,*

2. $f(\theta)$ *is a permissible prior function,*

   then $f(\theta)$ *is a reference prior for this model* $\mathcal{M}$ *and prior class* $\mathcal{P}$*.*

Since the analytical derivation of a reference prior may be technically demanding due to a complex model, we can use the following algorithm for the 1-parameter family of distributions.

Algorithm (Berger, Bernardo, Sun, 2009).

1. *Initial values:*

   Choose a moderate value for k*; Choose an arbitrary positive function* $\pi^*(\theta)$*;*

   Choose the number m *of samples to be simulated.*

2. *For any given* θ*, repeat, for* $j = 1, 2, ..., m$*:*

   Simulate a random sample $\{x_{1j}, x_{2j}, ..., x_{kj}\}$ *of size* k *from* $p(x \mid \theta)$*;*

   Compute numerically the integral $c_j = \int_\Theta \prod_{i=1}^k p(x_{ij} \mid \theta)\pi^*(\theta)d\theta$*;*

   evaluate $r_j(\theta) = log(\prod_{i=1}^k p(x_{ij} \mid \theta)\pi^*(\theta)/c_j)$*.*

3. *Compute* $\pi(\theta) = exp\{m^{-1}\sum_{j=1}^{m} r_j(\theta)\}$ *and store the pair* $\{\theta, \pi(\theta)\}$.

4. *Repeat routines* (2) *and* (3) *for all* θ *values for which the pair* $\{\theta, \pi(\theta)\}$ *is required.*

For the 1-parameter Erdos-Reyni ERGM model, we can compute the reference by using the notion of maximizing the $K - L$ divergence principle. Figure (2.1) shows the reference prior for θ in the Erdos-Reyni model. Now the Jeffrey's prior for θ in the one parameter Erdos-Reyni model is given by $I(\theta) = \sqrt{\frac{n}{\theta(1-\theta)}}$. It is then interesting to see the resemblance between the Jeffrey's prior and the reference prior θ in the 1-parameter model. It is also a nice example to see that the reference prior in one parameter model is nothing but Jeffrey's prior under certain regularity conditions.

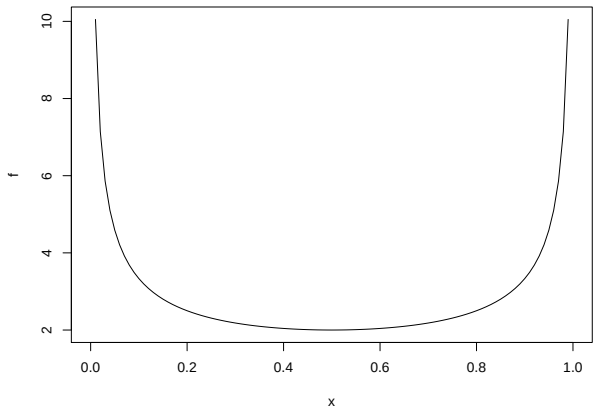


Figure 2.1: Reference Prior for θ in 1-Parameter (Erdos-Reyni) network Model

2.2.1 'Sampson's Monk Data' Implementation

We implement the reference prior approach for one parameter, Erdos-Reyni Model to Sampson's Monk dataset, which provides an adjacency structure representing the interaction between 18 monks in a monastery. Our target is to get the posterior estimate of the model

parameter η. Since the purpose of this paper is to catch the sensitivity due to informative priors, we compare the posterior through the reference prior with a non-informative uniform prior and with normal priors with a 0 mean and four different values for the standard deviation. The output shows the sensitivity of the posterior due to the informative normal priors.

It is important to address prior sensitivity in such a simple one-parameter modeling situation. If we incorporate more parameters in the model by introducing more complex structures in the network, we expect the sensitivity to increase further. The use of non-informative priors, such as a uniform prior, will resolve the issue of prior sensitivity, but it is not guaranteed to provide the maximized information with respect to the prior.

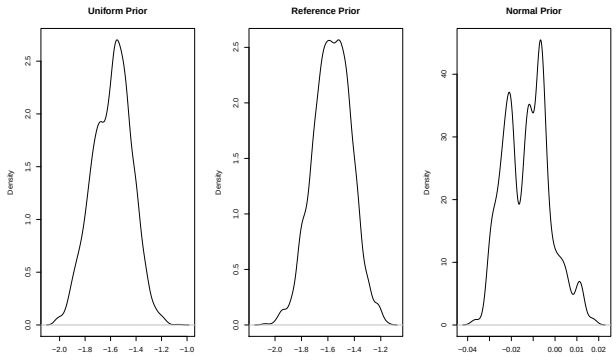


Figure 2.2: Posterior density of η for monk dataset with respect to Uniform prior, Reference prior and normal prior with $\sigma = 0.01$

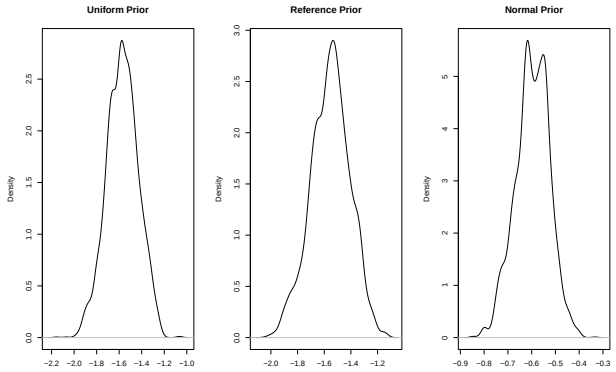


Figure 2.3: Posterior density of η for monk dataset with respect to Uniform prior, Reference prior and normal prior with $\sigma = 0.1$

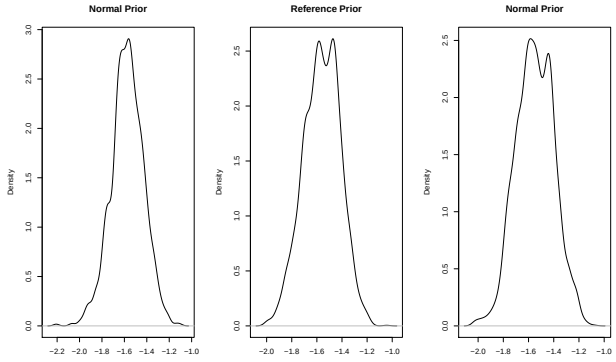


Figure 2.4: Posterior density of η for monk dataset with respect to Uniform prior, Reference prior and normal prior with $\sigma = 1$

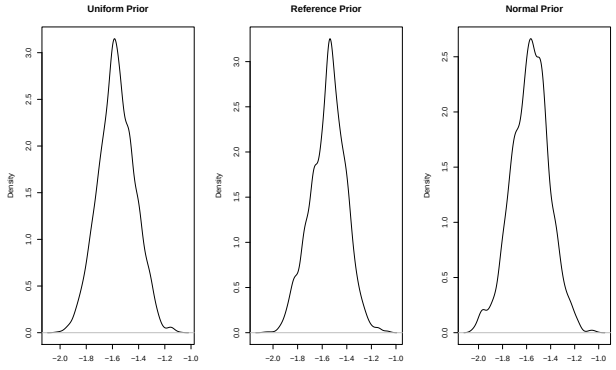


Figure 2.5: Posterior density of η for monk dataset with respect to Uniform prior, Reference prior and normal prior with $\sigma = 100$

2.3 Reference Prior for Two Parameter Dyadic independent network Model

To introduce a two-parameter dyadic independent network model, we first decompose $\mathbf{Y}$ into $\binom{n}{2}$ dyads of pairs, $D_{ij} = (y_{ij}, y_{ji})$ for $i < j$. To describe the joint distribution, we extend our independence assumption from the edge independent Erdos-Reyni model up to dyads. Here, we consider reciprocated edges along with single edges. We can then write our model as,

$$f(y \mid \theta_1, \theta_2) \sim exp(\theta_1 \sum_{i \neq j} y_{ij} + \theta_2 \sum_{i \leq j} y_{ij} y_{ji}) \tag{2.4}$$

Here, $\sum_{i \neq j} y_{ij}$ denotes the number of edges in the network and $\sum_{i \leq j} y_{ij} y_{ji}$ denotes the number of mutual ties. In this modeling scenario, we define θ_1 as the edge density parameter and θ_2 as the reciprocity parameter.

We can extend the one-parameter reference prior idea to a two-parameter case through the sequential scheme (Berger and Bernardo, 1992). That means we first arrange our parameters in terms of their inferential importance; in particular, the first parameter should be the parameter of interest. For the dyadic independent model, obviously our parameter of interest would be θ_2. Since the normalizing constant is unknown and hard to calculate for the dyadic independent model, we try to solve the problem in a Bayesian setup. Now the computation of a reference prior requires a closed form expression of the likelihood function that is intractable in a ERGM setup even for a moderately large network. Therefore, to facilitate the reference prior computation, we approximate the likelihood function through a pseudo-likehood that

is given by,

$$\begin{aligned} PL(\theta_1, \theta_2 \mid Y) &= \prod_{i \neq j} \mathbb{P}(y_{ij} = 1 \mid {y_{ij}}^C) \\ &= \prod_{i \neq j} \frac{e^{\theta_1 + \theta_2 y_{ij}}}{1 + e^{\theta_1 + \theta_2 y_{ij}}} \end{aligned} \tag{2.5}$$

where ${y_{ij}}^C$ denotes the network $\mathbf{Y}$, except nodes i and j. A theoretical validation of this approximation can be followed from Strauss & Ikeda (1990) and Besag (1974, 1975) where they have shown that likelihood maximization of (2.4) is equivalent to the maximization of (2.5).

2.3.1 Methodology for Derivation

We first consider the fact that a two-parameter dyadic ERGM is asymptotically normally distributed since it is nothing but a two-parameter exponential family.

Suppose $\Theta = \{\theta_1, \theta_2 : -\infty \leq \theta_1 \leq \infty; -\infty \leq \theta_2 \leq \infty\}$ is the parameter space of the two-parameter dyadic independent model.

Let us consider a nested sequence of $\{\Theta^l\}$ of compact subsets of Θ such that $\bigcup_{l=1}^{\infty} \Theta^l = \Theta$. The key technique relies on the fact that we arrange the model parameters by importance with respect to the model. We calculate the density $\pi_2^l(\theta_1 \mid \theta_2)$, as

$$\begin{aligned} &\pi_2^l(\theta_1 \mid \theta_2) \\ &\propto exp\Big\{ \sum_{\mathbf{Y}} p(\mathbf{Y} \mid \theta_1, \theta_2) logp(\theta_1 \mid \mathbf{Y}, \theta_2) \Big\} \end{aligned} \tag{2.6}$$

and,

$$\pi_1^l(\theta_1, \theta_2)$$
$$\propto \pi(\theta_2 \mid \theta_1) exp\Big\{ \sum_{\mathbf{Y}} p(\mathbf{Y} \mid \theta_2) logp(\theta_2 \mid \mathbf{Y}) \Big\} \tag{2.7}$$

The key advantage here is the pseudo-likelihood approximation and since we have the assumption of conditional independence over the dyads, the single observed network of size n induces a dyadic model with $\binom{n}{2}$ replications. In a sequential set up, we will be considering the distribution of θ_1 for fixed θ_2 and also distribution of θ_2. For a fixed θ_2, and each term of (2.5) is an exponential function of θ_1 and hence it is twice continuously differentiable w.r.t θ_1. Hence we have,

$$ln \prod_{i,j} PL(\theta_1 + \delta\theta_1 \mid y_{ij}, \theta_2)$$
$$\approx ln \prod_{i,j} PL(\theta_1 \mid y_{ij}, \theta_2) + \frac{1}{\sqrt{n}} \sum_{i,j} \frac{\partial lnPL(\theta_1 + \delta\theta_1 \mid y_{ij}, \theta_2)}{\partial \theta_1}$$
$$+ \frac{1}{2n} \sum_{i,j}^{n} \frac{\partial^2 lnPL(\theta_1 + \delta\theta_1 \mid y_{ij}, \theta_2)}{\partial \theta_1 \partial \theta_*}$$

where, $\theta_* = \frac{1}{\sqrt{n}}$ Now the second term is asymptotically normal by Central Limit Theorem and the third term converges to fisher information matrix $I(\underline{\theta})$ in probability. Hence, by LeCam (1960) we have the pseudolikelihood is locally asymptotically normal and hence regular which means pseudolikehood is asymptotically normal.

Similarly, each term of (2.5) is twice differentiable as a function of θ_2.

Suppose $I(\theta)$ be the Fisher information matrix where,

$$I_{ij}(\underset{\sim}{\theta}) = -E_{\theta}\Big(\frac{\partial^2}{\partial\theta_i\partial\theta_j} log f(\mathbf{Y} \mid \theta_1, \theta_2)\Big)$$

Now,

$$\frac{\partial^2 PL}{\partial\theta_1^2} = -\sum_{i\neq j}\Big\{\frac{y_{ij}^2 e^{\theta_1+\theta_2 y_{ij}}}{(1+e^{\theta_1+\theta_2 y_{ij}})^2}\Big\}$$

and

$$E\Big(-\frac{\partial^2 PL}{\partial\theta_1^2}\Big) = n(n-1)\frac{e^{\theta_1+\theta_2}}{\big(1+e^{\theta_1+\theta_2}\big)}\Big\{\frac{e^{\theta_1}}{1+e^{\theta_1}} + \frac{e^{\theta_1+\theta_2}}{1+e^{\theta_1+\theta_2}}\Big\}$$

Similarly,

$$E\Big(-\frac{\partial^2 PL}{\partial\theta_2^2}\Big) = n(n-1)\Big\{\frac{e^{\theta_1}}{(1+e^{\theta_1})^2} + \frac{e^{\theta_1+\theta_2}}{(1+e^{\theta_1+\theta_2})^2}\Big\}\Big\{\frac{e^{\theta_1}}{1+e^{\theta_1}} + \frac{e^{\theta_1+\theta_2}}{1+e^{\theta_1+\theta_2}}\Big\}^{n(n-1)-1}$$

and,

$$E\Big(-\frac{\partial^2 PL}{\partial\theta_2\partial\theta_1}\Big) = n(n-1)\frac{e^{\theta_1+\theta_2}}{(1+e^{\theta_1+\theta_2})^2}\Big\{\frac{e^{\theta_1}}{1+e^{\theta_1}} + \frac{e^{\theta_1+\theta_2}}{1+e^{\theta_1+\theta_2}}\Big\}^{n(n-1)}$$

Therefore we get

$$\mid I(\underset{\sim}{\theta}) \mid = n^2(n-1)^2\Big\{\frac{e^{\theta_1}}{1+e^{\theta_1}} + \frac{e^{\theta_1+\theta_2}}{1+e^{\theta_1+\theta_2}}\Big\}^{2n(n-1)-2}\frac{e^{\theta_1+\theta_2}}{(1+e^{\theta_1+\theta_2})^2}\frac{e^{\theta_1}}{(1+e^{\theta_1})^2}$$

Now we set, $S(\theta) = (I(\theta))^{-1}$.

In a two-parameter setup, we can set $S(\theta) = \{((a_{ij})) : 1 \leq i \leq 2, 1 \leq j \leq 2\}$.

Now suppose S_j be the upper-left $j \times j$ corners of S and set $H_j = S_j^{-1}$. Now if $h_j(\theta)$ be the lower-right $j \times j$ corner of H_j, then we have

$$h_1(\theta) = \frac{1}{a_{11}} = n(n-1)\frac{\left\{\frac{e^{\theta_1}}{1+e^{\theta_1}} + \frac{e^{\theta_1+\theta 2}}{1+e^{\theta_1+\theta_2}}\right\}^{n(n-1)-1}}{\frac{\left(1+e^{\theta_1+\theta_2}\right)^2}{e^{\theta_1+\theta_2}} + \frac{\left(1+e^{\theta_1}\right)^2}{e^{\theta_1}}}$$

and

$$h_2(\theta) = a_{11} = n(n-1)\Big\{\frac{e^{\theta_1}}{\left(1+e^{\theta_1}\right)^2} + \frac{e^{\theta_1+\theta 2}}{\left(1+e^{\theta_1+\theta_2}\right)^2}\Big\}\Big\{\frac{e^{\theta_1}}{1+e^{\theta_1}} + \frac{e^{\theta_1+\theta 2}}{1+e^{\theta_1+\theta_2}}\Big\}^{n(n-1)-1}$$

The above formulation gives the explicit result for the probabilities given by equations (2.5) and (2.7) by,

$$\pi_2^l(\theta_1 \mid \theta_2) = \frac{|h_2(\theta)|^{1/2} I_{\Theta^l(\theta_2)}(\theta_1)}{\int_{\Theta^l(\theta_2)} |h_2(\theta)|^{1/2} d\theta_1} \tag{2.8}$$

and,

$$\begin{aligned}&\pi_1^l(\theta_1, \theta_2)\\ &= \frac{\pi_2^l(\theta_1 \mid \theta_2) exp\{\frac{1}{2}E^l[(log|h_1(\theta)|)|\theta_2]\} I_{\Theta^l(\theta_1,\theta_2)}(\theta_2)}{\int_{\Theta^l(\theta_1,\theta_2)} exp\{\frac{1}{2}E_1^l[(log|h_1(\theta)|)|\theta_2] d\theta_2\}}\end{aligned}$$

where,

$$E_1[g(\theta)|\theta_2] = \int_{\{\theta_1:(\theta_1,\theta_2)\in\Theta^l\}} g(\theta)\pi_2^l(\theta_1 \mid \theta_2) d\theta_1$$

Typically, we determine the reference prior for the two-parameter dyadic independent model as,

$$\pi(\theta) = \lim_{l} \frac{\pi_1^l(\theta)}{\pi_1^l(\theta^*)} \tag{2.9}$$

where, θ^* is any fixed point on Θ for which, the following condition satisfies,

$$E_l^Y D(\pi_1^l(\theta \mid \mathbf{Y}), \pi(\theta \mid \mathbf{Y})) \longrightarrow 0 \quad as \quad l \longrightarrow \infty$$

where $D(g, h)$ defines the $K - L$ divergence between densities g and h.

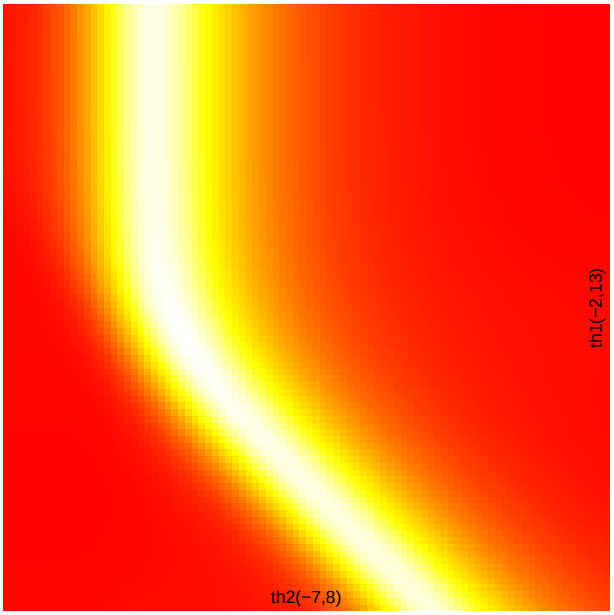


Figure 2.6: Reference Prior Heatmap for a Two-Parameter Dyadic Independent Model

2.4 Simulation Study for Dyadic Independent Model

In this section, we will be using a simulated network with node size $N = 50$. We restrict our simulation to a sparse network as most of the real-world examples generate sparse networks and the idea is to check how the reference prior can utilize its virtue of maximizing the information from the data to get the posterior estimates.

The key feature of Exponential Random Graph Model is that the specific structure is inherent in the model through specific parameters. In our application, we will try to see a graph where we believe that it has a dyadic feature, which means the network is directed and we are not going to see the network features beyond two nodes. In the current progress in network literature and computer science, the key issue is to estimate a large sparse network where most of the dyads are empty. This creates difficulty in the mixing of the MCMC chain since the chain involves most of its time to linking the empty dyads in case of a sparse network. The usual remedy (see Hunter et al., 2008) is to use a default "tie no tie" (TNT) sampler where we divide the whole set of dyads into two sets, one with all the links and another with no links, and then we pre-assign probabilities for these two sets and draw dyads from these two sets instead of drawing randomly from the whole network. The advantage of this technique is that most of the real-world networks are sparse and the sampler gives more chances to the set with edges to play most of the part in the MCMC chain. Moreover, we can tune the probability of the two sets to gain a better mixing since we try to control the degree of sparsity of the estimated network. In this technique, our computation does not depend on the size N of the whole network, but instead our computational complexity becomes $O(m)$ where m is the sample size to be selected from the two sets at each iteration step.

2.4.1 Scenario 1

In the first scenario, We set $\theta_1 = -3$ and $\theta_2 = 3$. We ran the iteration 40,000 times with a burn-in period of 20,000. We took a proposal distribution of $N(0, (0.25)^2)$, which makes the acceptance rate of the MCMC sampler stay around 20%. The posterior output of the MCMC chain, along with the auto-correlation plot and the MCMC iterations, is shown from Figure 2.7 to Figure 2.12b. We compute the posterior using two different default priors, first with $U(-\infty, \infty)$ and the second with the reference prior for the two-parameter dyadic independent model. For decreasing the autocorrelation between the MCMC samples, we apply a thinning process with $n = 5$.

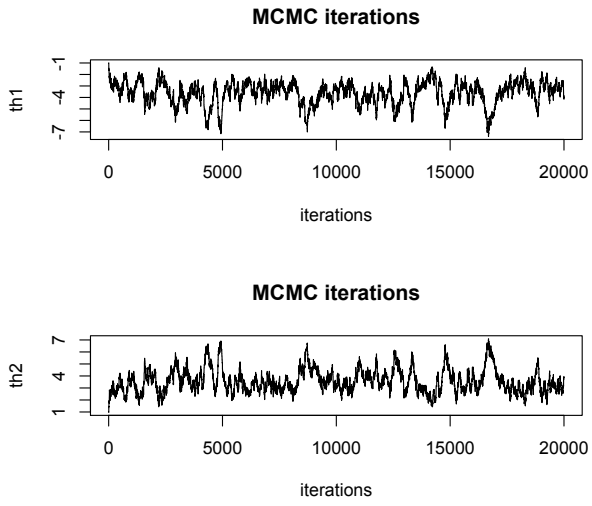
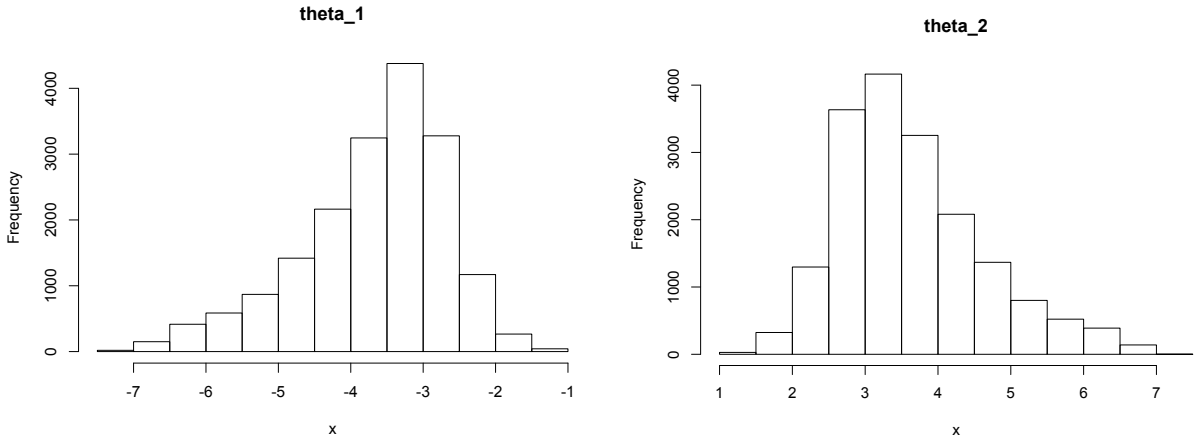
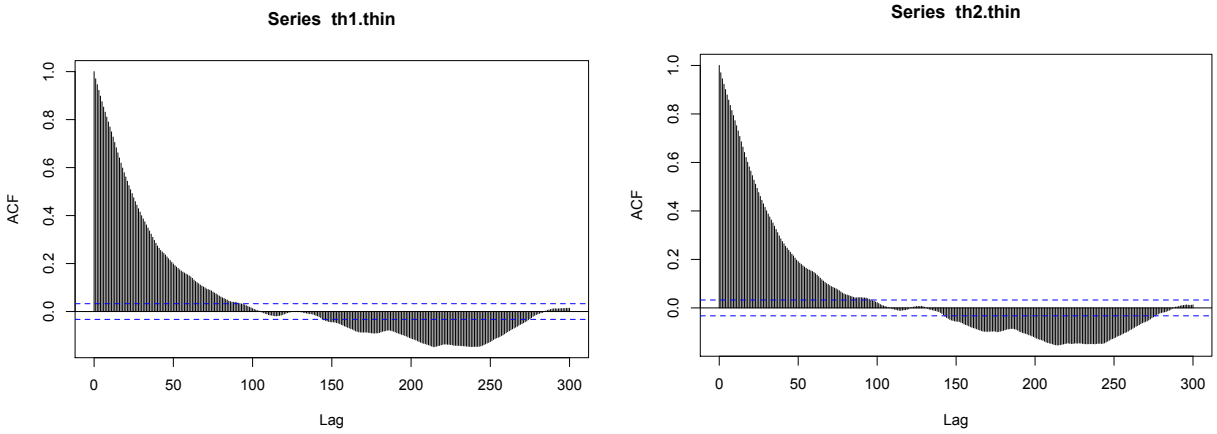


Figure 2.7: Scenario 1: MCMC iterations for θ_1 and θ_2 with uniform prior



(a) θ_1 (b) θ_2

Figure 2.8: Scenario 1: Histogram for posterior distribution of θ_1 & θ_2 with unoiform prior



(a) θ_1 (b) θ_2

Figure 2.9: Scenario 1: Auto-correlation plot for the MCMC iteration of θ_1 & θ_2 with uniform prior

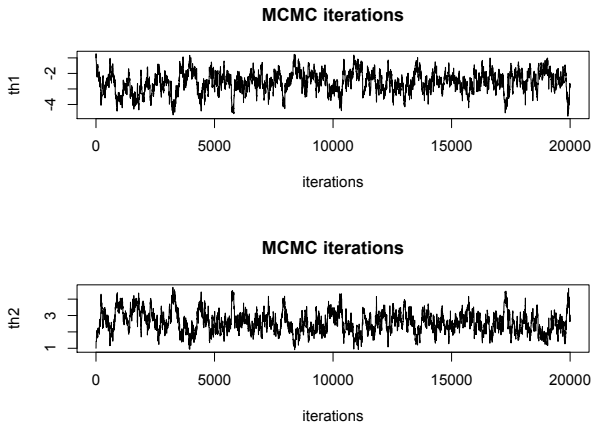
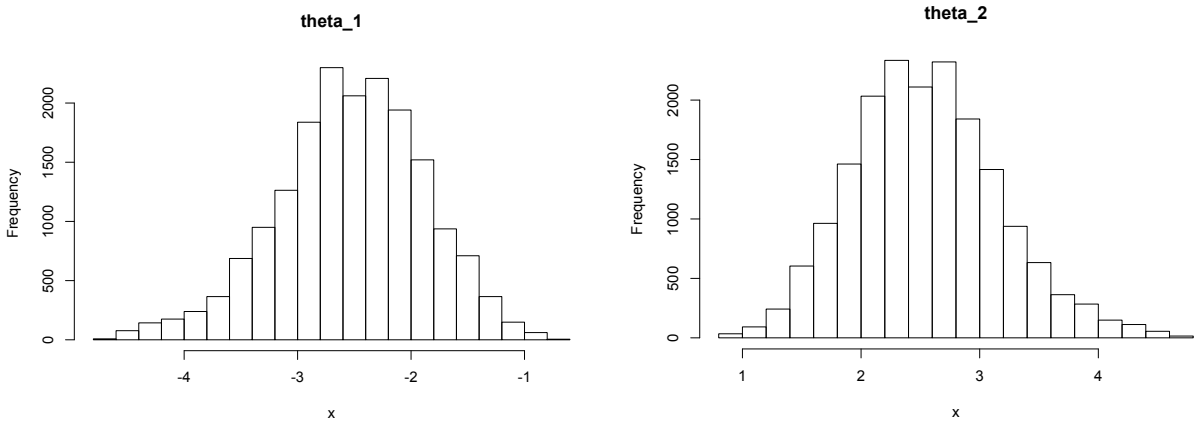
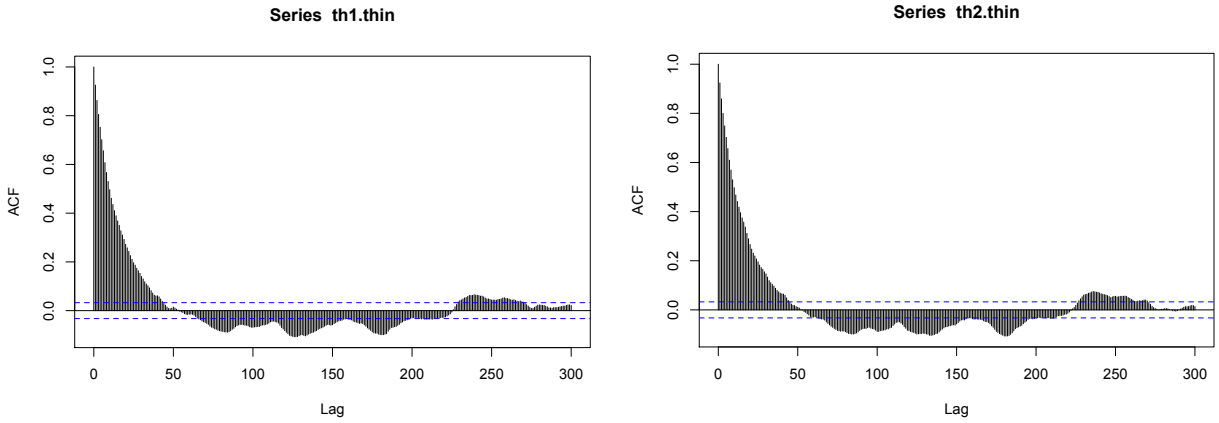


Figure 2.10: Scenario 1: MCMC iterations for θ_1 and θ_2 with Reference prior



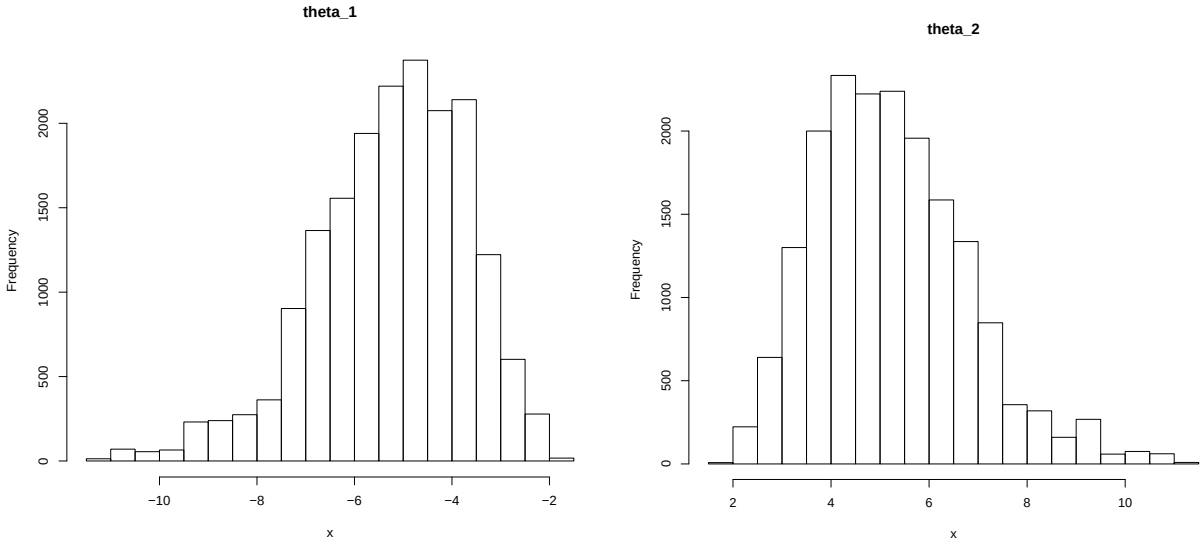
(a) θ_1 (b) θ_2

Figure 2.11: Scenario 1: Histogram for posterior distribution of θ_1 & θ_2 with reference prior



(a) θ_1 (b) θ_2

Figure 2.12: Scenario 1: Auto-correlation plot for the MCMC iteration of θ_1 & θ_2 with reference prior



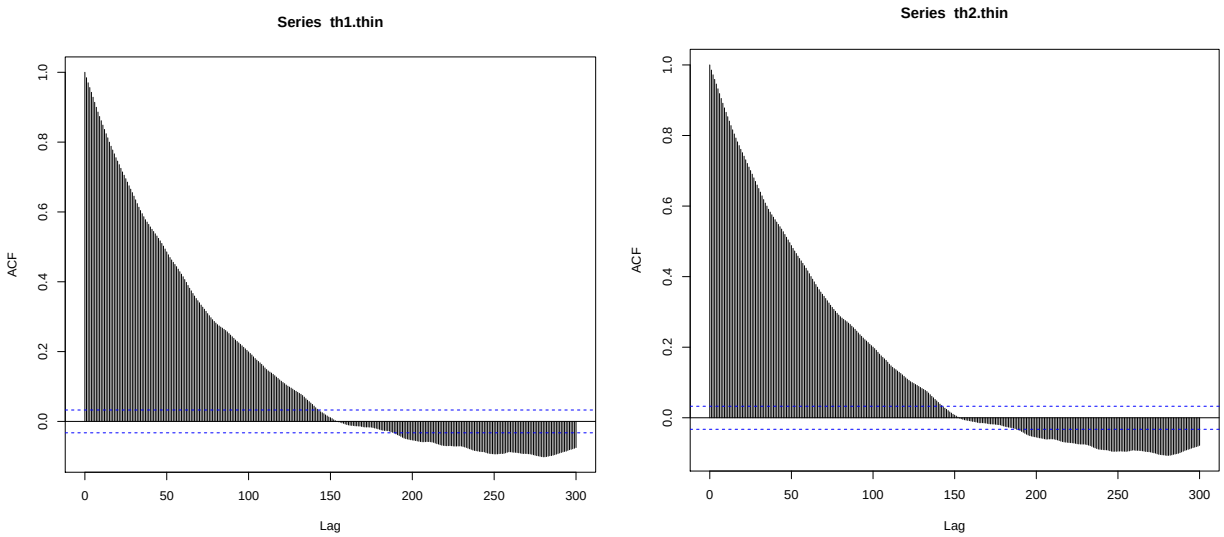
(a) θ_1 (b) θ_2

Figure 2.13: Scenario 2: Histogram for posterior distribution of θ_1 & θ_2 with unoiform prior

2.4.2 Scenario 2

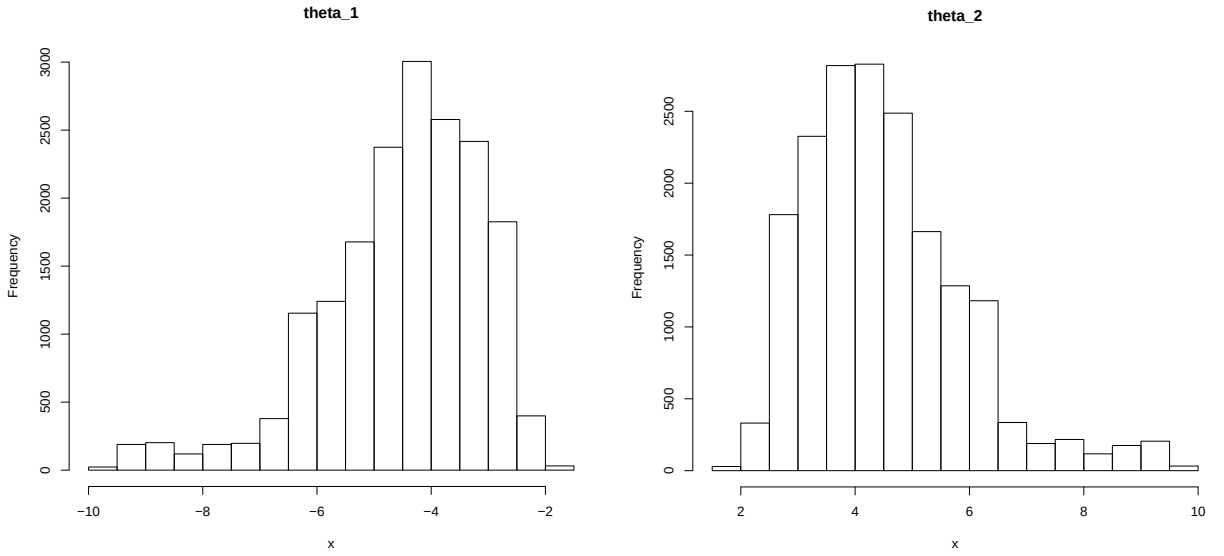
In the second scenario, we set $\theta_1 = -5$ and $\theta_2 = 5$. We ran the iteration 40,000 times with a burn-in period of 20,000. We took a proposal distribution of $N(0, (0.3)^2)$, which makes the acceptance rate of the MCMC sampler stay around 20%. The posterior output of the MCMC chain, along with the auto correlation plot and the MCMC iterations, is shown from Figure 2.13a to Figure 2.16b. We compute the posterior using two different default priors, first with $U(-\infty, \infty)$ and the second with the reference prior for the two-parameter dyadic independent model. For decreasing the autocorrelation between the MCMC samples, we apply a thinning process with $n = 5$.

It can be seen comparing the different prior setup in the two different scenarios that the autocorrelation for the MCMC iterations after thinning is decreasing faster (at around lag 50 in the first scenario and at around 100 in the second scenario) in the reference prior setup than the uniform setup (at around lag 100 at the first scenario and at around 150 in



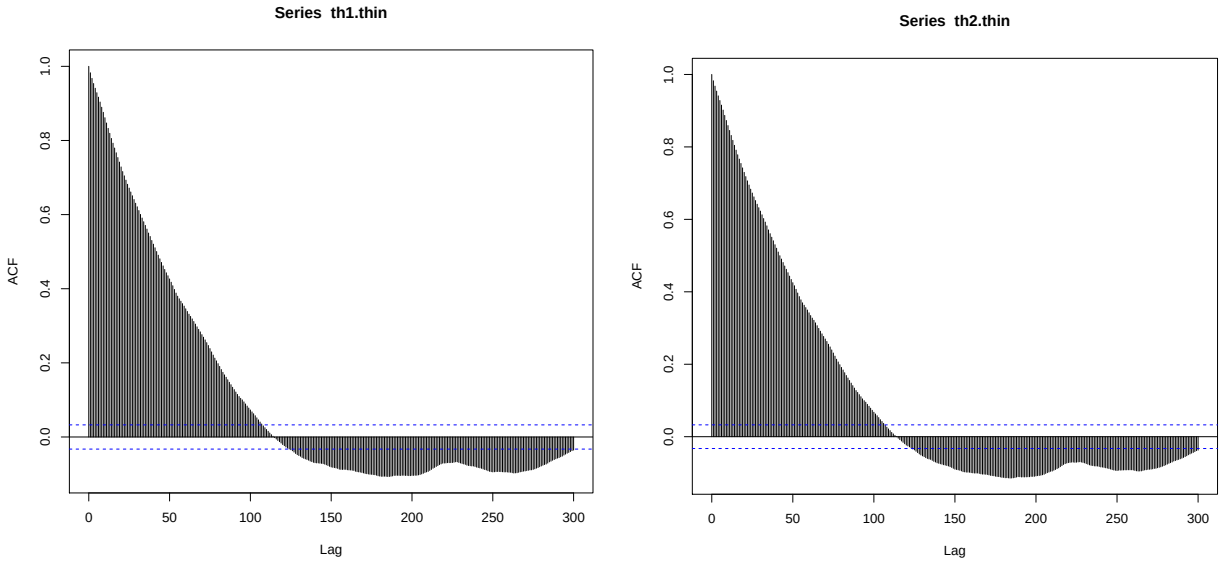
(a) θ_1 (b) θ_2

Figure 2.14: Scenario 2: Auto-correlation plot for the MCMC iteration of θ_1 & θ_2 with uniform prior



(a) θ_1 (b) θ_2

Figure 2.15: Scenario 2: Histogram for posterior distribution of θ_1 & θ_2 with reference prior



(a) θ_1 (b) θ_2

Figure 2.16: Scenario 2: Auto-correlation plot for the MCMC iteration of θ_1 & θ_2 with reference prior

Table 2.1: Table for Posterior Mean and Standard Deviation

Scenario's	Parameters	Posterior-Mean	Posterior-SD
SCENARIO 1			
Uniform Prior	θ_1	-3.672	0.982
	θ_2	3.619	0.974
Reference Prior	θ_1	-2.530	0.922
	θ_2	2.548	0.936
SCENARIO 2			
Uniform Prior	θ_1	-5.229	1.560
	θ_2	5.220	1.551
Reference Prior	θ_1	-4.490	1.464
	θ_2	4.510	1.448

the second scenario). Thus, we need more iterations compared to reference prior setup in uniform setup to achieve effectively independent draws.

2.5 Discussion

This paper is dedicated to developing a default prior setup in a Bayesian modeling scenario that not only overcomes the issue of sensitivity of the informative priors but also integrates to optimize the posterior estimates through an information theoretic setup. Although the paper has developed a two-parameter dyadic independent network model, it can also be extended to more complicated network models.

The Metropolis-Hastings algorithm in ERGM setup is much more challenging since ERGM suffers from the problem of model degeneracy where the iterations can converge to a full or empty graph. Poor choice of the initial values of the model parameters can lead the iterations to a full or empty network. To avoid this situation, we have used a weak thresholding for the number of edges of the network.

We have discussed that the Bayesian technique applied to a large network data can be effective since we can easily avoid the problem of obtaining the normalizing constant. Moreover, the application of the Tie-No-Tie algorithm in a large network becomes effective since it uses a random sample from the set of edges and empty dyads that are assumed to be a sub-sample of the observed network and are distributed as the proposed model. So TNT provides a strong theoretical background of an alternative to the traditional MCMC iteration. Along with that, it makes the MCMC iterations way faster since it always considers a subsample to perform each iterations no matter how big the network is. Our analysis also shows that use of a reference prior provides an added benefit by generating the independent samples from the posterior distribution faster than the other default priors.

Chapter 3

Big Data Application of ERGM through Reference Prior

3.1 Big Data Network

Recent advancements in computer, Internet and social media highlight the very important aspects of accessing and analyzing the user data for future developments. Network literature has broadened the area of user interfaces where not only can a user obtain his preferred products based on his inputs but also the manufacturers get an idea about what path they need to follow to get their product closer to each individual customer. For example, an Amazon purchaser can see the recommended product based on the product he has viewed or purchased. A probabilistic determination of the links is therefore important to optimize the sales of certain companies as well as to optimize the utility of the products of the purchasers.

Developments in online trading, shopping, and media services in the past few years have opened a gateway to analyze the characteristics of individual users through a massive data atmosphere. Careful analysis and handling of so much data is a challenging task that requires massive space and time and is still not applicable in many circumstances.

Network estimation through good statistical properties of the estimates is a popular field of study in recent years. One of the important network models that can structurally model the network through nice statistical properties is the Exponential Random Graph Model (ERGM). Although a moderately large network can create trouble in estimation of parameters in ERGM model, Bayesian estimation can overcome this issue through bypassing the calculation of the normalizing constant. But since the estimation procedure involves MCMC iterations through the Metropolis-Hastings technique, the estimation procedure becomes very slow and a good mixing becomes challenging even for a sparse network of size 500×500.

Not much work has been done to handle large sparse networks in an ERGM setup.

Thiemichen and Kauermann (2016) have proposed an ERGM model with non-parametric components to estimate large networks through subsampling techniques, but this method is limited to dense networks only. He and Zheng (2015) proposed an estimating technique of large social network through graph limits for ERGM, but their method also fails for large sparse networks where it is shown that the graph limit tends to zero in a sparse situation.

This chapter is dedicated to estimate a sparse Big Data Amazon co-purchasing network with 262111 nodes through the application of TNT procedure and through tuning the probability of edge set and empty dyad set. We implement the reference prior technique as we have seen that the ACF function drops faster in the reference prior scenario and so we obtained the random samples from the posteriors faster than the other prior case.

3.2 Data and Model

For a Big Data implementation of ERGM, we consider an Amazon co-purchasing network from the Stanford SNAP data repository where the network is based on the "Customers Who Bought This Item Also Bought" feature of the Amazon website. We consider 262111 nodes, and it can be considered as a sparse network since the observed network has 1048575 directed links and 562240 reciprocated links. The main idea of the Amazon recommendations is to pick buyers for similar types of items. For the purpose of the data analysis, we to assign a probability to each possible link based on the observed adjacency structure. That means if a buyer purchases or visits some item webpage, then the system would incorporate the information from the other buyers to pick an item similar to what they have purchased to show in the recommendation list.

For the implementation purposes of the ERGM model, we introduce the dyadic indepen-

dent ERGM model where the one directional and the reciprocated link are being captured. That means a pair of items listed on Amazon to be considered under modeling and directional and reciprocated links between them is going to be estimated. The influence of any third item will be considered to be conditionally independent of these two.

Hence consider the model,

$$f(\mathbf{Y} \mid \theta_1, \theta_2) \sim exp(\theta_1 \sum_{i \neq j} y_{ij} + \theta_2 \sum_{i \leq j} y_{ij} y_{ji}) \tag{3.1}$$

where y_{ij} being the observed adjacency between i^{th} and j^{th} item listed on Amazon. θ_1 is the single link parameter and θ_2 is the reciprocation parameter. Hence, $\sum_{i \neq j} y_{ij} = 1048575$ and $\sum_{i \leq j} y_{ij} y_{ji} = 562240$.

3.3 Estimation

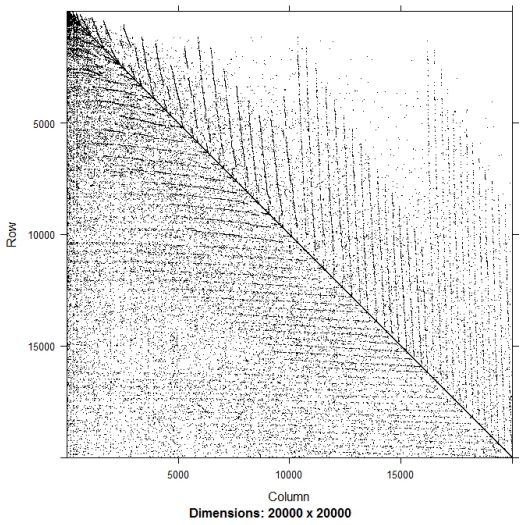


Figure 3.1: Adjacency plot for first 20,000 nodes for the observed Amazon Co-Purchasing network

Table 3.1: Table for Posterior Mean and SD for Amazon Data

Parameters	Posterior-Mean	Posterior-SD
θ_1	-6.581	0.229
θ_2	6.582	0.231

The key advantage of the TNT procedure is that it uses a small subset of the observed network at each MCMC iteration instead of observing the whole network, which makes the whole computation time much faster than the traditional MCMC. This method is applicable to large networks and facilitates good estimates based on information theoretic sense along with faster computation.

We divided the dyads into two sets, as we discussed before, and assigned equal probability to the set with edges and to the set of empty dyads. This specific assignment gives a decent level of mixing with a sparse estimate of the network. We take 5,000 MCMC iterations with a burn-in period of 4,000. We set our proposal distribution to be $N(0, (0.075)^2)$. The posterior mean of θ_1 is given by -6.581 and the posterior mean of θ_2 is 6.582.

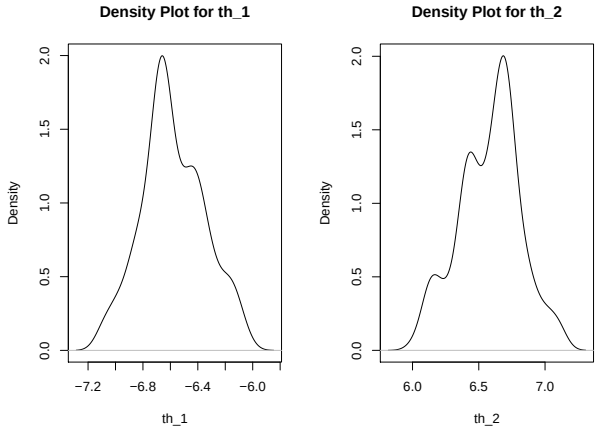


Figure 3.2: Density Plot for Dyadic independent Network Model parameters to estimate the Amazon Co-purchasing Network through Reference Prior

A key feature of the ERGM model is the nodes are considered equivalent. To check whether the sparseness of the observed network is sustained in the estimate, the idea is to randomly pick dyads from the network and try to re-assign or drop the edges based on the estimated model. For that purpose, we picked 100000 dyads and tried to reassign the directed and the reciprocated edges. When we tried to reassign those edges, we saw that the error rate in terms of sparsity was 5.233% for the directed edges and 5.226% for the reciprocated edges.

Although the network has 262111 nodes, which means the adjacency matrix can have 262111×262111 edges, the computation time was very fast compared to the size of the adjacency matrix. Although we had to choose the size of the sample, we needed to select at each iteration step, as a moderate size of samples gives a good mix along with fast computation time. In our example, we ran the iterations by taking 1000 random sample of dyads at each step and assigning equal probability to the edges set and the empty dyad set. We also can tune the probability based on the degree of sparsity, along with our predefined value of the sample to be chosen at each iteration that might have better mixing.

3.4 Discussion

It is important to observe that such a low estimate of θ_1 shows the sparsity property of the network is valid for the assumed family. Also we observed a high value for the estimate of θ_2, which meant most of the estimated directional links were essentially reciprocating and if i^{th} the item is being shown in the recommendation for the j^{th} item, then it is true in the other way, which makes sense in a product recommendation scenario.

This idea can be extended to more complex parameter situation where we bring more

structural ties in the model. Bringing some interesting ties between the node can explore a detailed social characteristics of the buyer and help to decide the optimized recommendations for him. In our modeling scenario, a low value of the directional edge parameter indicates some specific tendencies of the buyers and it reveals that a set of buyer can easily be separated into various categories of the products and random recommendations can be irrelevant according to the nuyers choice. For example, recommending kids toys could be irrelevant to a student who bought books or a laptop in his last couple of purchases.

Again a very high value of reciprocating parameter is very meaningful in the sense that it can help clustering a group of buyers who are buying different kind of products of similar categories. For example, Buyer1 who bought a desktop computer in his last purchase may end up buying a printer of it. Also, Buyer2 who bought a printer in his last purchase might have an old computer and he may be interested in replacing his old computer and may be interested in the computer that Buyer1 purchased. Hence Buyer1 and Buyer2 may be purchasing different products on that day but the network characteristic reveals that they belong to the same cluster of buyers who are buying computer related products. Hence, a very simple two parameter dyadic independent model gives us a direction for the companies need to make to optimize their product recommendation. Hence, inclusion of more complex graphical structure could be benificial for optimizing the sells for the manufacturers as well as the product utilities for the buyers.

The key advantage of applying the TNT algorithm in the ERGM and reference prior is that it makes the computation fast since, even for a large sparse matrix, the computation mostly depends the size of the random samples drawn from the two sets as well as the prespecified probabilities for these two sets. Allowing us to tune the set probabilities helps with the controling the mixing of the MCMC chain since the chain does not spend most of its time

on the empty sets by being sparse. Although it is very challanging to obtain the optimized value of the sample size of the dyads at each iteration steps along with the optimized value of the tunning probability that makes the posterior MCMC samples closest to the actual posterior and makes the computation very fast.

Although the nodes are considered equivalent in ERGM, the key advantage of the implementation of the ERGM technique is that we can pre-specify the graph structure we are interested in and can determine if a specific structure is dominating or has no influence in the formation of the links.

Chapter 4

Latent Space Network for three US Auto Manufacturing Giant

4.1 Main Idea and Methodology

There is growing interest in the use of network models to represent relations between actors or nodes. In social networks, these actors are individuals, while in the inter-firm linkages context, they are firms. Sometimes the two come together, for example, in the literature on director networks; for a recent discussion from a more general context, see Borgatti et al. (2009). This paper examines interfirm networks in the US automobile industry, focusing on the 3 manufacturing giants - Chrysler, Ford and General Motors - and 21 of the most prominent intermediaries in the sector. The analysis is based on a model of operating leverage (Bhattacharjee et al. 2014), asking the question: how much do firms benefit (or lose) from network externalities in sharing firm-specific financial risk? The results point to a combination of explanations: corporate governance linkages, supply chain networks and potentially demand side linkages as well. Therefore, this chapter suggests a wholistic approach in the analysis if inter-firm networks, combining a number of channels (or drivers) and disciplinary approaches.

Recent studies on inter-firm networks have revealed various characteristics of the firms, its managers and its ownership, and how this influences it's neighboring firms. In turn, neighborhood has been captured by linkages along the supply chain (Hertzel et al., 2008; Wang, 2012; Ahern and Harford, 2014; Itzkowitz, 2015), director networks (Renneboog and Zhao, 2011, 2014), or joint ventures and investment syndication (Wang and Wang, 2012). Questions have been asked about whether supply-side or demand-side linkages drive interactions between firms (Ellis and McGuire, 1993; Venables 1996), and about the nature of the networks themselves - whether cohesive networks of socially embedded ties or sparse networks rich in structural holes (Grandori and Soda 1995; Hite and Hesterly 2001). Recent

research in finance has explored the functioning of the financial market placing emphasis on strategic interactions between firms and their industry rivals and suppliers. In this setting, economic or financial shocks are transmitted through the customer supplier network and the whole industry can be affected by those shocks since these can move through the links of firms in that industry, and beyond; see, for example, Ahern and Harford (2014). In particular, Vickery et al. (1999) and Narasimhan and Jayram (1998) argue that sales volatility, disruptions and opportunities in the firm's supply and demand environments are important aspects of the financial market.

In terms of methodology, network methods based on graph theoretic frameworks have helped document a positive link between network structures and firm performance (Geletka nycz and Boyd 2011; Larcker et al., 2013). The key mechanism is that a strong network provides better access to information which then brings benefits to a firm in its decision making (Larcker and Tayan, 2010; Omer et al., 2012). Then, this framework helped researchers reveal previously hidden relationships between the connection of the corporate elite and board room issues such as decision making on managerial compensation, investments, and hiring and firing of top management. Rebbeboog and Zhao (2011) and Horton et al. (2012) demonstrated that a CEO's direct and indirect connections affect his power and the value of his information-connections, which is reflected in higher remuneration.

Indeed, a relationship between inter-firm linkages and business performance has been the focus of many studies. Vickery et al. (1999) found significant relationships between supply chain flexibility and different measures of performance in the US furniture industry. Likewise, Vonderembse and Tracy (1999) observed a link between supplier selection criteria and manufacturing performance. Narasimhan and Jayram (1998) argued that research in supply chain management tends to focus on the individual functions and fails to examine

the causal linkages that comprise the supply side of the economy. Swaminathan et al. (1998) emphasized the importance of demand forcasting in the supply chain dynamics.

In this paper we develop a method based on network data that captures the latent linkages between firms based on a model of operating leverage, and which then influences various financial decision making issues and financial strategies for those firms. Network data typically consist of a set of N nodes and a relational tie y_{ij} measured on each ordered pair of nodes. This framework has many applications in social network literature. The most simplest situation is when y_{ij} is a dichotomous variable indicating the presense or absense of some connection between nodes (in our case, firms) i and j. The data are often represented by an $N \times N$ social-interaction matrix or the adjacency matrix $\mathbf{W}$.

In our proposed model, the adjacency structure of the network with N firms is not observed but it depends on the latent positions $Z = (z_1, z_2, ..., z_N)$. Following Bhattacharjee et al. (2014), we apply a model of financial leverage where a firm can anticipate part of the variation in its sales turnover and the reaction of costs to these fluctuations in sales. This is because there is an equilibrium profit margin and an error correction model that captures partial adjustment to this equilibrium. We estimate this panel error correction model and extract residuals which in turn capture the reaction to an unanticipated change in sales. What remains in the error after systematic effects have been removed are the effect of inter-firm linkages that bring positive or negative risk management externalities to the firm. Then, in the second step, we use the covariance structure of these errors between firms to implement the latent space algorithm. We implement our model to estimate the network structure between the US auto industry firms, where our data constitute the costs and sales turnover for Ford, GM and Chrysler, together with 21 suppliers. The objective is to estimate the operational linkages between the three automobile manufacturers and their

main independent suppliers. In effect, we go beyond the regression modelling approach of Ramcharran (2001) to infer on interactions between firms using their latent positions in two dimentional Euclidean space. The results provide exciting new evidences on inter-firm networks that can then be interpreted based on the operating financial, organizational and governance structures of these firms.

The US automobile industry is large and consists of hundreds of firms, a small nuber of which are auto-manufacturers, and the vast majority auto-ancilliaries that supply various components. Our empirical objective is to analyze the structure and interaction between firms in the industry focusing on the latent inter-firm network. Following Ramcharran (2001), we focus attention on 3 major manufacturers - Chrysler, Ford and General Motors, together with the top 21 suppliers that were listed most frequently in the various issues of Ward's Automotive Yearbook. Annual data on sales turnover and costs (of sales) are collected for the period of 1950 through 2013 from the Compustat database. These constitute the basic data for our empirical work. In addition, we also use information from Bloomberg SPLC database on the supply chain network for Ford.

Economic links of manufactures to their suppliers and customers constitutes our baseline characterization of inter-firm networks. Inter-firm linkages influence the actions of suppliers and customers of firms in distress (Hertzel et al., 2008; Wang, 2012). Suppliers can impose costs by failing supply trade credit, backing away from entering into long term contracts, delaying shipments, sourcing new customers or shifting sales away from the distressed firm and existing customers. Likewise, inter-firm links through corporate governance channels (for example, director networks) can enhance or reduce credit constraints on firms (Renneboog and Zhao, 2011, 2014). Many important aspects of the auto industry market can be influenced by the corporate policies by the firms which could be driven by the latent linkages

between the firms. Firms financial decisions can be a direct consequence by a negative or positive links they have with the other firms.

Inter firm linkages can also influence the suppliers to switch to different customers and also can have a significant stock price effect when industry rivals have a positive link with the same customer. Finally, inter-firm networks can be related to dividend policies (Wang, 2012) and relationship specific investment (Wang and Wang, 2012).

4.1.1 Error Correction Model

We have the cost of goods sold data Y_{it} observed for $i = 1, 2, ..., N$ companies and $t = 1, 2, ..., K$ years. X_{it} is the sales turnover for the company i for $t = 1, 2, ..., K$ years. To handle the non-stationarity of the data, consider the panel error correction model,

$$\Delta Y_{it} = \alpha_i + \lambda_i \Delta X_{it} + (1 - \lambda_i)(Y_{i,t-1} - \theta_i X_{i,t-1}) + \eta_{it} \tag{4.1}$$

$$\underset{\sim}{\eta_t} = \underset{\sim}{w}' \eta_t + \underset{\sim}{\epsilon_t} \tag{4.2}$$

where $\mathbf{W} = (\underset{\sim}{w}'_1, \underset{\sim}{w}'_2, ..., \underset{\sim}{w}'_N)'$ is a symmetric adjacency matrix with $w_{ij} = 1$ if node i and j has edge between them and 0 otherwise, with $w_{ii} = 0$ for $1 \leq i, j \leq N$ and $(I - \mathbf{W})$ being singular. The parameter $(1 - \lambda_i)$ determines the speed at which the system corrects back to the equilibrium relationship $Y_{i,t-1} - \theta_i X_{i,t-1}$ after a sudden shock. We assume the errors ϵ_{it} are iid across time. Under the non-singularity assumption of $(I - \mathbf{W})$ we have,

$$E(\underset{\sim}{\eta}.\underset{\sim}{\eta}') = (I - \mathbf{W})^{-1} \mathbf{\Sigma} (I - \mathbf{W})^{-1'}$$

where $E(\epsilon' \epsilon) = diag({\sigma_1}^2, {\sigma_2}^2, ..., {\sigma_N}^2)$.

Under the spatial error model described by (4.1) and (4.2), the population spatial autocovariance matrix $E(\eta^{'}\eta)$ is unknown and positively definite with probability one. There exists a consistent estimator, $\hat{\mathbf{\Gamma}}$, of the population spatial auto covariance matrix $E(\eta^{'}\eta)$.

Thus, Bhattacharjee and Jensen-Butler (2013) has made an estimation method for the underlying regression model. Based on the residual from these estimates, a consistent estimator for the spatial auto-covariance matrix is first obtained. This estimator is then used to estimate the spatial weight matrix. They showed that under the previous assumptions, the matrix

$$\mathbf{V} = (I - \mathbf{W})^{'}.\left(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, ..., \frac{1}{\sigma_N}\right)$$

is consistently estimated up to an orthogonal transformation by

$$\hat{\mathbf{\Gamma}}^{-1/2} = \hat{\mathbf{E}}\big(\hat{\mathbf{\Lambda}}^{-1/2}\big)\hat{\mathbf{E}}^{'}$$

where $\hat{\mathbf{E}}$ and $\hat{\mathbf{\Lambda}}$ contain the eigenvectors and the eigen values respectively, of the estimated spatial autocovariance matrix $\hat{\mathbf{\Gamma}}$.

After getting an estimate of the spatial weight matrix, we use it to generate Bootstrap samples for the spatial weight matrix to perform the following multiple testing problem:

$$H_{0ij} :\mid w_{ij} \mid = 0 \quad vs \quad H_{1ij} :\mid w_{ij} \mid = 1 \quad \forall \quad 1 \leq i < j \leq N. \tag{4.3}$$

We use this as an initial estimate of $\mathbf{W}$ for implementing the latent space model.

4.1.2 Latent Space Model

In the second stage, to assign a latent position for each company in the Euclidean space, we take a conditional independence approach to modeling by assuming the presence or absence of a tie between two nodes is independent of all other ties, given the unobserved positions in the latent space of the two nodes.

$$\mathbb{P}(\mathbf{W} \mid Z_1, \alpha_1, \beta_1, \gamma_1) = \prod_{i,j} \mathbb{P}(w_{i,j} \mid z_{1i}, z_{1j}, \alpha_1, \beta_1, \gamma_1)$$

Here Z is the unknown latent positions.

We fit a logistic regression model as below,

$$logodds(w_{i,j} = 1 \mid z_{1i}, z_{1j}, \alpha_1, \beta_1, \gamma_1) = \alpha_1 - \beta_1 f(\underset{\sim}{\eta_i}, \underset{\sim}{\eta_j}) - \gamma_1 f(z_{1i}, z_{1j})$$

$f(z_i,z_j)$ can be replaced by any arbitrary set of distances $d_{i,j}$ satisfying the triangle inequality. In general, we prefer to model the $d_{i,j}$'s as distances in some low-dimensional Euclidean space for reasons of parsimony and ease of model interpretability.

The latent position model is inherently reciprocal and transitive. If $i \to j$ and $j \to k$ then d_{ij} and d_{jk} are not too large, which implies the event $j \to i$ (reciprocity) and $i \to k$ (transitivity).

Here we set our distance model as,

$$logodds(\mid w_{i,j} \mid = 1 \mid z_{1i}, z_{1j}, \alpha_1, \beta_1, \gamma_1) = \alpha_1 - \beta_1 \mid\mid \underset{\sim}{\eta_i} - \underset{\sim}{\eta_j} \mid\mid - \gamma_1 \mid z_{1i} - z_{1j} \mid \qquad (4.4)$$

Here α_1 is the additive constant. The model implies that reducing the latent distance

between node i and j will increase the log odds of node i and j to be connected. Similarly, reducing the Euclidean distance between the model errors of node i and j will increase the log odds of node i and j to be connected. β_1 and γ_1 are the scaling factors for the corresponding distance measures.

To identify the positive and the negative linkages among the significant links, we set up a nested model where we assume having a positive and negative connection between two nodes is independent of all other ties, given the latent positions of the two nodes and on the significance of the linkages. We can write,

$$logodds(w_{i,j} = 1 \mid z_{2i}, z_{2j}, \alpha_2, \beta_2, \gamma_2, \mid w_{i,j} \mid = 1) = \alpha_2 - \beta_2 \mid\mid \underset{\sim}{\eta_i} - \underset{\sim}{\eta_j} \mid\mid - \gamma_2 \mid z_{2i} - z_{2j} \mid \quad (4.5)$$

4.2 Estimation

Distances between a set of points in Euclidean space are invariant under rotation, reflection, and transition. Hence, an infinite number of latent positions Z give the same log-likelihood. Specifically, $logP(\mathbf{W} \mid Z, \theta) = logP(W \mid Z^*, \theta)$ for any Z^* that is equivalent to Z under the operations of reflection, rotation, or transition. This creates an identifiability issue of for the latent positions. Our model involves a one-dimensional Euclidean space, but since each dimension is being identified separately through two different models, we need to fix two z_i's to overcome the reflection, rotation, or transition issue for the latent positions.

We assume,

$$z_{i1},...,z_{in} \sim i.i.d\ N(0, \sigma_{z_i})\ \forall i = 1, 2$$

We formulate mutually independent priors for α_1, α_2, β_1, β_2, γ_1, γ_2, ${\sigma_{z_1}}^2$, ${\sigma_{z_2}}^2$ as,

$$\pi(\alpha_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{{\alpha_i}^2}{\sigma^2_{\alpha_i}}} \qquad -\infty < \alpha_i < \infty \qquad \forall i = 1, 2$$

$$\pi(\beta_i) = \frac{1}{2}\frac{1}{\sqrt{2\pi}} e^{-\frac{{\beta_i}^2}{\sigma^2_{\beta_i}}} \qquad 0 \leq \beta_i < \infty \qquad \forall i = 1, 2$$

$$\pi(\gamma_i) = \frac{1}{2}\frac{1}{\sqrt{2\pi}} e^{-\frac{{\gamma_i}^2}{\sigma^2_{\gamma_i}}} \qquad 0 \leq \gamma_i < \infty \qquad \forall i = 1, 2$$

$$\pi({\sigma_{z_i}}^2) = \frac{2^{-\frac{\nu_1}{2}}}{\Gamma(\frac{\nu_1}{2})} e^{-\frac{1}{2{\sigma_{z_i}}^2}} {\sigma_{z_i}}^{-\nu_1 - 2} \qquad {\sigma_{z_i}}^2 > 0 \qquad \forall i = 1, 2$$

The detailed Gibbs steps for the posterior computations are given in section 4.5.

4.3 Data Analysis

We consider COMPUSTAT data for past 64 years starting from 1950 to 2013. The data consists of Cost of Goods Sold and Sales Turnover for three major U.S.-based auto manufacturers: (1) GM, (2) Ford, and (3) Chrysler and their 20 major suppliers.

We first divide the two variables with the corresponding consumer price index to remove the scale factors of dollar values over the years. We then implement the error correction model (4.1) and extract the model errors and use the errors to calculate the correlation matrix (shown in Figure 4.1), and then we use the correlation matrix to calculate a bootstrap sample from $\mathbf{W}$. We need to get the bootstrap since we have to test for each element of the $\mathbf{W}$ matrix. We assign 0 to w_{ij} if we fail to reject the ij^{th} test and 1 otherwise. We use this estimated W as the adjacency structure between the companies for our latent space

model. For the issues of rotation, reflection, and transition, we keep the latent position of GM and Ford to be fixed to some predefined values. We Run the Gibbs update for the parameter values for 3,000 times with a burn-in period of 2,000. We achieved an overall nice convergence at around 1,000 iterations.

Figure 4.2 shows the probability of having no connection (shown in blue), a positive connection (shown in green), and a negative connection (shown in red) by the three major auto manufacturers and their key suppliers. Here a link between company i and j signifies the business impact of i^{th} company and j^{th} company either ways with respect to the auto industry market. It is evident from Figure 4.2 that the network that the three big manufacturers have with their suppliers is moderately dense with a high probability of being connected with their suppliers. Also, a similarity between Chrysler and Ford can be seen with respect to their connectedness. This is evident in Figure 4.3, which provides the latent positions of the companies, and it can be seen that Ford and Chrysler are close to each other in the latent space. The probabilities for GM imply that, although having the same set of suppliers as Ford or Chrysler, GM relies on a select number of suppliers with respect to the auto industry market with a very large probability of being disconnected from a few of these suppliers.

A careful inspection of the latent positions of the suppliers and three major companies reveals an important aspect of the auto industry market. If we look at the position of Ford and Chrysler, they are sitting in the middle of their suppliers and moderately depend on almost all of them with respect to their business. Again, the position of GM in the latent space reveals that GM has a specific subset of suppliers that it relies on for its business.

	ALCAN	HONEYWEL	ARVIN	CHRYSLER	COOPER TIRE	DANA	DEERE	EATON	FORD	GE	GM	SPX	GOODRICH	GOODYEAR	JOHNSON (	KUHLMAN	AEROQUIP-	OWENS	PPG	ROCKWELL	SMITH (A O)	TRW	UNITED TECH	WALBRO
ALCAN	1	0.28	0	-0.18	-0.27	0.13	-0.11	0.18	-0.1	-0.09	-0.14	0.09	0.33	0.03	0	-0.02	-0.01	0.3	0.25	0.02	0.04	0.16	0	0.09
HONEYWEL	0.28	1	-0.03	-0.35	0.08	0.19	0.26	-0.11	0.1	-0.17	-0.01	0.14	0.08	0.03	-0.05	-0.08	0.23	0.08	0.18	-0.07	0.14	-0.33	0.02	-0.09
ARVIN	0	-0.03	1	0.08	0.22	0.02	0.33	0.15	-0.08	0.12	0.07	0.22	0.09	-0.04	0.36	0.08	0.32	0.11	0.36	0.13	0.27	0.12	-0.03	0.05
CHRYSLER	-0.18	-0.35	0.08	1	0.32	0.17	0.35	0.13	0.28	0.36	0.47	0.04	0.07	0.14	0.03	-0.4	0.08	0.01	0.14	0.18	0.24	0.32	0.17	0.05
COOPER TIRE	-0.27	0.08	0.22	0.32	1	0.2	0.34	-0.09	0.16	0.04	0.13	-0.04	-0.06	0.12	0.07	-0.19	0.36	0.12	0.3	-0.11	0.15	0.04	0.05	0
DANA	0.13	0.19	0.02	0.17	0.2	1	0.23	0.11	0.27	-0.24	0.21	0.11	-0.19	-0.1	0.04	0.39	0.23	0.25	0.32	-0.07	0.2	0.04	0.13	-0.3
DEERE	-0.11	0.26	0.33	0.35	0.34	0.23	1	0.15	0.23	0.2	0.41	0.04	0.11	-0.02	0.3	-0.09	0.22	0.1	0.19	-0.03	0.32	-0.13	0.23	0.13
EATON	0.18	-0.11	0.15	0.13	-0.09	0.11	0.15	1	0.08	0.11	0.05	0.12	0.35	0.11	0.13	-0.03	-0.11	0.27	0.07	0.18	0.01	0.15	0.34	0.4
FORD	-0.1	0.1	-0.08	0.28	0.16	0.27	0.23	0.08	1	0.05	0.1	0.19	0.23	0.08	0.2	-0.35	0.15	0.11	0.22	0.16	0.28	0.1	0.05	0.03
GE	-0.09	-0.17	0.12	0.36	0.04	-0.24	0.2	0.11	0.05	1	0.07	-0.09	0.17	-0.05	0.13	-0.24	0.13	0.2	0.23	0.28	0.01	0.02	0.09	0.24
GM	-0.14	-0.01	0.07	0.47	0.13	0.21	0.41	0.05	0.1	0.07	1	0.05	-0.05	0.02	0.01	-0.21	0.26	0.25	0.08	-0.13	-0.11	-0.05	0.03	-0.17
SPX	0.09	0.14	0.22	0.04	-0.04	0.11	0.04	0.12	0.19	-0.09	0.05	1	0.05	-0.04	0.09	-0.05	0.17	0.02	0.12	-0.12	0.22	-0.05	-0.03	-0.43
GOODRICH	0.33	0.08	0.09	0.07	-0.06	-0.19	0.11	0.35	0.23	0.17	-0.05	0.05	1	0.22	0.04	-0.32	0.14	0.42	0.13	0.08	0.05	0.02	0.03	0.31
GOODYEAR	0.03	0.03	-0.04	0.14	0.12	-0.1	-0.02	0.11	0.08	-0.05	0.02	-0.04	0.22	1	0.2	-0.11	0.11	0.09	-0.07	0.15	-0.15	0.05	-0.18	0.01
JOHNSON CONTROLS	0	-0.05	0.36	0.03	0.07	0.04	0.3	0.13	0.2	0.13	0.01	0.09	0.04	0.2	1	0.02	0.2	0.28	0.03	0.11	0.2	-0.12	0.02	-0.1
KUHLMAN	-0.02	-0.08	0.08	-0.4	-0.19	0.39	-0.09	-0.03	-0.35	-0.24	-0.21	-0.05	-0.32	-0.11	0.02	1	-0.03	-0.1	-0.09	0.49	-0.07	0.13	-0.18	-0.47
AEROQUIP-VICKERS	-0.01	0.23	0.32	0.08	0.36	0.23	0.22	-0.11	0.15	0.13	0.26	0.17	0.14	0.11	0.2	-0.03	1	0.27	0.38	0.11	0.07	-0.42	-0.23	-0.19
OWENS CORNING	0.3	0.08	0.11	0.01	0.12	0.25	0.1	0.27	0.11	0.2	0.25	0.02	0.42	0.09	0.28	-0.1	0.27	1	0.39	-0.01	0.12	0.01	0.08	0.22
PPG	0.25	0.18	0.36	0.14	0.3	0.32	0.19	0.07	0.22	0.23	0.08	0.12	0.13	-0.07	0.03	-0.09	0.38	0.39	1	0.16	0.52	0.08	0.01	0.19
ROCKWELL	0.02	-0.07	0.13	0.18	-0.11	-0.07	-0.03	0.18	0.16	0.28	-0.13	-0.12	0.08	0.15	0.11	0.49	0.11	-0.01	0.16	1	0.1	0.31	0.05	-0.11
SMITH (A O)	0.04	0.14	0.27	0.24	0.15	0.2	0.32	0.01	0.28	0.01	-0.11	0.22	0.05	-0.15	0.2	-0.07	0.07	0.12	0.52	0.1	1	0.38	0.03	0.06
TRW	0.16	-0.33	0.12	0.32	0.04	0.04	-0.13	0.15	0.1	0.02	-0.05	-0.05	0.02	0.05	-0.12	0.13	-0.42	0.01	0.08	0.31	0.38	1	0.25	0.1
UNITED TECH	0	0.02	-0.03	0.17	0.05	0.13	0.23	0.34	0.05	0.09	0.03	-0.03	0.03	-0.18	0.02	-0.18	-0.23	0.08	0.01	0.05	0.03	0.25	1	0.49
WALBRO	0.09	-0.09	0.05	0.05	0	-0.3	0.13	0.4	0.03	0.24	-0.17	-0.43	0.31	0.01	-0.1	-0.47	-0.19	0.22	0.19	-0.11	0.06	0.1	0.49	1

Figure 4.1: Correlation Matrix

	Chrysler			Ford			GM		
	No connection	-1	1	No Connection	-1	1	No Connection	-1	1
ALCAN INC	0.291517297	0.197938233	0.51054447	0.283682277	0.198989743	0.51732798	0.156482066	0.274752536	0.568765398
HONEYWELL INTERNATIONAL INC	0.302061396	0.188080514	0.50985809	0.30788057	0.181106924	0.511012505	0.596504181	0.12552593	0.277969889
ARVIN INDUSTRIES INC	0.2233028	0.202005358	0.574691842	0.230466266	0.19996442	0.569569314	0.496734143	0.152956576	0.350309281
COOPER TIRE & RUBBER CO	0.424483339	0.149557083	0.425959578	0.418657797	0.156772403	0.4245698	0.178219435	0.256184303	0.565596262
DANA HOLDING CORP	0.513307516	0.125071289	0.361621195	0.504245092	0.12719639	0.368558518	0.234618984	0.232353896	0.53302712
DEERE & CO	0.484339548	0.133672629	0.381987823	0.476194464	0.137575602	0.386229934	0.213559192	0.238145197	0.548295611
EATON CORP PLC	0.296078748	0.18077951	0.523141743	0.304999332	0.179882089	0.515118579	0.592189664	0.123879027	0.283931309
GENERAL ELECTRIC CO	0.391101639	0.155886863	0.453011498	0.387136757	0.165369356	0.447493886	0.159575248	0.260249269	0.580175483
SPX CORP	0.438707751	0.147707465	0.413584784	0.446985	0.144634211	0.408380788	0.728999061	0.083698861	0.187302078
GOODRICH CORP	0.442677261	0.14450348	0.412819259	0.452143811	0.142250629	0.40560556	0.733237962	0.082148423	0.184613615
GOODYEAR TIRE & RUBBER CO	0.176738633	0.211836917	0.611424451	0.183379231	0.211820496	0.604800273	0.426279112	0.175470436	0.398250452
JOHNSON CONTROLS INC	0.4461163	0.142767313	0.411116387	0.455157273	0.140740013	0.404102713	0.735154945	0.080141915	0.18470314
KUHLMAN CORP	0.40026834	0.177783516	0.421948145	0.435670427	0.204954472	0.359375101	0.717530676	0.116891419	0.165577905
AEROQUIP-VICKERS INC	0.398316432	0.164347992	0.437335575	0.388291259	0.165862668	0.445846073	0.159849102	0.264104092	0.576046807
OWENS CORNING	0.357748892	0.169901284	0.472349824	0.365991599	0.166803176	0.467205225	0.655351639	0.105230611	0.239417749
PPG INDUSTRIES INC	0.333440904	0.172397222	0.494161874	0.324885239	0.174407806	0.500706954	0.127076551	0.267277579	0.605645869
ROCKWELL AUTOMATION	0.312176033	0.178259717	0.50956425	0.305743284	0.182019007	0.512237709	0.137553601	0.27039764	0.592048759
SMITH (A O) CORP	0.319396592	0.176080447	0.504522961	0.31156077	0.177814038	0.510625192	0.134144977	0.269698702	0.596156322
TRW INC	0.325370013	0.172101211	0.502528775	0.31850731	0.175389682	0.506103008	0.128498827	0.26358802	0.607913153
UNITED TECHNOLOGIES CORP	0.437424578	0.146701731	0.415873691	0.448030161	0.146077345	0.405892494	0.72865246	0.083671866	0.187675674
WALBRO CORP	0.493257883	0.133179593	0.373562524	0.502116276	0.129860892	0.368022831	0.769925945	0.070816385	0.15925767

Figure 4.2: Probability of linkages for Chrysler, Ford, and GM with their Suppliers

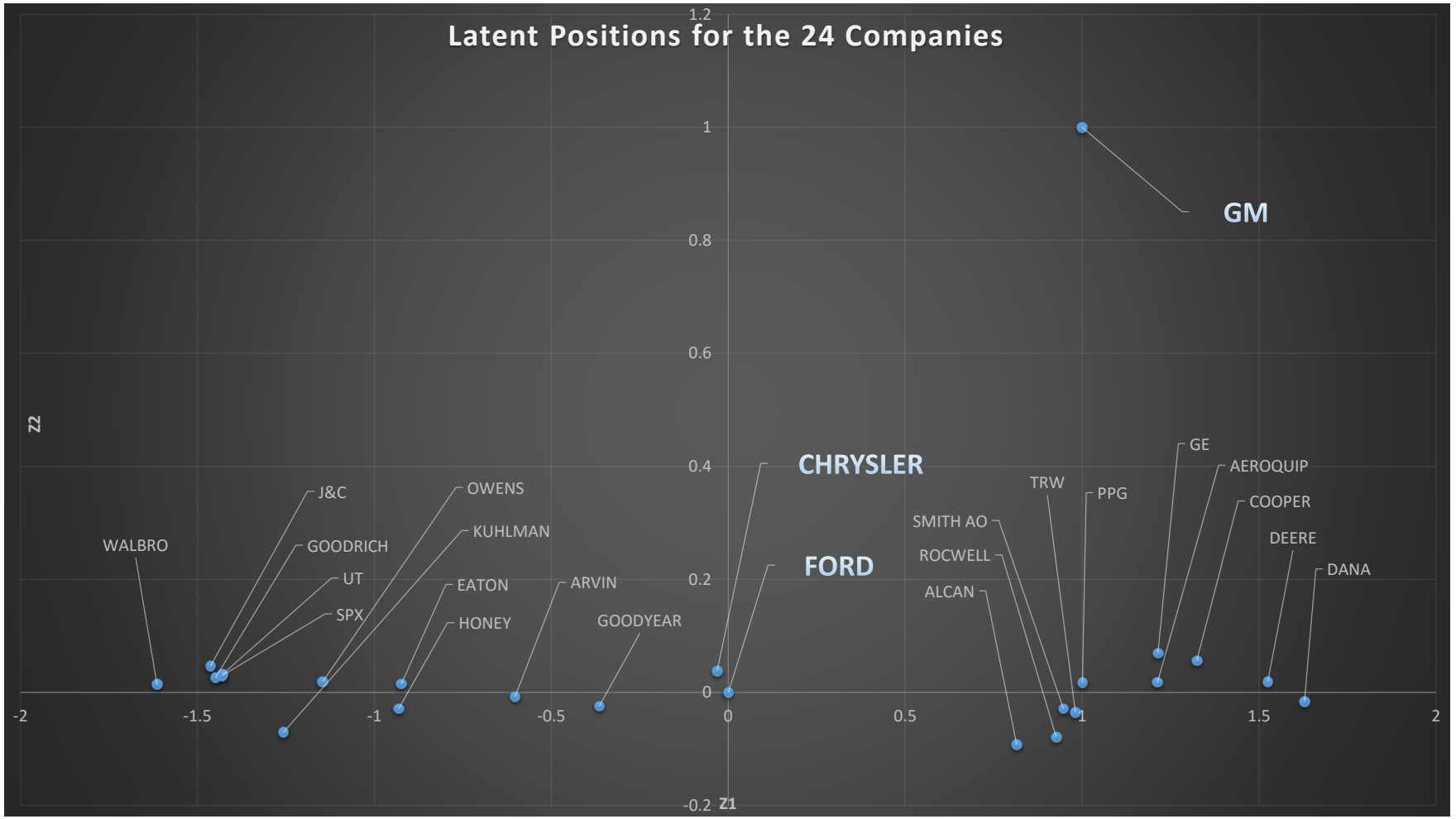


Figure 4.3: Position of the 24 companies in the latent space

4.4 Conclusion

The current literature in corporate finance highlights the importance of inter-firm networks in financial management of firms. Such networks can be based on supply chains, but equally may reflect director networks, joint ventures or demand side linkages. Hence, analysis of network structure requires a broad perspective that allows each of these potential drivers of network interactions to act and interact. We develop new Bayesian methodology to analyze latent inter-firm networks. Applied to data on the US auto industry, the estimated inter-firm networks reflect a strong influence of the supply chain, but also governance links between firms. Importantly, the estimated networks also point to both positive (complementary) and negative (competitive) interactuions between the firms. A lot of interesting questions emerge, relating to the impact of inter-firm networks on corporate finance issues.

4.5 Some Posterior Calculations

We set initial values for the parameters as $\alpha_{01}, \alpha_{02}, \beta_{01}, \beta_{02}, \gamma_{01}, \gamma_{02}, {\sigma_{0z_1}}^2, {\sigma_{0z_2}}^2$ and we update the parameters according to the following Gibbs steps.

4.5.1 $(s+1)$th Gibbs Step for updating $(\alpha_1, \beta_1, \gamma_1)$:

The full conditional distribution of $\alpha_1, \beta_1, \gamma_1$ is given by:

$$P(\alpha_1, \beta_1, \gamma_1 \mid \sigma^2_{z_1}, Z_1, \mathbf{W}, \eta)$$

$$\propto \left(\prod_{i>j} \frac{e^{(\alpha_1 - \beta_1 ||\underset{\sim}{\eta_i} - \underset{\sim}{\eta_j}|| - \gamma_1 |z_{1i} - z_{1j}|)|w_{ij}|}}{1 + e^{\alpha_1 - \beta_1 ||\underset{\sim}{\eta_i} - \underset{\sim}{\eta_j}|| - \gamma_1 |z_{1i} - z_{1j}|}} \right) \times e^{-(\frac{\alpha_1^2}{2\sigma^2_{\alpha_1}} + \frac{\beta_1^2}{2\sigma^2_{\beta_1}} + \frac{\gamma_1^2}{2\sigma^2_{\gamma_1}})}$$

$$= K_{\alpha_1, \beta_1, \gamma_1 | \sigma^2_{z_1}, Z_1, \mathbf{W}, \eta}$$

which is not a closed form expression of any distribution. Hence we need to perform Metropolis-Hastings with a symmetric proposal distribution for $\alpha_1, \beta_1, \gamma_1$. The r-th step for the Metropolis-Hastings algorithm is given by,

1. Generate ${\alpha_1}_s{}^{r*}$ from $N({\alpha_1}_s{}^{r-1}, \sigma_{met})$, ${\beta_1}_s{}^{r*}$ from $N({\beta_1}_s{}^{r-1}, \sigma_{met})$ and ${\gamma_1}_s{}^{r*}$ from $N({\gamma_1}_s{}^{r-1}, \sigma_{met})$.

2. Calculate:

$$u = \frac{K_{\alpha_{1s}^{r*}, \beta_{1s}^{r*}, \gamma_{1s}^{r*} | {\sigma^2}_{z_1 s}^{r-1}, Z_{1s}, \mathbf{W}, \eta}}{K_{\alpha_{1s}^{r-1}, \beta_{1s}^{r-1}, \gamma_{1s}^{r-1} | {\sigma^2}_{z_1 s}^{r-1}, Z_{1s}, \mathbf{W}, \eta}}$$

3. Set $\alpha_{1s}^r = \alpha_{1s}^{r*}$, $\beta_{1s}^r = \beta_{1s}^{r*}$, $\gamma_{1s}^r = \gamma_{1s}^{r*}$ with probability u otherwise continue with the value of step $r-1$ with probability $1-u$.

4. Repeat the above steps n_{met} times.

   Hence, we get n_{met} simulated samples from the distribution of $\alpha_1, \beta_1, \gamma_1 \mid \sigma^2_{z_1 s}, Z_{1s}$,

$\mathbf{W}, \eta$. We can use this simulated distribution update $\alpha_{1s}, \beta_{1s}, \gamma_{1s}$ to $\alpha_{1(s+1)}$, $\beta_{1(s+1)}, \gamma_{1(s+1)}$.

4.5.2 $(s+1)$th Gibbs Step for updating $(\alpha_2, \beta_2, \gamma_2)$:

The full conditional distribution of $\alpha_2, \beta_2, \gamma_2$ is given by:

$$P(\alpha_2, \beta_2, \gamma_2 \mid \sigma^2_{z_2}, Z_2, \{w_{ij} :| \ w_{ij} \ |= 1\}, \eta)$$

$$\propto \left(\prod_{i>j} \frac{e^{(\alpha_2 - \beta_2 ||\underset{\sim}{\eta_i} - \underset{\sim}{\eta_j}|| - \gamma_2 |z_{2i} - z_{2j}|)(1+w_{ij})/2}}{1 + e^{\alpha_2 - \beta_2 ||\underset{\sim}{\eta_i} - \underset{\sim}{\eta_j}|| - \gamma_2 |z_{2i} - z_{2j}|}} \right) \times e^{-(\frac{\alpha_2^2}{2\sigma^2_{\alpha_2}} + \frac{\beta_2^2}{2\sigma^2_{\beta_2}} + \frac{\gamma_2^2}{2\sigma^2_{\gamma_2}})}$$

$$= K_{\alpha_2, \beta_2, \gamma_2 | \sigma^2_{z_2}, Z_2, \mathbf{W}, \eta}$$

which is not a closed form expression of any distribution. Hence we need to perform Metropolis-Hastings with a symmetric proposal distribution for $\alpha_2, \beta_2, \gamma_2$. The r-th step for the Metropolis-Hastings algorithm is given by,

1. Generate α^{r*}_{2s} from $N(\alpha^{r-1}_{2s}, \sigma_{met})$, β^{r*}_{s2} from $N(\beta^{r-1}_{2s}, \sigma_{met})$ and γ^{r*}_{2s} from $N(\gamma^{r-1}_{2s}, \sigma_{met})$.

2. Calculate:

$$u = \frac{K_{\alpha^{k*}_{2s}, \beta^{r*}_{2s}, \gamma^{r*}_{2s} | \sigma^{2\,r-1}_{z_2 s}, Z_2, \mathbf{W}, \eta}}{K_{\alpha^{r-1}_{2s}, \beta^{r-1}_{2s}, \gamma^{r-1}_{2s} | \sigma^{2\,r-1}_{z_2 s}, Z_{2s}, \mathbf{W}, \eta}}$$

3. Set $\alpha^{r*}_{2s} = \alpha^{r*}_{2s}$, $\beta^{r}_{2s} = \beta^{r*}_{2s}$, $\gamma^{r*}_{2s} = \gamma^{r*}_{2s}$ with probability u otherwise continue with the value of step $r-1$ with probability $1-u$.

4. Repeat the above steps n_{met} times.

   Hence, we get n_{met} simulated samples from the distribution of $\alpha_2, \beta_2, \gamma_2 \mid \sigma^2_{z_2 s}, Z_{2s}, \mathbf{W}, \eta$. We can use this simulated distribution to update $\alpha_{2s}, \beta_{2s}, \gamma_{2s}$ to $\alpha_{2(s+1)}, \beta_{2(s+1)}, \gamma_{2(s+1)}$.

4.5.3 $(s+1)$th Gibbs Step for updating $\sigma^2_{z_1}$:

The posterior distribution of $\sigma^2_{z_1}$ is given by,

$$\sigma^2_{z_1} \mid Z_1 \sim (\sum_{i=1}^{N} {z_{1i}}^2 + 1) inv\chi^2_{\nu_1+4}$$

$\sigma^2_{z_1(s+1)}$ can be generated from the conditional $\sigma^2_{z_1} \mid Z_{1s}$

4.5.4 $(s+1)$th Gibbs Step for updating $\sigma^2_{z_2}$:

The posterior distribution of $\sigma^2_{z_2}$ is given by,

$$\sigma^2_{z_2} \mid Z_2 \sim (\sum_{i=1}^{N} {z_{2i}}^2 + 1) inv\chi^2_{\nu_1+4}$$

$\sigma^2_{z_2(s+1)}$ can be generated from the conditional $\sigma^2_{z_2} \mid Z_{2s}$

4.5.5 $(s+1)$th Gibbs Step for updating Z_1, Z_2:

The full conditional distribution of z_{1i} is given by:

$$\begin{aligned}
&P(z_{1i} \mid \alpha_1, \beta_1, \gamma_1, {\sigma_{z_1}}^2, \mathbf{W}, \eta) \\
&\propto \left(\prod_{j \neq i} \frac{e^{(\alpha_1 - \beta_1 ||\underset{\sim}{\eta_i} - \underset{\sim}{\eta_j}|| - \gamma_1 |z_{1i} - z_{1j}|)|w_{ij}|}}{1 + e^{\alpha_1 - \beta_1 ||\underset{\sim}{\eta_i} - \underset{\sim}{\eta_j}|| - \gamma_1 |z_{1i} - z_{1j}|}} \right) \times e^{\frac{z_i^2}{\sigma^2_{z_1}}} \\
&= K_{z_{1i} | \alpha_1, \beta_1, \gamma_1, {\sigma_{z_1}}^2, \mathbf{W}, \eta)}
\end{aligned}$$

which is not a closed form expression of any distribution. Hence we need to perform Metropolis-Hastings with a symmetric proposal distribution for z_{1i}. The r-th step for the Metropolis-Hastings algorithm is given by,

1. Generate z^{r*}_{1is} from $N(z^{r-1}_{1is}, \sigma^2_{met})$.

2. Calculate:

$$u = \frac{K_{z^{r*}_{1is} | \alpha_{1(s+1)}, \beta_{1(s+1)}, \gamma_{1(s+1)}, \sigma^2_{z_1(s+1)}, \mathbf{W}, \eta)}}{K_{z^{r-1}_{1is} | \alpha_{1(s+1)}, \beta_{1(s+1)}, \gamma_{1(s+1)}, \sigma^2_{z_1(s+1)}, W, \eta)}}$$

3. Set $z^r_{1is} = z^{r*}_{1is}$ with probability u otherwise continue with the value of step $r-1$ with probability $1-u$.

4. Repeat the above steps n_{met} times.

Hence, we get n_{met} simulated samples from the distribution of,

$$z_{1i} \mid \alpha_{1(s+1)}, \beta_{1(s+1)}, \gamma_{1(s+1)}, \sigma^2_{z_1(s+1)}, \mathbf{W}, \eta$$

We can use this simulated distribution update z_{1is} to $z_{1i(s+1)}$.

We can update z_{2is} similarly.

Chapter 5

Region Wise Variable Selection with Bayesian Group LASSO

5.1 Region-wise Variable Selection

In various spatial-economic analyses, the problem is to select variables that are located spatially. Sometimes the interest lies on estimating the relevance of individual variable over different locations where a Bayesian framework can be really effective provided the fact that we have the idea of the adjacency structure or rather the network structure between the location. Hence, the main idea is to incorporate the adjacency structure between the nodes (locations) so that we can incorporate this fact that relevance of a variablein a certain location is influenced on its status on the adjacent locations. For example, if we can assume that annual snowfall rate is a key factor on deciding the auto insurance premium rates in Michigan, then it would also be relevant or would have some impact on deciding the auto insurance rates in Ohio or Indiana. It is very important to select variables that are relevant to each location and the the problem becomes a bi-level selection where we not only select the variable overall but we also inspect whether it is significant to each location.

Spatial or cross-sectional dependency is a common feature in present econometric applications. To capture spatial dependency, a popular approach is to introduce a spatial weight matrix $\mathbf{W}$ containing the spatial weights over its elements (Giacomini and Granger, 2004). There are various ways to retrieve the spatial weights: from geographic distances, notions of economic distances(Conley, 1999; Pesaran, 2004; Holly et al., 2010), socio-cultural distances (Conley and Topa, 2002; Bhattacharjee and Jensen-Butler, 2005) etc. An alternative and increasingly popular approach is to estimate spatial panel regression models under multi-factor error structures. Factor models are potentially powerful in the sense that they do not require strong and unverifiable assumptions on the nature of spatial dependence.

In a location variable selection problem, it is important to consider the variable selection

procedure to be dependent spatially. Consider the following model for each location $i \in \{1, 2, ..., N\}$:

$$\underset{\sim}{y_i} = \mathbf{X}_i \underset{\sim}{\beta_i} + \underset{\sim}{\epsilon_i} \tag{5.1}$$

where $\underset{\sim}{y_i}$ is a $(R \times 1)$ vector of response variables, $\mathbf{X}_i$ is the $(R \times p)$ design matrix containing p variables, $\underset{\sim}{\beta_i}$ is the $(p \times 1)$ coefficient vector and $\underset{\sim}{\epsilon_i} \sim N(0, \sigma^2 \mathbf{I})$.

Smith and Kohn (1996), Smith and Fahrmeir (2015) have considered the variable selection problem by attaching an indicator vector $\underset{\sim}{\gamma_i} = (\gamma_{i1}, \gamma_{i2}, ..., \gamma_{ip})'$ corresponding to $\underset{\sim}{\beta_i}$ where we set $\beta_{ij} = 0$ if $\gamma_{ij} = 0$ and set $\beta_{ij} \neq 0$ if $\gamma_{ij} = 1$.

The above model can alternatively be expressed as:

$$\underset{\sim}{y_i} = \mathbf{X}_i(\underset{\sim}{\gamma_i}) \underset{\sim}{\beta_i}(\underset{\sim}{\gamma_i}) + \underset{\sim}{\epsilon_i}$$

To undertake the posterior computation, Kohn et al. (2001) have considered a proper conditional prior by setting it proportional to the likelihood:

$$\underset{\sim}{\beta_i}(\underset{\sim}{\gamma_i}) \mid \underset{\sim}{y_i}, \sigma^2, \underset{\sim}{\gamma_i} \sim N(\underset{\sim}{\hat{\beta}_i}(\underset{\sim}{\gamma_i}, R\sigma^2 (\mathbf{X}_i(\underset{\sim}{\gamma_i})' \mathbf{X}_i(\underset{\sim}{\gamma_i}))^{-1})$$

where,

$$\underset{\sim}{\hat{\beta}_i}(\underset{\sim}{\gamma_i}) = (\mathbf{X}_i(\underset{\sim}{\gamma_i})' \mathbf{X}_i(\underset{\sim}{\gamma_i}))^{-1} \mathbf{X}_i(\underset{\sim}{\gamma_i})' \underset{\sim}{y_i}$$

If we assume,

$$P(\sigma^2 \mid \underset{\sim}{\gamma_i}) \propto \frac{1}{\sigma^2}$$

then by Smith and Kohn ((1996), we can show that $P(\underset{\sim}{\gamma_i} \mid \underset{\sim}{y_i}) \propto P(\underset{\sim}{y_i} \mid \underset{\sim}{\gamma_i}) P(\underset{\sim}{\gamma_i})$

We need to set a prior distribution on $\underset{\sim}{\gamma_i}$ to estimate the above model. After we decide on the prior knowledge, we can run the MCMC sampling schemes (Smith and Kohn, 1996) and the Metropilos-Hastings technique to figure out the posterior estimates.

Smith and Fahrmeir (2015) have considered the fact that the variable selection procedure should have a spatial impact, and they addressed this issue by introducing the Ising prior technique where for $\underset{\sim}{\gamma_{(j)}} = (\gamma_{1j}, \gamma_{2j}, ..., \gamma_{Nj})$, they have considered the prior knowledge on $\boldsymbol{\gamma}$ as, $P(\boldsymbol{\gamma}) = \prod_{j=1}^{p} P(\underset{\sim}{\gamma_{(j)}})$, where,

$$P(\underset{\sim}{\gamma_{(j)}}) \propto exp\Big\{ \sum_{i=1}^{N} \alpha_{ij}\gamma_{ij} + \sum_{i \sim k} \theta_{ikj} w_{ik} I(\gamma_{ij} = \gamma_{kj}) \Big\}$$

Here, $I(\cdot)$ is an indicator function, w_{ij} is the pre-specified weight due to adjacency between location i and j. The term $\sum_{i \sim k} \theta_{ikj} w_{ik} I(\gamma_{ij} = \gamma_{kj})$ evaluates the interaction between the effects of the elements $\underset{\sim}{\gamma_{(j)}}$ for all pairwise neighboring sites.

A critical issue of this technique is to specify the external field $\sum_{i=1}^{N} \alpha_{ij}\gamma_{ij}$ where the parameter α_{ij} is fixed apriori. The usual technique is to use a pre-estimate of α_{ij} which depends on the type of the problem, and the posterior estimates are much sensitive over the choice of α_{ij}'s (see Smith and Fahrmeir, 2015).

This paper is focused on proposing an alternative technique of the location-wise variable selection that not only involves the impact of the adjacency structure on the variable selection but also overcomes the ambiguity of pre-specification of the hyper-parameters. The variable selection process is carried out by implementing the Bayesian Group Lasso technique where we put an emphasis on a similar bi-level variable selection approach that incorporates a cross-sectional dependency among the coefficients over the various locations. We use the

spike and slab prior on the group level and within the group level where a group means the model covariates over several locations. Our purpose is to select within group level, while keeping in mind that the relevance of a covariate in a certain location depends on its relevance on the other locations. We introduce a conditional autoregressive structure among the model covariates to incorporates this fact. The median thresholding technique (Xu and Ghosh, 2015) facilitates having exact zero estimates of the non-relevant variables for the corresponding locations since it has a slightly better model selection accuracy as well as a better prediction performance than the traditional LASSO method. The key factor of the technique introduced by Xu and Ghosh (2015)is to use the posterior median estimator that derives that under an orthogonal design and works as a soft thresholding estimator, and the median thresholding is consistent in model selection and has an optimal asymptotic estimation rate.

5.2 Region-wise Variable Selection with Bayesian Group LASSO

Suppose we observe responses $\underset{\sim}{y}_{ir}$ on $i = 1, 2, ..., N$ locations and on $r = 1, 2, ..., R$ independent replications. We setup the following linear regression model as:

$$y_{ir} = \underset{\sim}{X}_{ir}\underset{\sim}{\beta}_i + \epsilon_{ir} \quad i = 1, 2, ..., N \quad r = 1, 2, ..., R. \tag{5.2}$$

where $\underset{\sim}{X}_{ir}$ is a $p \times 1$ vector of predictors for the i^{th} location and for r^{th} replicate. $\underset{\sim}{\beta}_i = (\beta_{i1}, \beta_{i2}, ..., \beta_{ip})'$ is a vector of model coefficients corresponding to the i^{th} location.

We assume the spatial errors are independently and identically distributed ($i.i.d$) over

time and a homoscedastic structure across locations as $E(\underset{\sim}{\epsilon}_r \underset{\sim}{\epsilon}_r^{'}) = \sigma^2 \mathbf{I}_n$.

Now to divide the set of coefficients in to different group, we can rewrite our model (5.2) as,

$$\underset{\sim}{y}_r = \sum_{g=1}^{p} \underset{\sim}{\underset{\sim}{X}}_{gr} \otimes \underset{\sim}{\beta}_g + \underset{\sim}{\epsilon}_r \quad , \quad r = 1, 2, ..., R \tag{5.3}$$

where $\underset{\sim}{y}_r^{'} = (y_{1r}, y_{2r}, ..., y_{Nr})$ is the response vector at replicate r over N locations, $\underset{\sim}{\underset{\sim}{X}}_{gr} = (x_{1gr}, x_{2gr}, ..., x_{ngr})^{'}$ and $\underset{\sim}{\beta}_g = (\beta_{1g}, \beta_{2g}, ..., \beta_{Ng})^{'}$ $\forall g = 1, 2, ..., p$ and $\forall r = 1, 2, ..., R$.

The purpose of this paper is to perform a variable selection where spatial dependence is driven by observed structural interactions. Since our model involves a variable selection over a fixed set of covariates in multiple locations, we propose the group LASSO method that generalizes the LASSO in order to select the grouped variables for accurate prediction of regression. The group LASSO estimator can be obtained by solving the following minimization problem,

$$\min_{\beta} \left(\sum_{r=1}^{R} (||\underset{\sim}{y}_r - \sum_{g=1}^{p} \underset{\sim}{\underset{\sim}{X}}^{*}_{gr} \otimes \underset{\sim}{\beta}_g|| + \lambda_1 ||\underset{\sim}{\beta}||_1 + \lambda_2 \sum_{g=1}^{p} ||\underset{\sim}{\beta}_g||_2) \right) \tag{5.4}$$

The Bayesian formulation provides shrinkage of the coefficients in the group and within the group's level. But the classical group LASSO technique does not provide exact zero estimates for the coefficients that are not relevant. Thus, we introduce sparsity at the group and within the group level by assuming spike and slab prior for the model covariate that brings sparsity in the model coefficients. Johnstone and Silverman (2004) showed that posterior median with a random thresholding estimator provides good estimate along with some desirable properties under spike and slab priors for normal means. We use the posterior median instead of the posterior mean as our posterior estimates of the model coefficients.

5.2.1 Spike and Slab Prior for Model Coefficients

We propose the following Bayesian hierarchical model that we refer to as Bayesian Sparse group LASSO to enable shrinkage both at group level and within a group.

$$\underset{\sim}{y}_r \mid \underset{\sim}{\beta}_1, \underset{\sim}{\beta}_2, ..., \underset{\sim}{\beta}_g, \sigma \quad \sim N(\sum_{g=1}^{p} \underset{\sim}{X}_{gr} \otimes \underset{\sim}{\beta}_g, \sigma^2 \mathbf{I}_N) \tag{5.5}$$

$$\underset{\sim}{\beta}_g \mid \underset{\sim}{\tau}_g, \sigma \quad \sim N(0, \sigma^2 \mathbf{V}_g) \quad g = 1, 2, ..., p \tag{5.6}$$

Here $\mathbf{V}_g^{1/2} = diag\{\tau_{g1}, ..., \tau_{gN}\}$, $\tau_{gj} \geq 0$, $g = 1, 2, ..., p$; $j = 1, 2, ..., N$.

Xu and Ghosh (2015) have introduced a sparse group LASSO modeling where they have represented the model coefficients as a scaled version of a sparse diagonal matrix that helps to select variables within group level along with the group selection.

To introduce sparsity in the model and to select relevant variables at the group and within the group level, we reparametrize the coefficient vectors as,

$$\underset{\sim}{\beta}_g = \mathbf{V}_g^{1/2} \underset{\sim}{b}_g \tag{5.7}$$

Here $\underset{\sim}{b}_g$, when nonzero has a 0 mean and dispersion matrix $\mathbf{I}_N$. The diagonal elements of $\mathbf{V}_g^{1/2}$ control the magnitude of elements of $\underset{\sim}{\beta}_g$. A hierarchical Bayesian modeling using Spike and Slab prior have been introduced by Inswaran and Rao (2005) in which they have considered an inflated probability structure at 0 that helps bringing sparsity in the model. One key advantage of the Spike Slab model is we can show that the prior variance can be dependent on the sample size and hence an appropriate shrinkage level can be achieved and a strong selection consistency can be shown (Narisetty, 2014).

To have a sparse estimate of the model coefficients, we define the following multivariate

spike and slab prior to selecting variables at group level:

$$\underset{\sim}{b}_g \overset{iid}{\sim} (1-\pi_0)N_n(0, I_n) + \pi_0\delta_0(\underset{\sim}{b}_g), \quad g = 1, 2, ..., p \tag{5.8}$$

Note that when $\tau_{gj} = 0$, β_{gj} is dropped out of the model even when $b_{gj} \neq 0$, which means $\underset{\sim}{\tau}_g$ drives a within group level selection for a selected group of $\underset{\sim}{\beta}_g$. Now selecting the elements of $\underset{\sim}{\beta}_g$ means selecting the g^{th} covariate over N different locations. Here, we use the fact that importance of the g^{th} variable on the i^{th} location should depend on the relevance of the g^{th} variable on the adjacent location. Consider a spatial adjacency structure among the N spatial locations that can be represented through a known spatial weight matrix $\mathbf{W} = ((w_{ij}))$, $i = 1, 2, ..., N$ *and* $i = 1, 2, ..., N$. Here, w_{ij} is the weight corresponding to the strength of adjacency between location i and j.

For the prior selection of the within group, we assume a spatial cross-sectional dependence is convoluted within the covariate structure of the model. We would use this fact later on to define the prior structure of the model's coefficients.

To perform a group lasso variable selection in the above model, we need to consider a proper prior for the beta that considers the spatial relationships among the covariates. We therefore assumed a Conditional Autoregressive Prior for the prior on τ as,

$$\tau_{gj} \mid \tau_{gi} : i \neq j \sim (1-\pi_1)N^+(\sum_{i=1, i\neq j}^{N} \frac{w_{ij}}{w_{j+}}\tau_{gi}, \frac{s^2}{w_{i+}}) + \pi\delta_0(\tau_{gj}), \quad g = 1, 2, ..., p; \;\; j = 1, 2, ..., N \tag{5.9}$$

Here $w_{j+} = \sum_{i=1}^{N} w_{ij}$.

where N^+ denotes a folded normal towards the positive side of the real line.

Remarks. *In chapter 4, we have considered a Spatial Error Correction Model where the*

adjacency structure $\boldsymbol{W}$ *is unobserved. The chapter shows an estimation technique of the link probabilities in a two-step procedure where an error correction model is considered to have a pre-estimate of the* $\boldsymbol{W}$ *martix and then the latent metwork model is incorporated where a Bayesian estimate of a connection (positive or negative) or no connection is obtained. The above modeling technique can also be carried out with an unobserved structure of the* $\boldsymbol{W}$ *and then a two-step variable selection technique by estimating* $\boldsymbol{W}$ *and using that* $\boldsymbol{W}$ *as observed in our model to consider it for the variable selection purpose. We will be following a similar technique in the data analysis part of this paper where we will consider the* $\boldsymbol{W}$ *to be pre-estimated.*

Instead of specifying fixed values for hyperparameters, we set,

$$\sigma^2 \sim IG(\alpha, \gamma), \quad \alpha = 0.1, \gamma = 0.1 \tag{5.10}$$

$$\pi_0 \sim Beta(a_1, a_2), \qquad \pi_1 \sim Beta(c_1, c_2) \tag{5.11}$$

$$s^2 \sim IG(1, k) \tag{5.12}$$

5.3 Hellinger Consistency for the Posterior Distribution of β

In this section, we will show that the posterior density of β_{gj} i.e. $(\beta_{gj} \mid rest) = (\tau_{gj} \mid rest) \cdot (b_{gj} \mid rest)$ is Hellinger consistent under a true density is $\mathbb{P}_0$. Suppose the true value of the ij^{th} model coefficient is β_{ij}^0. We will apply the Schwartz theorem to show that

$(\beta_{gj} \mid rest)$ is consistent for the true density under $\boldsymbol{\beta}^0$.

Theorem (Due to 'Schwartz (1965)'). *Let the model* $\mathbb{P} = \Big\{ f(\cdot \mid \beta, \sigma) : \beta \in \mathbb{R}^{N \times p}, \sigma > 0 \Big\}$ *be totally bounded relative to the Hellinger metric* H *and let* $y_{i1}, y_{i2}, ..., y_{iR}$ *be iid* $P_0 = \Big\{ f(\cdot \mid \beta_0, \sigma) : \sigma > 0 \Big\}$ *for some* $P_0 \in \mathbb{P}$. *If* Π *is a Kullback-Leibler prior, i.e., for all* $\delta > 0$

$$\Pi\Big(P \in \mathbb{P} : -P_0 log \frac{dP}{dP_0} < \delta\Big) > 0$$

then the posterior is Hellinger consistent at P_0, *that is,*

$$\Pi\Big(H(P, P_0) > \epsilon \mid y_{i1}, y_{i2}, ..., y_{iR}\Big) \xrightarrow[P_0 a.s]{} 0.$$

An equivalent formulation of totally bounded model can be found in LeCam (1986) where it has been shown involving an unbiased test for testing $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta_0}$ vs $H_1 : \boldsymbol{\beta} \in U^c$ for every neighbourhood U of $\boldsymbol{\beta}_0$. More generally, existance of uniformly consistent test for $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta_0}$ vs $H_1 : \boldsymbol{\beta} \in U^c$ implys Hellinger consistency at $\boldsymbol{\beta_0}$.

Let us assume $\pi(K_\epsilon(\boldsymbol{\beta_0})) > 0 \ \forall \ \epsilon > 0$, where $K_\epsilon(\boldsymbol{\beta_0})$ is the $K - L$ neighborhood of $\boldsymbol{\beta_0}$ denoted by $\{\boldsymbol{\beta} : K(\boldsymbol{\beta_0}, \boldsymbol{\beta}) < \epsilon\}$ and $K(\cdot, \cdot)$ is the $K - L$ divergence.

Lemma 1. *Suppose* U *be the* ϵ *neighbourhood around* $\boldsymbol{\beta_0}$. *If,*

1. $\boldsymbol{\beta_0}$ *is in the* $K - L$ *support of* π.
2. $\int \sqrt{f_{\boldsymbol{\beta}}(y) f_{\boldsymbol{\beta_0}}(y)} dy < \delta \ \forall \ \boldsymbol{\beta} \in U^c$.
3. $\sup\limits_{\boldsymbol{\beta} \in U^c} \left| \int \sqrt{\frac{f_{\boldsymbol{\beta}}(y)}{f_{\boldsymbol{\beta_0}}(y)}} dy - \int \sqrt{f_{\boldsymbol{\beta}}(y) f_{\boldsymbol{\beta_0}}(y)} dy \right| \longrightarrow 0$ *a.s. under the joint distribution of* $y_{i1}, y_{i2}, ..., y_{iR}$ *as* $R \longrightarrow \infty$.

then, there exists a uniformly consistent test for $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta_0}$ *vs* $H_1 : \boldsymbol{\beta} \in U^c$ *(see Van De Geer (1993); Choi and Ramamoorthi (2008)).*

The proof of ***Lemma 1*** *is given in section 5.9.*

Lemma 2. *Suppose* B *is a normal with mean* μ_B *and a standard deviation of* σ_B*, and suppose* T *is a positive normal with mean* μ_T *and a standard deviation of* σ_T*. Then the distribution of* $Z = TB$ *is given by*

$$f(Z) = \frac{e^{-\frac{\rho_1^2+\rho_2^2}{2}}}{\pi\Phi(\rho_2)}\left[\Sigma_0 K_0 + (\rho_1^2+\rho_2^2)\frac{|Z|}{Z!}\Sigma_2 K_1 + (\rho_1^4+\rho_2^4)\frac{|Z|^2}{4!}\Sigma_4 K_2 + ...\right]$$

where,

$K_\gamma(Z) = \frac{1}{2}(\frac{Z}{2})^\gamma \int_0^\infty \frac{e^{-y-\frac{Z^2}{4y}}}{y^{\gamma+1}}dy$

$\rho_1 = \frac{\mu_B}{\sigma B}$ *and* $\rho_1 = \frac{\mu_T}{\sigma_T}$,

$\Sigma_r(\rho_1\rho_2 Z) = 1 + \frac{\rho_1\rho_2 Z}{r+1} + \frac{(\rho_1\rho_2 Z)^2}{(r+2)^{(2)}2!} + ...$

with $(r+k)^{(k)} = (r+k)(r+k-1)...(r+1)$.

Here, $K_\gamma(Z)$ *is called the modified Bessel function of second kind.*

The proof of ***Lemma 2*** *is given in section 5.9.*

Sparsity Assumption. *If we have a sparsity assumption for a large network, as well as on the model, then we can approximate the prior mean of* τ_{gj} *close to* 0*. This would be our key assumption to show the Hellinger consistency. That means we would assume that each node has a very small number of edges with respect to the whole set of nodes, and we set most of the* τ_{gj}*'s at* 0 *ending up with a very small number of significant covariates. In other words, if we set* $\lim_{N\to\infty}\sum_{i=1,i\neq j}^{N}\frac{w_{ij}}{w_{j+}}\tau_{gi} = 0$*, then we have a large* N*, the prior mean for* $b_{gj} = 0$*, and prior mean for* $\tau_{gj} \approx 0$*. It is shown that the prior distribution of* β_{gj} *is* $\frac{K_0(|\beta_{gj}|)}{\pi\Phi(\rho_2)}$*.*

Lemma 3. *Suppose B is a normal with mean μ_B and standard deviation σ_B and suppose T is a positive normal with mean μ_T and standard deviation σ_T. Then the mean of $Z = TB$ is given by*

$$\mu_Z = \left\{ \frac{\phi(\frac{\mu_T}{\sigma_T})}{\Phi(\frac{\mu_T}{\sigma_T})} \frac{\sigma_T}{\mu_B} + \mu_B \mu_T \right\} \frac{\mu_T}{\sigma_T} \Phi(\frac{\mu_T}{\sigma_T}) exp\left\{ -\frac{1}{2} \frac{\mu_T^2}{\sigma_T^2} \right\}$$

*The Proof of **Lemma 3** is given in section 5.9.*

Thus, it is easy to observe that the prior distribution of β_{gj} has a mean of 0 and so a sparsity assumption of τ_{gj} gives the prior mean of $\tau_{gj} \approx 0$, and so we have $\mu_{\beta_{gj}} = 0$. As the prior distribution of beta is a modified Bessel function of second kind, the *Lemma* 2 shows it is symmetric around 0.

Proof (**Theorem**). *According to Lemma 1, prior density of $\beta_{gj} = \tau_{gj} b_{gj}$ is given by*

$$\pi(\beta_{ij}) = l_\beta \delta_0(\beta_{gj}) + (1 - l_\beta) \frac{K_0(|\beta_{ij}|)}{\pi \Phi(\rho_2)} \tag{5.13}$$

where $l_\beta = 1 - (1 - \pi_0)(1 - \pi_1)$ and $\rho_2 = \frac{\sum_{i \neq j} w_{ij} \tau_{gi}}{s^2}$.

Set $\underset{\sim}{\beta}_i^0$ be the true value of $\underset{\sim}{\beta}_i$. According to Schwartz's Theorem, we calculate,

$$\begin{aligned} & -\prod_{i=1}^{N} f_{\underset{\sim}{\beta}_i^0, \sigma} log \frac{\prod_{i=1}^{N} f_{\underset{\sim}{\beta}_i^0, \sigma}}{\prod_{i=1}^{N} f_{\underset{\sim}{\beta}_i, \sigma}} \\ & = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{ -\frac{1}{2} \sum_{i=1}^{N} \frac{y_{ir} - \underset{\sim}{x}_{ir}' \underset{\sim}{\beta}_i^0}{\sigma} \right\}^2 \cdot \frac{1}{2} \sum_{i=1}^{N} \left\{ \left(\frac{y_{ir} - \underset{\sim}{x}_{ir}' \underset{\sim}{\beta}_i}{\sigma} \right)^2 - \left(\frac{y_{ir} - \underset{\sim}{x}_{ir}' \underset{\sim}{\beta}_i^0}{\sigma} \right)^2 \right\} \end{aligned}$$

Suppose there exists an ϵ_1 for which the KL condition does not holds, i.e.,

$$\pi_{\boldsymbol{\beta}}\Bigg\{\boldsymbol{\beta}: e^{-\frac{1}{2}\sum_{t=1}^{R}\sum_{i=1}^{N}\left(\frac{y_{ir}-\underset{\sim}{x}_{ir}^{'}\underset{\sim}{\beta}_i^0}{\sigma}\right)^2}\cdot\left[\sum_{r=1}^{R}\sum_{i=1}^{N}\left(\frac{y_{ir}-\underset{\sim}{x}_{ir}^{'}\underset{\sim}{\beta}_i}{\sigma}\right)^2 - \sum_{r=1}^{R}\sum_{i=1}^{N}\left(\frac{y_{ir}-\underset{\sim}{x}_{ir}^{'}\underset{\sim}{\beta}_i^0}{\sigma}\right)^2\right] < \epsilon_2\Bigg\} = 0$$

$$\Big(Here,\quad \epsilon_2 = \epsilon_1\cdot 2(\sigma)^R(\sqrt{2\pi})^{R/2} > 0\Big)$$

$$\implies \quad \pi_{\boldsymbol{\beta}}\Bigg\{\boldsymbol{\beta}: \sum_{r=1}^{R}\sum_{i=1}^{N}\Big\{\Big(\frac{y_{ir}-\underset{\sim}{x}_{ir}^{'}\underset{\sim}{\beta}_i}{\sigma_0}\Big)^2 - \Big(\frac{y_{ir}-\underset{\sim}{x}_{ir}^{'}\underset{\sim}{\beta}_i^0}{\sigma_0}\Big)^2\Big\} < \epsilon_3\Bigg\} = 0$$

$$\Big(Here,\quad \epsilon_3 = \epsilon_2\cdot exp\left[\frac{1}{2}\sum_{r=1}^{R}\sum_{i=1}^{N}\left(\frac{y_{ir}-\underset{\sim}{x}_{ir}^{'}\underset{\sim}{\beta}_i^0}{\sigma}\right)^2\right]\Big)$$

$$\implies \quad \pi_{\boldsymbol{\beta}}\Bigg\{\boldsymbol{\beta}: 2\sum_{r=1}^{R}\sum_{i=1}^{N}y_{it}\sum_{j=1}^{p}x_{ijr}(\beta_{ij}^0-\beta_{ij}) - \sum_{r=1}^{R}\sum_{i=1}^{N}\sum_{j=1}^{p}x_{ijr}^2({\beta_{ij}^0}^2-{\beta_{ij}}^2) < \epsilon_3\Bigg\} = 0$$

$$\implies \quad \pi_{\boldsymbol{\beta}}\Bigg\{\boldsymbol{\beta}: \sum_{i=1}^{N}\sum_{j=1}^{p}\Big[\beta_{ij}^2\sum_{r=1}^{R}x_{ijr}^2 + \frac{2}{p}(\beta_{ij}^0-\beta_{ij})\sum_{r=1}^{R}y_{ir}x_{ijr}\Big] < \epsilon_4\Bigg\} = 0$$

$$(Where,\quad \epsilon_4 = \epsilon_3 + \sum_{i=1}^{N}\sum_{j=1}^{p}{\beta_{ij}^0}^2\sum_{r=1}^{R}x_{ijr}^2 > 0)$$

$$\implies \quad \pi_{\boldsymbol{\beta}}\Bigg\{\boldsymbol{\beta}: \sum_{i=1}^{N}\sum_{i=1}^{p}(\beta_{ij}-\beta_{ij}^0)a_{ij}) < \epsilon_5\Bigg\} = 0$$

where, $\epsilon_5 = -\frac{p}{2}\epsilon_4 < 0$ and $a_{ij} = \sum_{r=1}^{R}y_{ir}x_{ijr}$. The above inequality inside $\pi_{\underset{\sim}{\beta}_i}\{\underset{\sim}{\beta}_i:\cdot\}$ holds since $\sum_{i=1}^{N}\sum_{j=1}^{p}\beta_{ij}^2\sum_{r=1}^{R}x_{ijr}^2 > 0$.

Hence,

$$\pi_{\boldsymbol{\beta}}\Bigg\{\boldsymbol{\beta}: \sum_{i=1}^{N}\sum_{j=1}^{p}\beta_{ij}a_{ij} > \epsilon_5 + \sum_{i=1}^{N}\sum_{j=1}^{p}\beta_{ij}^0 a_{ij}\Bigg\} = 0$$

$$\implies \prod_{i=1}^{N}\prod_{j=1}^{p}\pi_{\beta_{ij}}\left\{\beta_{ij}:\beta_{ij}a_{ij}>\epsilon_6+\beta_{ij}^0a_{ij}\right\}=0 \tag{5.14}$$

where $\epsilon_6=\epsilon_5/Np<0$.

Equation (5.14) is true since

$$\left\{\beta_{ij}a_{ij}>\epsilon_6+\beta_{ij}^0a_{ij},\forall i\geq 1,j\geq 1\right\}\implies\left\{\sum_{i=1}^{N}\sum_{j=1}^{p}\beta_{ij}a_{ij}>\epsilon_5+\sum_{i=1}^{N}\sum_{j=1}^{p}\beta_{ij}^0a_{ij}\right\}$$

Equation (5.14) means that $\forall\ i\in\{1,2,...,N\}$ *and* $j\in\{1,2,...,p\}$ *and for* $\delta>0$ *there exists an* $a_{ij}:a_{ij}=\delta$ *(OR,* $a_{ij}=-\delta$*) where* $\frac{\epsilon_6}{\delta}=-M$ *(OR,* $\frac{\epsilon_6}{-\delta}=M$*) where* M *is a large positive integer with* $|M|>\beta_{ij}^0$ *(we can assume) such that,*

$$\pi_{\beta_{ij}}\left\{\beta_{ij}:\beta_{ij}>-M+\beta_{ij}^0\right\}=0 \tag{5.15}$$
$$\left(OR,\quad\pi_{\beta_{ij}}\left\{\beta_{ij}:\beta_{ij}<M+\beta_{ij}^0\right\}=0\right)$$

Now (5.15) is a contradiction since $\beta_{ij}\sim l_\beta\delta_0(\beta_{ij})+(1-l_\beta)\frac{K_0(|\beta_{ij}|)}{\pi\Phi(\rho_2)}$, $0<l_\beta<1$ *and Lemma 2 shows the density of* β_{ij} *is symmetric with respect to zero and* $l_\beta>0$.

5.4 Posterior Distributions and Gibbs Sampling for Group LASSO

Let us denote $\mathbf{X}_r=(\underset{\sim}{X}_{1r},\underset{\sim}{X}_{2r},...,\underset{\sim}{X}_{pr})$.

The joint posterior of $\mathbf{b}=\{\underset{\sim}{b}_i:i=1,2,...,p\}$, $\boldsymbol{\tau}^2=\{\tau_{ij}^2:i=1,2,...,p;j=1,2,...,N\}$,

σ^2, π_0, π_1, s^2 conditional on the observed data is:

$$
\begin{aligned}
&P(\mathbf{b}, \boldsymbol{\tau}^2, \sigma^2, \pi_0, \pi_1, s^2 \mid \underset{\sim}{y}_r, \mathbf{X}_r; r = 1, 2, ..., R) \\
&\propto (\sigma^2)^{-NR/2} exp\Bigg\{ -\frac{1}{2\sigma^2} \sum_{r=1}^{R} ||\underset{\sim}{y}_t - \sum_{i=1}^{p} \underset{\sim}{X}_{gr} \otimes \mathbf{V}_g^{1/2} \underset{\sim}{b}_g||_2^2 \Bigg\} \\
&\times \prod_{g=1}^{p} \Bigg[(1-\pi_0)(2\pi)^{-N/2} exp\Bigg\{ -\frac{1}{2} \underset{\sim}{b}_g' \underset{\sim}{b}_g \Bigg\} I(\underset{\sim}{b}_g \neq \underset{\sim}{0}) + \pi_0 \delta_0(\underset{\sim}{b}_g) \Bigg] \\
&\times \prod_{g=1}^{p} \prod_{j=1}^{N} \Bigg[(1-\pi_1) 2(2^2)^{-1/2} exp\Bigg\{ -\frac{(\tau_{gj} - \frac{w_{ij}}{w_{i+}} \tau_{gi})^2}{2\frac{s^2}{w_{i+}}} \Bigg\} I(\tau_{gj} > 0) + \pi_0 \delta_0(\tau_{gj}) \Bigg] \\
&\times (\sigma^2)^{-\alpha - 1} exp - \frac{\gamma}{2} \\
&\times \pi_0^{a_1 - 1} (1-\pi_0)^{a_2 - 1} \\
&\times \pi_1^{c_1 - 1} (1-\pi_0)^{c_2 - 1} \\
&\times k (s^2)^{-2} exp\Bigg\{ -\frac{k}{s^2} \Bigg\}
\end{aligned}
$$

5.4.1 Gibbs Sampler

- The posterior distribution of $\underset{\sim}{b}_g$ conditional on everything else is given by:

$$\underset{\sim}{b}_g \mid rest \sim l_g \delta_0(\underset{\sim}{b}_g) + (1 - l_g) N(\underset{\sim}{\mu}_g, \boldsymbol{\Sigma}_g) \tag{5.17}$$

where l_g is the posterior probability of $b_g$being equal to $\underset{\sim}{0}$ given the other parameters, i.e.

$$
\begin{aligned}
l_g \quad &= P(\underset{\sim}{b}_g = 0 \mid rest) \\
&= \frac{\pi_0}{\pi_0 + (1-\pi_0)|\boldsymbol{\Sigma}_g|^{1/2} exp\{ \frac{1}{2\sigma^4} ||\boldsymbol{\Sigma}_g^{1/2} (\sum_{r=1}^{R} \mathbf{V}_g^{1/2} \underset{\sim}{X}_{tg}' (\underset{\sim}{y}_r - \mathbf{X}_{r(g)} \otimes \mathbf{V}_{(g)}^{1/2} \mathbf{b}_{(g)}))||_2^2 \}}
\end{aligned}
$$

where, $\mathbf{X}_{r(g)} = (\underset{\sim}{X}r1, ..., \underset{\sim}{X}_{r(g-1)}, \underset{\sim}{X}_{r(g+1)}, ..., \underset{\sim}{X}rp)$,

$\mathbf{b}_{(g)} = (\underset{\sim}{b}_1^{'}, ..., \underset{\sim}{b}_{g-1}^{'}, \underset{\sim}{b}_{g+1}^{'}, ..., \underset{\sim}{b}_p^{'})^{'}$.

Similarly $\mathbf{V}_{(g)} = diag(\mathbf{V}_1, ..., \mathbf{V}_{g-1}, \mathbf{V}_{g+1}, ..., \mathbf{V}_p)$ matrix after deleting the g^{th} row and g^{th} column. Also,

$$\underset{\sim}{\mu}_g = \frac{1}{\sigma^2}\mathbf{\Sigma}_g \sum_{r=1}^{R} \left\{ \mathbf{V}_g^{1/2} \otimes \underset{\sim}{X}_{rg}^{'}(\underset{\sim}{y}_r - \mathbf{X}_{r(g)} \otimes \mathbf{V}_{(g)}^{1/2}\mathbf{b}_{(g)}) \right\}$$

$$\mathbf{\Sigma}_g = \left(\mathbf{I}_N + \frac{1}{\sigma^2} \sum_{r=1}^{R} \left\{ \mathbf{V}_g^{1/2} \otimes \underset{\sim}{X}_{rg}^{'} \underset{\sim}{X}_{rg} \otimes \mathbf{V}_g^{1/2} \right\} \right)^{-1}$$

- The conditional posterior of τ_{gj} is given by,

$$\tau_{gj} \mid rest \sim q_{gj}\delta_0(\tau_{gj}) + (1 - q_{gj})N^{+}(u_{gj}, v_{gj}^2), \quad g = 1, 2, ..., p; \;\; j = 1, 2, ..., N \quad (5.15)$$

where,

$$u_{gj} = \frac{v_{gj}^2}{\sigma^2} \sum_{r=1}^{R} (\underset{\sim}{y}_r - \mathbf{X}_{r(gj)} \otimes \mathbf{V}_{(gj)}^{1/2} \underset{\sim}{b}_{(gj)}) x_{rgj} b_{gj} + \frac{v_{gj}^2 w_{i+}}{s^2} \sum_{i \neq j} \frac{w_{ij}}{w_{i+}} \tau_{gi}$$

$$v_{gj}^2 = (\frac{w_{i+}}{s^2} + \frac{b_{gj}^2}{\sigma^2} \sum_{r=1}^{R} x_{rgj}^2)^{-1}$$

$$q_{gj} = \frac{\pi_1}{\pi_1 + 2(1 - \pi_1) \frac{v_{gj}\sqrt{w_{i+}}}{s} exp\{ \frac{1}{2} \frac{\mu_{gj}^2}{v_{gj}^2} - \frac{w_{i+}}{2s^2} (\sum_{i \neq j} \frac{w_{ij}}{w_{i+}} \tau_{gi})^2 \} \Phi(\frac{u_{gj}}{v_{gj}})}$$

Here we define $\mathbf{X}_{r(gj)}, \mathbf{V}_{(gj)}, \underset{\sim}{b}_{(gj)}$ similarly by removing the corresponding gj^{th} element.

- $\sigma^2 \mid rest \sim IG(\frac{NR}{2} + \alpha, \frac{1}{2} \sum_{r=1}^{R} ||\underset{\sim}{y}_r - \mathbf{X}_r \otimes \boldsymbol{\beta}||_2^2 + \gamma)$

Here $\mathbf{X}_r = (\underset{\sim}{X}r1, , ..., \underset{\sim}{X}rp)$, $\boldsymbol{\beta} = (\underset{\sim}{\beta}_1^{'} ..., \underset{\sim}{\beta}_p^{'})^{'}$.

- $\pi_0 \mid rest \sim beta(\#(b_g = 0) + a_1, \#(b_g \neq 0) + a_2)$
- $\pi_1 \mid rest \sim beta(\#(\tau_{gj} = 0) + c_1, \#(\tau_{gj} \neq 0) + c_2)$
- $s^2 \mid rest \sim IG(1 + \frac{1}{2}\#(\tau_{gj} = 0), t + \frac{1}{2}\sum_{g=1}^{p}\sum_{j=1}^{N}\left[\tau_{gj} - \sum_{i \neq j}\frac{w_{ij}}{w_{i+}}\tau_{gi}\right]^2)$.

We consider our posterior values of the model coefficients over the Gibbs sampler to be $\hat{\beta}_{gj} = (\underset{\sim}{gj} \mid rest) \cdot (\tau_{gj} \mid rest)$. To have our posterior estimate of the β_{gj}, we use the same approach followed by Xu and Ghosh (2015) who have used the posterior median instead of the posterior mean. They have shown in a paper that the posterior median works as a random thresholding estimator that satisfies the oracle property with a faster convergence than the general group LASSO estimator under an orthogonal design.

5.5 Variable Selection for Temporal Data

Variable selection for spatially dependent data can be carried out along the method we discussed in the last few sections. But we have ignored the fact that the temporal dependence might also effect in the sense of having dependency over the responses that are closer with respect to time. Assume we have response vector for N locations $\underset{\sim}{y_t}$ over time points $t = 1, 2, ..., T$. Consider the spatial-temporal regression model as,

$$\underset{\sim}{y_t} = \mathbf{\Phi}\underset{\sim}{y_{t-1}} + \sum_{g=1}^{p}\underset{\sim}{X}_{gt} \otimes \underset{\sim}{\beta}_g + \underset{\sim}{\epsilon_t} \tag{5.13}$$

where,

$$\mathbf{\Phi} = diag(\phi_1, \phi_2, ..., \phi_N)$$

The model we considered under equation (5.13) is nothing but the AR(1) process. Stationarity of the AR(1) process requires the assumption: $|\phi_i| \leq 1$, $\forall 1 \leq i \leq N$. The reason that we have chosen AR(1) over higher-order autoregressive processes is that we want to avoid the abstract restrictions on the autoregressive model parameters due to the stationarity of the process. AR(1) process allows the temporal dependence to decrease gradually as the time lag increases.

We consider the autoregressive component parameters ϕ_i to be independently uniform between -1 and 1, i.e.,

$$\pi(\phi_i) = \prod_{i=1}^{N} I(-1 \leq \phi_i \leq 1)$$

5.5.1 Posterior Distribution of ϕ_i

We can write the posterior probability of $\phi_i; i = 1(1)N \mid rest$ as,

$$\begin{aligned}
&P(\phi_i; i = 1(1)N \mid rest) \\
&\propto exp\Big\{ -\frac{1}{2\sigma^2} \sum_{t=2}^{T} ||\underset{\sim}{y}_t - \mathbf{\Phi}\underset{\sim}{y}_{t-1} - \sum_{g=1}^{p} \underset{\sim}{X}_{gt} \otimes \mathbf{V}_g^{1/2} \underset{\sim}{b}_g||^2 \Big\} \prod_{i=1}^{N} U(-1,1) \\
&\propto exp\Big\{ -\frac{1}{2\sigma^2} \sum_{t=2}^{T} \big[(\underset{\sim}{y}_t - \sum_{g=1}^{p} \underset{\sim}{X}_{gt} \otimes \mathbf{V}_g^{1/2} \underset{\sim}{b}_g)' (\underset{\sim}{y}_t - \sum_{g=1}^{p} \underset{\sim}{X}_{gt} \otimes \mathbf{V}_g^{1/2} \underset{\sim}{b}_g) \\
&-2 \sum_{i=1}^{N} \phi_i \underset{\sim}{m}_{it} y_{\overline{t-1}i} + \sum_{i=1}^{N} \phi_i^2 y^2_{\overline{t-1}i} \big] \Big\}
\end{aligned}$$

Here, $\underset{\sim}{m}_{it}$ is the i^{th} row of $(\underset{\sim}{y}_t - \sum_{g=1}^{p} \underset{\sim}{X}_{gt} \otimes \mathbf{V}_g^{1/2} \underset{\sim}{b}_g)$.

Hence,

$$
\begin{aligned}
&P(\phi_i; i = 1(1)N \mid rest) \\
&\propto exp\Big\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{N} \big[-2\phi_i \sum_{t=2}^{T} \underset{\sim}{m}_{it} \underset{\sim}{y}_{\overline{t-1}i} + \phi_i^2 \sum_{t=2}^{T} \underset{\sim}{y}^2_{\overline{t-1}i} \big] \Big\} \\
&\propto exp\Big\{ \frac{\sum_{t=2}^{T} \underset{\sim}{y}^2_{\overline{t-1}i}}{\sigma^2} \sum_{i=1}^{N} \big(\phi_i - \frac{\sum_{t=2}^{T} \underset{\sim}{m}_{it} \underset{\sim}{y}_{\overline{t-1}i}}{\sum_{t=2}^{T} \underset{\sim}{y}^2_{\overline{t-1}i}}\big)^2 \Big\}
\end{aligned}
$$

Hence, $\phi_i \mid rest \sim N\Big(\frac{\sum_{t=2}^{T} \underset{\sim}{m}_{it} \underset{\sim}{y}_{\overline{t-1}i}}{\sum_{t=2}^{T} \underset{\sim}{y}^2_{\overline{t-1}i}}, \frac{\sigma^2}{\sum_{t=2}^{T} \underset{\sim}{y}^2_{\overline{t-1}i}}\Big) \quad \forall 1 \leq i \leq N$

5.5.2 Gibbs Sampler

Gibbs' sampling steps for the spatio-temporal model would be similar to the situation for the spatial modeling. We would follow the same update procedure along with the update of autoregressive component matrix $\mathbf{\Phi}$. We would replace $\underset{\sim}{y}_t$ with $\underset{\sim}{y}_t - \mathbf{\Phi}\underset{\sim}{y}_{t-1}$ for the Gibbs Sampler steps in section (5.4.1).

5.6 Simulation Study

5.6.1 A Sample Simulation with Prefixed β

In this setting, we preselect some values for β_{ij}s and we compare the BGL-SS-CAR with the simple BGL-SS and the variable selection with ISING prior based on the RMSE and the TPR/FPR values. We set $p = 5$, $T = 10$, and $N = 7$. The comparison is done based on only one simulation. We will increase the number of simulations in the subsequent section to have a better view of the prediction error measurement.

It is evident from table 5.1 and table 5.2 that when a CAR structure is being considered

Table 5.1: RMSE, TPR and FPR comparison for BGL-SS, Ising and BGL-SS-CAR model

Methods	BGL-SS	ISING	BGL-SS-CAR
RMSE	0.868	1.12	1.06
TPR	1	0.83	0.81
FPR	0.368	0.26	0.21

Table 5.2: BGL-SS and BGL-SS-CAR estimates for prefixed β's

Methods	True	BGL-SS	ISING	BGL-SS-CAR
β_{11}	0	0.15	0.30	0.38
β_{21}	2.5	1.63	1.11	1.92
β_{31}	-2.25	-0.70	-1.06	-0.84
β_{41}	0	0.19	0	0
β_{51}	3	2.30	1.33	0.81
β_{61}	0	0.33	0.10	0.36
β_{71}	-1	-1.57	0.36	0
β_{12}	0	0.17	0	0
β_{22}	0	0	0	0
β_{32}	0	0	0	0
β_{42}	0	0	0	0
β_{52}	0	0	0	0
β_{62}	0	0.99	0.37	0
β_{72}	0	0	0	0
β_{13}	2	0.16	1.28	0.34
β_{23}	2	2.89	2.23	1.50
β_{33}	-3	-0.96	-2.58	-2.43
β_{43}	3	1.31	1.78	1.01
β_{53}	1.5	0	0.96	0.71
β_{63}	1	0.16	0	0
β_{73}	-2	-1.79	0.88	0
β_{14}	0	0	0	0
β_{24}	0	0	0	0
β_{34}	0	0	0	0
β_{44}	0	0	0	0
β_{54}	0	0	0	0
β_{64}	0	0	0	0
β_{74}	0	0	0	0
β_{15}	0	0.19	0.64	0.38
β_{25}	2	1.05	0.88	0.25
β_{35}	-1	-2.12	-1.92	-1.35
β_{45}	-3	-2.08	1.68	0.65
β_{55}	0	0	0.47	0.23
β_{65}	2	0	0	1.82
β_{75}	-1.5	-1.75	0	-2.65

within the model coefficients, it is clearly out-performing the regular Sparse Group LASSO and ISING modeling technique both in terms of RMSE as well as in terms of FPR or TPR.

5.6.2 Scenario 1: $N = 7$, $p = 5$ and $T = 10$

In the first scenario, we take a simulated data over 7 spatial location with a pre-specified adjacency structure ($\mathbf{W}$). We take 5 variables for each of the locations, and we take the data over 10 time points. We set σ in to three specified values: $\sigma = 0.5$, $\sigma = 1$ and $\sigma = 3$. We ran the Gibbs iterations for 10,000 times and we took the burn-in period to be 8,000. We replicated the simulation 20 times to have a better measure of errors.

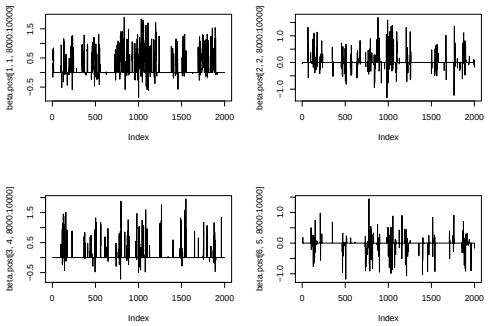


Figure 5.1: Gibbs iterations of β's for the first scenario under BGL-SS-CAR when $\sigma = 0.5$

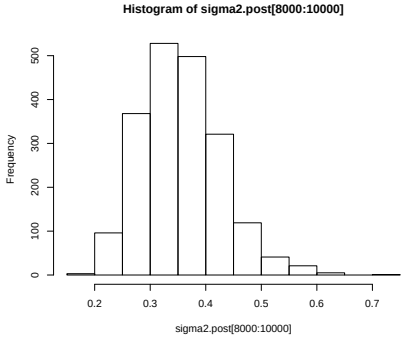


Figure 5.2: Posterior Distribution of σ^2 for the first scenario underBGL-SS-CAR when the true σ^2 is 0.25

5.6.3 Scenario 2: $N = 14$, $p = 15$ and $T = 50$

In the second scenario, we take a simulated data over 14 spatial locations with a pre-specified adjacency structure (**W**). We take 10 variables for each of the locations, and we take the data over 50 time points. We set σ to three specified values:$\sigma = 0.5$, $\sigma = 1$ and $\sigma = 3$. We ran the Gibbs iterations 10,000 times, and we took the burning period to be 8,000. We replicate the simulation 20 times.

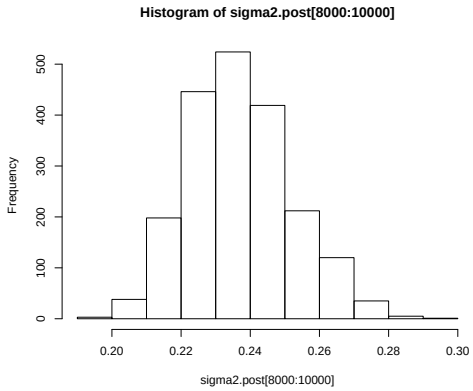


Figure 5.3: Posterior Distribution of σ^2 for the second scenario underBGL-SS-CAR when the true σ^2 is 0.25

In table 5.3, we are comparing simple Bayesian Sparse Group Lasso and ISING model vs the Bayesian Sparse Group Lasso with a CAR structure. The comparison is being done using the False Positive Rates and the True Positive Rates as the two methods based on the two scenarios we considered before.

It can be observed from the tables above is Bayesian sparse group LASSO technique is mostly outperforming the simple sparse group LASSO as well as the ISING model in terms of the RMSE and TPR. Which means when a spatial data is considered and when it is known or expected for the variables to have a dependency structure convoluted in the joint distribution of the covariates, we can expect that a CAR structure an catch the relevant variable more

Table 5.3: Table for RMSE and True / False Positive Rates

	Scenario 1			Scenario 2		
Methods	BGL-SS	ISING	BGL-SS-CAR	BGL-SS	ISING	BGL-SS-CAR
$\sigma = 0.5$						
RMSE	1.02	0.92	0.80	0.42	0.31	0.28
TPR	0.56	0.70	0.62	0.71	0.78	0.77
FPR	0.08	0.06	0.01	0.13	0.15	0.04
$\sigma = 1$						
RMSE	0.70	0.72	0.60	0.71	0.76	0.51
TPR	0.61	0.66	0.71	0.68	0.68	0.72
FPR	0.20	0.10	0.03	0.23	0.07	0.02
$\sigma = 2$						
RMSE	1.03	0.75	0.78	0.95	1.08	0.92
TPR	0.69	0.61	0.73	0.72	0.82	0.88
FPR	0.18	0.15	0.03	0.31	0.10	0.07

efficiently than the other competative methods and is better in terms of lowering the model errors. An advantage of using BGL-SS-CAR over ISING model in terms of computations is that unlike the ISING model it is free from the ambiguity of prespecifying values for some hyperparameter. Also, a bi-level shrinkage brings more control on the sparsity of the model through the two acting variabilities, one between group levels and another within group levels.

It is also very interesting to observe that incorporating the CAR structure in the prior setup of the Bayesian sparse group LASSO technique results in a very efficient variable selection in terms of the False positive rates. We can see from table 5.3 that FPR is close to zero in all the simulation scenarios and also very low compared to the other two competative methods. This means bringing CAR structure in the modeling scenario for a spatially related data allows the model to be very efficient in identifying the covariates which are not relevant for a given location.

5.7 Data Analysis

In this section, we consider compustat data over the U.S. auto industry market. The data consist of 20 U.S. auto manufacturers and the suppliers, including three 3 U.S. auto manufacturing giants GM, Ford, and Chrysler. The data span from 1960 to 1987 and include data for nine variables that consist of some key factors of the manufacturing industry like actual costs, cost of goods sold, total sales figures, revenue total etc. In chapter 4, we have considered this data to determine the latent network structure within the U.S. auto manufacturing industry.

The method discussed in chapter 4 is dedicated to obtaining the probability of a connection or no connections between the companies, i.e., existence of a connection between company i and company j means $w_{ij} = 1$ and $w_{ij} = 0$ stands for no connection between company i and j.

In our application, we will consider the estimated adjacency matrix from chapter 4 to be the observed $\mathbf{W}$ and the purpose is to run a variable selection among the available model covariates. A key thing to note is that in chapter 4, we have used revenue total as the response and sale as the covariate since those two are theoretically the key factors for finding the actual relationships among the companies. In our problem here, we consider 8 covariates, and we ran a variable selection to see which variables are most important.

The data might have the issue of non-stationarity over time since we are not considering a time parameter to handle the dependence over time. Instead, we use the first difference of the response and the covariate as out actual model response and the covariates, i.e.,

$$\Delta ln(y_{it}) = \mathbf{\Phi}\Delta ln(\underset{\sim}{y}_{t-1}) + \Delta ln(\underset{\sim}{\underset{\sim}{X}}_{it})\underset{\sim}{\beta}_i^* + \epsilon_{it} \;\; i = 1, 2, ..., N \;\; t = 2, 3, ..., T.$$

where $\Delta ln(y_{it}) = ln(y_{it}) - ln(y_{it-1})$ and $\Delta ln(x_{itj}) = ln(x_{itj}) - ln(x_{it-1j})$.

Table 5.4: Coefficient estimates through BGL-SS-CAR for Auto Industry Data

Ticker	act	at	cogs	gp	lct	lt	ppegt	sale
$AL.1$	0 (0.03)	0 (0.12)	0.28 (0.13)	0.08 (0.07)	0 (0.03)	0 (0.10)	0 (0.11)	0.62 (0.26)
HON	0 (0.11)	0 (0.05)	0.16 (0.06)	0.04 (0.03)	0 (0.03)	0 (0.11)	0 (0.08)	0.81 (0.38)
ARV	0 (0.03)	0 (0.4)	0.50 (0.44)	0.12 (0.10)	0 (0.03)	0 (0.02)	0 (0.01)	0.37 (0.22)
$C.3$	0 (0.11)	0 (0.04)	0.40 (0.33)	0.03 (0.03)	0 (0.02)	0 (0.01)	0 (0.09)	0.51 (0.32)
CTB	0 (0.01)	0 (0.06)	0.61 (0.34)	0.08 (0.12)	0 (0.05)	0 (0.06)	0 (0.05)	0.28 (0.13)
DAN	0 (0.09)	0 (0.09)	0.41 (0.19)	0.12 (0.11)	0 (0.04)	0 (0.08)	0 (0.03)	0.43 (0.26)
DE	0 (0.05)	0 (0.04)	0.27 (0.15)	0.06 (0.04)	0 (0.10)	0 (0.07)	0 (0.05)	0.66 (0.51)
ETN	0 (0.03)	0 (0.05)	0.26 (0.19)	0.11 (0.09)	0 (0.05)	0 (0.04)	0 (0.11)	0.63 (0.46)
F	0 (0.18)	0 (0.07)	0.41 (0.28)	0.08 (0.13)	0 (0.06)	0 (0.11)	0 (0.04)	0.41 (0.20)
GE	0 (0.11)	0 (0.09)	0.30 (0.23)	0.14 (0.9)	0 (0.10)	0 (0.05)	0 (0.04)	0.57 (0.56)
GM	0 (0.11)	0 (0.09)	0.35 (0.17)	0.08 (0.06)	0 (0.08)	0 (0.12)	0 (0.08)	0.55 (0.07)
$SPXC$	0 (0.07)	0 (0.13)	0.33 (0.17)	0.09 (0.11)	0 (0.05)	0 (0.08)	0 (0.09)	0.55 (0.54)
GR	0 (0.13)	0 (0.10)	0.20 (0.14)	0.07 (0.05)	0 (0.06)	0 (0.09)	0 (0.03)	0.73 (0.47)
GT	0 (0.10)	0 (0.06)	0.19 (0.11)	0.07 (0.05)	0 (0.02)	0 (0.10)	0 (0.05)	0.72 (0.41)
JCL	0 (0.07)	0 (0.05)	0.33 (0.35)	0.14 (0.12)	0 (0.13)	0 (0.08)	0 (0.12)	0.53 (0.45)
$ANV.1$	0 (0.07)	0 (0.06)	0.36 (0.29)	0.21 (0.09)	0 (0.11)	0 (0.08)	0 (0.12)	0.37 (0.21)
OC	0 (0.11)	0 (0.06)	0.25 (0.17)	0.10 (0.05)	0 (0.01)	0 (0.04)	0 (0.16)	0.64 (0.49)
PPG	0 (0.05)	0 (0.07)	0.32 (0.25)	0.13 (0.09)	0 (0.08)	0 (0.13)	0 (0.10)	0.52 (0.45)
AOS	0 (0.06)	0 (0.09)	0.47 (0.23)	0.07 (0.05)	0 (0.10)	0 (0.07)	0 (0.10)	0.43 (0.45)
UTX	0 (0.05)	0 (0.09)	0.31 (0.14)	0.12 (0.10)	0 (0.03)	0 (0.12)	0 (0.04)	0.56 (0.40)

We consider modeling the data using $'revt'$ (Revenue Total) as our response and use the other 8 variables as our model covariates to perform a variable selection. We consider the spatio-temporal modeling (BGL-SS-CAR) by considering an AR(1) process over time. Table 5.4 shows the variable selection along with the posterior median estimates of the coefficients. The table corresponds to the BGL-SS-CAR modeling scenario. The 'Ticker' symbol shows the company tickers in the US stock market. Our model includes 8 covariates are 'act', 'at', 'cogs' etc. The three U.S. auto manufacturing giants are given by $C.3 = Chrysler$, $F = Ford$, $GM = General\ Motors$. Values in the brackets showing the standard errors of the estimates. It is clear to see from the table above according to our Bayesian Group LASSO model that $'cogs'$= Cost of Goods Sold, $'gp'$ = Gross Profit Loss Property, Plant and Equipment - Total (Gross) and $'Sale'$ = Sales total is coming out to be significant variables.

which are heuristically and theoretically make sense and go consistently with the selected covariates in chapter 4.

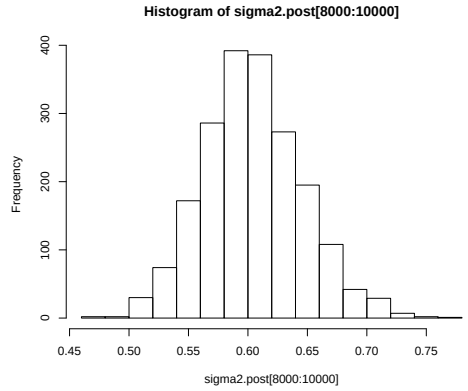


Figure 5.4: Posterior Distribution of σ^2 for the Data

5.8 Discussion

The topic presented in this chapter uses a variable selection technique that takes information from the spatially located covariates as well as the spatial adjacency structure among the nodes that facilitate in selecting the covariates over several spatial locations.

Since this paper uses the spatial adjacency structure among the nodes, it is important to have reliable information on the adjacency structure among the nodes. Since the data is observed in a spatio-temporal fashion, it is important to undergo the test if there is any spatial or temporal non-stationarity. In practice, we take the first difference of the responses to be our initial data to remove the non-stationarity.

Assumption of a CAR structure among the covariates is an important assumption in this paper. This not only facilitates the variable selection technique through a betterment of the RMSE or the TPR but it also provides a profound heuristic and theoretical validation since

it is must be expected that the variable selection in a spatial situation must depend on its neighboring spatial locations.

Since the posterior mean does not provide an exact 0 estimate for the non-relevant covariates, Geweke (1994), Kuo and Mallick (1998), and George and McCullough (1997) suggested the highest posterior probability Model via Gibbs sampling calculates the highest posterior probability around 0 and rejects those variables that have a very significant posterior probability around 0. FDR-based variable selection has been proposed to select variables if marginal inclusion probability is larger than some pre-controlled threshold. The posterior estimation is distinctive in the sense that it directly gives the zero or non-zero estimates without going to a second-step estimation.

5.9 Proof of the Lemmas

5.9.1 Proof of Lemma 1

Let us consider, w.l.g, $y \sim N(x\beta, \sigma^2)$. Let us assume $\beta \in (\beta_0 - \epsilon, \beta_0 + \epsilon)^c$

Now we have,

$$\int \sqrt{f_\beta(y) f_{\beta_0}(y)} dy \qquad exp\Big\{ -\frac{1}{8} x^2 (\beta - \beta^0)^2 \Big\}$$
$$< exp\Big\{ -\frac{1}{8} x^2 \epsilon^2 \Big\} = \delta$$

For the second part of the proof, let us assume $y \in (-M, M)$ where M is a large positive integer so that we can truncate the distribution in $(-M, M)$. Let us take R replicates of y

as $y_{i1}, y_{i2}, ..., y_{iR}$. So, we have,

$$\int \sqrt{\frac{f_\beta(y_{i1}, y_{i2}, ..., y_{iR})}{f_{\beta_0}(y_{i1}, y_{i2}, ..., y_{iR})}} dy \quad \approx \Big(\frac{\sigma^2}{x(\beta - \beta^0)}\Big)^R \Big(e^M - e^{-M}\Big)^R e^{-\frac{R}{2}x^2(\beta^2 - \beta^{0^2})}$$

Hence $\sup_{\beta \in U^c} \left| \int \sqrt{\frac{f_\beta(y_{i1}, y_{i2}, ..., y_{iR})}{f_{\beta_0}(y_{i1}, y_{i2}, ..., y_{iR})}} dy \right| = \Big(\frac{\sigma^2}{x\epsilon}\Big)^R \Big(e^M - e^{-M}\Big)^R e^{-\frac{R}{2}x^2(\beta^{0^2} + 2\epsilon\beta_0)}$ for fixed R.

So, for each ϵ, if we assume $M \approx \frac{(x\epsilon)\frac{1}{\sigma}}{2} exp\Big\{\frac{x^2}{2\sigma}(\beta^{0^2} + 2x\beta^0)\Big\}$, then we can show,

$$\sup_{\beta \in U^c} \left| \int \sqrt{\frac{f_\beta(y_{i1}, y_{i2}, ..., y_{iR})}{f_{\beta_0}(y_{i1}, y_{i2}, ..., y_{iR})}} dy \right| \longrightarrow 0 \quad as \quad R \longrightarrow \infty$$

Again,

$$\int \sqrt{f_\beta(y_{i1}, y_{i2}, ..., y_{iR}) f_{\beta_0}(y_{i1}, y_{i2}, ..., y_{iR})} dy \quad = exp\Big\{ - \frac{R}{8}x^2(\beta - \beta^0)^2 \Big\}$$

So, $\sup_{\beta \in U^c} \left| \int \sqrt{f_\beta(y_{i1}, y_{i2}, ..., y_{iR}) f_{\beta_0}(y_{i1}, y_{i2}, ..., y_{iR})} dy = exp\Big\{ - \frac{R}{8}x^2\epsilon^2 \Big\}$ for fixed R.

Now from triangle inequality,

$$\begin{aligned}
&\sup_{\beta \in U^c} \left| \int \sqrt{\frac{f_\beta(y_{i1}, y_{i2}, ..., y_{iR})}{f_{\beta_0}(y_{i1}, y_{i2}, ..., y_{iR})}} dy - \int \sqrt{f_\beta(y_{i1}, y_{i2}, ..., y_{iR}) f_{\beta_0}(y_{i1}, y_{i2}, ..., y_{iR})} dy \right| \\
&= \sup_{\beta \in U^c} \left| \sqrt{\frac{f_\beta(y_{i1}, y_{i2}, ..., y_{iR})}{f_{\beta_0}(y_{i1}, y_{i2}, ..., y_{iR})}} dy \right| + \sup_{\beta \in U^c} \left| \int \sqrt{f_\beta(y_{i1}, y_{i2}, ..., y_{iR}) f_{\beta_0}(y_{i1}, y_{i2}, ..., y_{iR})} dy \right| \\
&\longrightarrow 0 \quad as \quad R \longrightarrow \infty
\end{aligned}$$

5.9.2 Proof of Lemma 2

Consider a random variable $Z = TB$ and set $Z^{'} = B$.

The Jacobian of transformation is $J(\frac{T,B}{Z,Z^{'}}) = \frac{1}{Z^{'}}$.

Hence the pdf of Z is given by,

$$\begin{aligned} f(Z) \quad &= \int_{-\infty}^{\infty} \frac{exp\Bigg\{ -\frac{1}{2}\Big[\Big(\frac{Z^{'}-\mu_B}{\sigma_B}\Big)^2 + \Big(\frac{\frac{Z^{'}}{Z}-\mu_T}{\sigma_T}\Big)^2\Big]\Bigg\}}{2\pi\sigma_B\sigma_T\Phi(\frac{\mu_T}{\sigma_T})}\frac{1}{Z^{'}}dZ^{'} \\ &= \frac{exp\Bigg\{ -\frac{1}{2}\Big(\frac{\mu_B^2}{\sigma_B^2}+\frac{\mu_T^2}{\sigma_T^2}\Big)\Bigg\}}{2\pi\sigma_B\sigma_T\Phi(\frac{\mu_T}{\sigma_T})}\int_{-\infty}^{\infty} exp\Bigg\{ -\frac{1}{2}\Bigg[\frac{{Z^{'}}^2-2Z^{'}\mu_B}{\sigma_B^2} \\ &+\frac{\frac{Z^2}{{Z^{'}}^2}-2\frac{Z}{Z^{'}}\mu_T}{\sigma_T^2}\Bigg]\Bigg\}\frac{1}{Z^{'}}dZ^{'} \end{aligned}$$

Set,

$$\Psi(Z,Z^{'}) = exp\Bigg\{ -\frac{1}{2}\Bigg[\frac{{Z^{'}}^2-2Z^{'}\mu_B}{\sigma_B^2}+\frac{\frac{Z^2}{{Z^{'}}^2}-2\frac{Z}{Z^{'}}\mu_T}{\sigma_T^2}\Bigg]\Bigg\}$$

Hence,

$$f(Z) = \frac{exp\Bigg\{ -\frac{1}{2}\Big(\frac{\mu_B^2}{\sigma_B^2}+\frac{\mu_T^2}{\sigma_T^2}\Big)\Bigg\}}{2\pi\sigma_B\sigma_T\Phi(\frac{\mu_T}{\sigma_T})}\Bigg[\int_0^{\infty}\Psi(Z,Z^{'})\frac{dZ^{'}}{Z^{'}}+\int_{-\infty}^{0}\Psi(Z,Z^{'})\frac{dZ^{'}}{Z^{'}}\Bigg]$$

Hence we have the similar setting as of Craig (1936). Hence using the same technique we can show that the above function has a closed form expression with an infinite polynomial function of $|Z|$ with the coefficients are being a scaled version of the Bessel function of second kind.

$$f(Z) = \frac{e^{-\frac{\rho_1^2+\rho_2^2}{2}}}{\pi\Phi(\rho_2)}\Bigg[\Sigma_0K_0+(\rho_1^2+\rho_2^2)\frac{|Z|}{Z!}\Sigma_2K_1+(\rho_1^4+\rho_2^4)\frac{|Z|^2}{4!}\Sigma_4K_2+...\Bigg]$$

where,

$K_\gamma(Z) = \frac{1}{2}(\frac{Z}{2})^\gamma \int_0^\infty \frac{e^{-y-\frac{Z^2}{4y}}}{y^{\gamma+1}} dy$

$\rho_1 = \frac{\mu_B}{\sigma B}$ and $\rho_1 = \frac{\mu_T}{\sigma_T}$,

$\Sigma_r(\rho_1\rho_2 Z) = 1 + \frac{\rho_1\rho_2 Z}{r+1} + \frac{(\rho_1\rho_2 Z)^2}{(r+2)^{(2)}2!} + ...$

with $(r+k)^{(k)} = (r+k)(r+k-1)...(r+1)$.

5.9.3 Proof of Lemma 3

$$\begin{aligned}\mathbb{M}_{TB}(\xi) &= \int_{-\infty}^{\infty}\int_0^{\infty} e^{\xi tb}\frac{1}{\sigma_B\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{b-\mu_B}{\sigma_B}\right)^2}\frac{1}{\Phi(\frac{\mu_T}{\sigma_T})\sigma_T\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{t-\mu_T}{\sigma_T}\right)^2}dtdb \\ &= \frac{1}{2\pi\Phi(\frac{\mu_T}{\sigma_T})\sigma_T\sigma_B}\int_{-\infty}^{\infty}\int_0^{\infty} e^{-\frac{1}{2}\left[\left(\frac{b-\mu_B}{\sigma_B}\right)^2+\left(\frac{t-\mu_T}{\sigma_T}\right)^2-2\xi tb\right]}\end{aligned}$$

Now,

$$\begin{aligned}& \left(\frac{b-\mu_B}{\sigma_B}\right)^2 + \left(\frac{t-\mu_T}{\sigma_T}\right)^2 - 2\xi tb \\ = \;& \frac{\sigma_T^2 b^2 - 2b\sigma_T^2\mu_B + \mu_B^2\sigma_T^2 - 2\sigma_B^2\sigma_T^2\xi tb + \sigma_B^2(t^2 + -2t\mu_T + \mu_T^2)}{\sigma_B^2\sigma_T^2} \\ = \;& \frac{1}{\sigma_B^2\sigma_T^2}\{\sigma_T^2 b^2 + 2b(\sigma_T^2\mu_B + \sigma^{B2}\sigma_T^2\xi t) + (\sigma_T^2\mu_B + \sigma^{B2}\sigma_T^2\xi t)^2 \\ & -(\sigma_T^2\mu_B + \sigma^{B2}\sigma_T^2\xi t)^2 + \mu_B^2\sigma_T^2 + \sigma_B^2(t^2 + -2t\mu_T + \mu_T^2)\} \\ = \;& \left[\frac{b-(\mu_B+\sigma_B^2\xi t)}{\sigma_B^2}\right]^2 - \frac{(\sigma_T\mu_B + \sigma_B^2\sigma_T\xi t)^2 - \mu_B^2\sigma_T^2 - \sigma_B^2(t^2 - 2t\mu_T + \mu_T^2)}{\sigma_B^2\sigma_T^2} \\ = \;& \left[\frac{b-(\mu_B+\sigma_B^2\xi t)}{\sigma_B}\right]^2 + \frac{\mu_T^2}{\sigma_T^2} + \left[\frac{t-(\frac{\sigma_T^2\mu_B\xi+\mu_T}{1-\sigma_B^2\sigma_T^2\xi^2})}{\frac{\sigma_T}{\sqrt{1-\sigma_B^2\sigma_2^2\xi^2}}}\right]^2 - \frac{(\sigma_T^2\mu_B\xi+\mu_T)^2}{\sigma_T^2(1-\sigma_B^2\sigma_T^2\xi^2)}\end{aligned}$$

(5.10)

Hence,

$$M_{TB}(\xi) = \frac{\Phi(\frac{\sigma_T^2\mu_B\xi+\mu_T}{\sigma_T\sqrt{1-\sigma_B^2\sigma_T^2\xi^2}})}{\Phi(\frac{\mu_T}{\sigma_T})} \cdot \frac{exp\left\{\frac{1}{2}\frac{(\sigma_T^2\mu_B\xi+\mu_T)^2}{\sigma_T^2(1-\sigma_B^2\sigma_T^2\xi^2)} - \frac{1}{2}\frac{\mu_B^2}{\sigma_B^2}\right\}}{(1-\sigma_B^2\sigma_T^2\xi^2)^{1/2}}$$

Now the first moment of Z can be obtained from the first derivative of $M_{TB}(\xi)$ at $\xi = 0$:

$$\frac{\partial}{\partial\xi}M_{TB}(\xi)|_{\xi=0} = \left\{\frac{\phi(\frac{\mu_T}{\sigma_T})}{\Phi(\frac{\mu_T}{\sigma_T})}\frac{\sigma_T}{\mu_B} + \mu_B\mu_T\right\}\frac{\mu_T}{\sigma_T}\Phi(\frac{\mu_T}{\sigma_T})exp\left\{-\frac{1}{2}\frac{\mu_T^2}{\sigma_T^2}\right\}$$

5.10 Some Posterior Calculations

5.10.1 Posterior Calculation for $\underset{\sim}{b}_g$

$$\begin{aligned}
&P(\underset{\sim}{b}_g \mid rest)\\
&\propto exp\left\{-\frac{1}{2\sigma^2}\sum_{r=1}^{R}||\underset{\sim}{y}_r - \sum_{g=1}^{p}\underset{\sim}{X}_{gr}\otimes\mathbf{V}_g^{1/2}\underset{\sim}{b}_g||_2^2\right\}\\
&\times\prod_{g=1}^{p}\left[(1-\pi_0)(2\pi)^{-N/2}exp\{-\frac{1}{2}\underset{\sim}{b}_g'\underset{\sim}{b}_g\}I(\underset{\sim}{b}_g\neq\underset{\sim}{0}) + \pi_0\delta_0(\underset{\sim}{b}_g)\right]\\
&= exp\left\{-\frac{1}{2\sigma^2}\sum_{r=1}^{R}(\underset{\sim}{y}_r - \mathbf{X}_r\mathbf{V}^{1/2}\underset{\sim}{b})(\underset{\sim}{y}_r - \mathbf{X}_r\mathbf{V}^{1/2}\underset{\sim}{b})\right\}\\
&\times(1-\pi_0)(2\pi)^{-p/2}exp\left\{-\frac{1}{2}\underset{\sim}{b}_g'\underset{\sim}{b}_g\right\}I(\underset{\sim}{b}_g\neq\underset{\sim}{0})\\
&+exp\left\{-\frac{1}{2\sigma^2}\sum_{r=1}^{R}(\underset{\sim}{y}_r - \mathbf{X}_r\mathbf{V}^{1/2}\underset{\sim}{b})(\underset{\sim}{y}_r - \mathbf{X}_r\mathbf{V}^{1/2}\underset{\sim}{b})\right\}\pi_0\delta_0(\underset{\sim}{b}_g)
\end{aligned}$$

where, $\mathbf{X}_r = (\underset{\sim}{x}_{1r}, \underset{\sim}{x}_{2r}, ..., \underset{\sim}{x}_{pr})^{n\times p}$, $\quad \underset{\sim}{b} = (\underset{\sim}{b}_1', \underset{\sim}{b}_2', ..., \underset{\sim}{b}_p')^{np\times 1}$,

$\mathbf{V} = diag(\mathbf{V}_1, \mathbf{V}_2, ..., \mathbf{V}_p)^{np\times np}$

Hence,

$$
\begin{aligned}
&P(\underset{\sim}{b}_g \mid rest)\\
&= exp\Bigg\{ -\frac{1}{2\sigma^2}\sum_{r=1}^{R} ||\underset{\sim}{y}_r - \mathbf{X}_{r(g)}\mathbf{V}_{(g)}^{1/2}\underset{\sim}{b}_{(g)}||_2^2\\
&+\frac{1}{2\sigma^2}\sum_{r=1}^{R}(\underset{\sim}{y}_r - \mathbf{X}_{r(g)}\mathbf{V}_{(g)}^{1/2}\underset{\sim}{b}_{(g)})'\mathbf{X}_{rg}\otimes\mathbf{V}_g^{1/2}\underset{\sim}{b}_g\\
&+\frac{1}{2\sigma^2}\sum_{r=1}^{R}\underset{\sim}{b}_g\mathbf{V}_g^{1/2\prime}\otimes\underset{\sim}{X}_{rg}'(\underset{\sim}{y}_r - \mathbf{X}_{r(g)}\mathbf{V}_{(g)}\underset{\sim}{b}_{(g)})'\\
&-\frac{1}{2}\underset{\sim}{b}_g\Bigg(\mathbf{I}_p - \frac{1}{\sigma^2}\sum_g \mathbf{V}_g^{1/2}\otimes\underset{\sim}{X}_{rg}'\underset{\sim}{X}_{rg}\otimes\mathbf{V}_g^{1/2}\Bigg)\underset{\sim}{b}_g'\Bigg\}\\
&\times(1-\pi_0)(2\pi)^{-p/2}I(\underset{\sim}{b}_g \neq \underset{\sim}{0})\\
&+exp\Bigg\{ -\frac{1}{2\sigma^2}\sum_{r=1}^{R} ||\underset{\sim}{y}_r - \mathbf{X}_r\mathbf{V}^{1/2}\underset{\sim}{b}||_2^2\pi_0\delta_0(\underset{\sim}{b}_g)I(\underset{\sim}{b}_g \neq \underset{\sim}{0})\Bigg\}
\end{aligned}
$$

Set $\mathbf{\Sigma}_g = \Bigg(\mathbf{I}_p - \frac{1}{\sigma^2}\sum_g \mathbf{V}_g^{1/2}\otimes\underset{\sim}{X}_{rg}'\underset{\sim}{X}_{rg}\otimes\mathbf{V}_g^{1/2}\Bigg)$
& $\boldsymbol{\mu}_g = \frac{1}{\sigma^2}\mathbf{\Sigma}_g\sum_{r=1}^{R}\mathbf{V}_g^{1/2\prime}\otimes\underset{\sim}{X}_{rg}'\Bigg(\underset{\sim}{y}_r - \mathbf{X}_{r(g)}\mathbf{V}_{(g)}^{1/2}\underset{\sim}{b}_{(g)}\Bigg)$

So,

$$
\begin{aligned}
&P(\underset{\sim}{b}_g \mid rest)\\
&= (1-\pi_0)|\mathbf{\Sigma}_g|^{1/2}N(\boldsymbol{\mu}_g, \mathbf{\Sigma}_g)exp\Bigg\{ -\frac{1}{2\sigma^2}||\underset{\sim}{y}_r - \mathbf{X}_{r(g)}\mathbf{V}_{(g)}\underset{\sim}{b}_g||_2^2\\
&+\frac{1}{2\sigma^4}||\mathbf{\Sigma}_g^{1/2}\Bigg(\sum_{r=1}^{R}\mathbf{V}_g^{1/2\prime}\otimes\underset{\sim}{X}_{rg}'(\underset{\sim}{y}_r - \mathbf{X}_{r(g)}\mathbf{V}_g^{1/2}\underset{\sim}{b}_g)\Bigg)||_2^2\Bigg\}\times\mathbf{I}(\underset{\sim}{b}_g \neq \underset{\sim}{0})\\
&+\pi_0 exp\Bigg\{ -\frac{1}{2\sigma^2}\sum_{r=1}^{R} ||\underset{\sim}{y}_r - \mathbf{X}_r\mathbf{V}^{1/2}\underset{\sim}{b}||_2^2\Bigg\}\pi_0\delta_0(\underset{\sim}{b}_g)I(\underset{\sim}{b}_g \neq \underset{\sim}{0})
\end{aligned}
$$

5.10.2 Posterior Calculation for τ_{gj}

$$
\begin{aligned}
&P(\tau_{gj} \mid rest) \\
&\propto exp\Big\{ -\frac{1}{2\sigma^2}\sum_{r=1}^{R} ||\underset{\sim}{y}_r - \sum_{i=1}^{p} X_{rg}\otimes \mathbf{V}_g^{1/2}\underset{\sim}{b}_g||_2^2 \Big\} \\
&\times \prod_{i=1}^{p}\prod_{j=1}^{N}(1-\pi_1)2(2\pi s^2)^{-1/2} exp\Big\{ -\frac{1}{2}\frac{(\tau_{gj}-\sum_{i\neq j}\frac{w_{ij}}{w_{i+}}\tau_{gi})^2}{s^2/w_{i+}} \Big\} I(\tau_{gj}>0) + \pi_1\delta_0(\tau_{gj}) \\
&= exp\Big\{ -\frac{1}{2\sigma^2}\sum_{r=1}^{R}(\underset{\sim}{y}_r - \mathbf{X}_{r(gj)}\otimes \mathbf{V}_{gj}^{1/2}\underset{\sim}{b}_{(gj)})'(\underset{\sim}{y}_r - \mathbf{X}_{r(gj)}\otimes \mathbf{V}_{gj}^{1/2}\underset{\sim}{b}_{(gj)}) \Big\} \\
&exp\Big\{ \frac{1}{2\sigma^2}\sum_{r=1}^{R}(\underset{\sim}{y}_r - \mathbf{X}_{r(gj)}\otimes \mathbf{V}_{gj}^{1/2}\underset{\sim}{b}_{(gj)})\underset{\sim}{X}_{rgj}\tau_{gj}b_{gj} \\
&+\frac{1}{2\sigma^2}\sum_{r=1}^{R}\underset{\sim}{X}_{rgj}\tau_{gj}b_{gj}(\underset{\sim}{y}_r - \mathbf{X}_{r(gj)}\otimes \mathbf{V}_{gj}^{1/2}\underset{\sim}{b}_{(gj)}) - \frac{\tau_{gj}^2}{2}\Big[\frac{b_{gj}^2}{\sigma^2}\sum_{r=1}^{R}x_{rgj}^2 + \frac{w_{i+}}{s^2}\Big] \\
&+\frac{\tau_{gj}\sum_{i\neq j}\frac{w_{ij}}{w_{i+}}\tau_{gi}}{s^2/w_{i+}} - \frac{(\sum_{i\neq j}\frac{w_{ij}}{w_{i+}}\tau_{gj})^2}{2s^2/w_{i+}} \Big\}(1-\pi_1)2(\frac{2\pi s^2}{w_{i+}})^{-1/2} I(\tau_{gj}>0) \\
&+exp\Big\{ -\frac{1}{2\sigma^2}\sum_{r=1}^{R}(\underset{\sim}{y}_r - \mathbf{X}_r\otimes \mathbf{V}^{1/2}\underset{\sim}{b})'(\underset{\sim}{y}_r - \mathbf{X}_r\otimes \mathbf{V}^{1/2}\underset{\sim}{b}) \Big\}\pi_1\delta_0(\tau_{gj})
\end{aligned}
$$

Here, $\underset{\sim}{X}_{rgj} = (0,0,...,0,x_{rgj},0,...,0)'$.

Set, $v_{gj}^2 = \left(\frac{w_{i+}}{s^2} + \frac{b_{gj}^2}{\sigma^2}\sum_{r=1}^{R}x_{rgj}^2\right)^{-1}$

$u_{gj} = \frac{v_{gj}^2}{\sigma^2}\sum_{r=1}^{R}(\underset{\sim}{y}_r - \underset{\sim}{X}_{r(gj)}\otimes V_{gj}^{1/2}b_{gj})\underset{\sim}{X}_{rgj}b_{gj} + \frac{v_{gj}^2 w_{i+}}{s^2}\sum_{i\neq j}\frac{w_{ij}}{w_{i+}}\tau_{gj}$

Hence,

$$
\begin{aligned}
&P(\tau_{gj} \mid rest) \\
&= exp\Big\{ -\frac{1}{2\sigma^2}\sum_{r=1}^{R}(\underset{\sim}{y}_r - \mathbf{X}_{r(gj)} \otimes \mathbf{V}_{(gj)}^{1/2}\underset{\sim}{b}_{(gj)})'(\underset{\sim}{y}_r - \mathbf{X}_{r(gj)} \otimes \mathbf{V}_{(gj)}^{1/2}\underset{\sim}{b}_{(gj)})\Big\} \\
&exp\Big\{ -\frac{1}{2\sigma^2}(\sum_{i\neq j}\frac{w_{ij}}{w_{i+}}\tau_{gi})^2 + \frac{1}{2}\frac{u_{gj}^2}{v_{gj}^2}\Big\} \\
&\Phi(\frac{u_{gj}}{v_{gj}})N^+(u_{gj}, v_{gj}^2)(2\pi v_{gj}^2)^{1/2}(1-\pi_1)2(2\pi^2)^{-1/2}I(\tau_{gj} > 0) \\
&+exp\Big\{ -\frac{1}{2\sigma^2}\sum_{r=1}^{R}(\underset{\sim}{y}_r - \mathbf{X}_r \otimes \mathbf{V}^{1/2}\underset{\sim}{b})'(\underset{\sim}{y}_r - \mathbf{X}_r \otimes \mathbf{V}^{1/2}\underset{\sim}{b})\Big\}\pi_1\delta_0(\tau_{gj})
\end{aligned}
$$

5.10.3 Posterior Calculation for s^2

$$
\begin{aligned}
&P(s^2 \mid rest) \\
&\propto \prod_{g=1}^{p}\prod_{j=1}^{N}\Big[(1-\pi_1)2(2\pi s^2)^{-1/2}exp\{ -\frac{(\tau_{gj} - \sum_{i\neq j} w_{ij}\tau_{gj})^2}{2s^2}\}I(\tau_{gj} > 0) \\
&+\pi_1\delta_0(\tau_{gj})\Big] \times t(s^2)^{-2}exp\{ -\frac{t}{s^2}\} \\
&\propto (s^2)^{-(\frac{M}{2}+2)}exp\Big\{ -\frac{1}{s^2}\big[t + \frac{1}{2}\sum_{g=1}^{p}\sum_{j=1}^{N}(\tau_{gj} - \sum_{i\neq j}\frac{w_{ij}}{w_{i+}}\tau_{gj})^2\big]\Big\}
\end{aligned}
$$

where, $M = \#(\tau_{gj} \neq 0) \;\; \forall \;\; g = 1, 2, ..., p \;\; \& \;\; j = 1, 2, ..., N.$

5.11 Some Details for Data Analysis

Table 5.5: Company names with corresponding ticker

Ticker	**Company Names**
$AL.1$	ALCAN INC (RIO TINTO)
HON	HONEYWELL INTERNATIONAL INC
ARV	ARVIN INDUSTRIS INC (MERITOR)
$C.3$	CHRYSLER
CTB	COOPER TIRE & RUBBER COMPANY
DAN	DANA HOLDING CORP
DE	DEERE & CO
ETN	EATON CORP PLC
F	FORD
GE	GENERAL ELETRIC CO
GM	GENERAL MOTORS
$SPXC$	SPX CORP
GR	GOODRICH CORP
GT	GOODYEAR TIRE & RUBBER CO
JCL	JOHNSON CONTROLS INC
$ANV.1$	AEROQUIP-VICKERS INC
OC	OWENS CORNING
PPG	PPG INDUSTRIES INC
AOS	SMITH (A O) CORP
UTX	UNITED TECHNOLOGIES CORP

Table 5.6: Covariate List

Covariate Code	**Covariate Name**
act	CURRENT ASSETS - TOTAL
at	ASSETS - TOTAL
cogs	COST OF GOODS SOLD
gp	GROSS PROFIT (LOSS)
lct	CURRENT LIABILITIES - TOTAL
lt	LIABILITIES - TOTAL
ppegt	PROPERTY, PLANT AND EQUIPMENT - TOTAL (GROSS)
sale	SALES / TURNOVER (NET)

Table 5.7: Response Variable

Response Code	**Response Name**
revt	REVENUE TOTAL

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Anselin, L., Bera, A.K. (1998)
"Spatial dependence in linear regression models with an introduction to spatial econometrics". *Statistics Textbooks and Monographs,* **155** 237-290

[2] Banerjee, S., Carlin, B.P., Gelfand, A.E. (2004)
"Hierarchical Modeling and Analysis for Spatial Data". *Monograph on Statistics and Applied Probability 101.* Chapman & Hall/CRC

[3] Berger, J.O., Bernardo, J.M. (1992).
"On The Development of the reference Prior Method". *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting,* **4**, 35-60.

[4] Berger, J.O., Bernardo, J.M., Sun, D. (2009).
"The formal definition of Reference Priors". *The Annals of Statistics.* 905-938.

[5] Besag, J. (1974).
Spatial Interaction and the Statistical Analysis of Lattice Systems, *Journal of the Royal Statistical Society. Series B (Methodological).* **36**, 192-236.

[6] Bhattacharjee, A., Castro, E., Maiti, T., Marques, J. (2016).
"Endogenous spatial regression and delineation of submarkets: a new framework with application to housing markets". *Journal of Applied Econometrics* **31**, 32-57

[7] Bhattacharjee, A., Holly, S. (2013)
"Understanding interactions in social networks and committees". *Spatial Economic Analysis* **8**, 23-53

[8] Bhattacharjee, A. Jensen-Butler, C. (2013)
"Estimation of the Spatial Weights AMtrix under Structural Constraints". *Regional Science and Urban Economics,* **43**, 617 - 634

[9] Bondell, H.D. and Reich, B.J. (2012)
"Consistent high-dimensional Bayesian variable selection via penalized credible regions". *Journal of American Statistical Association,* **107**: 1610 - 1624

[10] Borgatti, S. P., Mehra, A., Brass, D. J., Labianca, G. (2009)
"Network analysis in the social sciences". *science,* **323**, 892-895

[11] Caimo, A., Friel, N. (2011).
Bayesian inference for exponential random graph models, *Social Networks,* 33, 41-55

[12] Castro, E.A., Zhang, Z., Bhattacharjee, A., Martins, J.M. and Maiti, T. (2015)
"Regional Fertility data Analysis: A small area Bayesian Approach". *Current Trends in Bayesian Methodology with Applications,* 203-224

[13] Casella, G., Giron, F.J., Martinez, M.L. and Moreno, E. (2009)
"Consistency of Bayesian Procedures for Variable Selection". *The Annals of Statistics,* **37**, 1207 - 1228

[14] Castillo, T., Schmidt-Heiber, J. and Van Der Vaart, A. (2015)
"Bayesian Linear Regression with Sparse Priors". *The Annals of Statistics,* **43**, 1986 - 2018

[15] Choi, T., Ramamoorthi, R. V. (2008)
"Remarks on consistency of posterior distributions. In Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh". *Institute of Mathematical Statistics,* 170-186

[16] Craig, C.C. (1936)
"On the Frequency Function Function of xy". *The Annals of Statistics,* **7**, 1-15

[17] Erds, P., Rnyi, A. (1959).
"On random graphs, I". *Publicationes Mathematicae (Debrecen),* **6**, 290-297

[18] Frank, O., Strauss, D. (1986).
"Markov Graphs". *Journal of the American Statistical Association,*. **81**, 832-842

[19] Ghosal, S. (1997).
"A Review of Consistency and Convergence of Posterior Distribution". *In Varanashi Symposium in Bayesian Inference*

[20] Ghosal, S., Ghosh, J.K., Van Der Vart, A.W. (2000)
"Convergence Rates of Posterior Distributions". *The Annals of Statistics,* **28**, 500 - 531

[21] Goodreau, S. M., Handcock, M. S., Hunter, D. R., Butts, C. T., Morris, M. (2008).
"A statnet Tutorial". *Journal of statistical software,* **24**, nihpa54860

[22] Grandori, A., Soda, G. (1995).
"Inter-firm networks: antecedents, mechanisms and forms". *Organization studies,* **16**, 183-214

[23] Holland, P.W. and Leinhardt,S. (1981).
"An exponential family of probability distributions for directed graphs". *Journal of the American statistical Association,*. **76**, 33-50.

[24] Hoff, P.D. (2003) Random Effect Models for Network Data.

[25] Hoff. P.D., Raftery, A.E., Handcock, A.E. (2002).
"Latent Space Approaches to Social Network Analysis". *Journal of the American statistical Association*. **97**, 1090-1098.

[26] Handcock, M.S. and Raftery, A.E. (2007).
"Model Based Clustering for Social Networks". *J.R.Statist.Soc.A*. **170**, 1-22.

[27] Hertzel, M.G, Li, Z, Officer, M.S, Rodgers, K.J. (2008)
Inter-firm linkages and the wealth effects of financial distress along the supply chain. *Journal of Corporate Finance* **87**, 374-387.

[28] Hunter, D. R., Handcock, M. S. (2012).
"Inference in curved exponential family models for networks". *Journal of Computational and Graphical Statistics,* **15**, 565-583

[29] Ishwaran, H. and Rao, J.S. (2005)
"Spike and Slab Variable Selection: Frequentist and Bayesian Stratigies". *The Annals of Statistics,* **33**, 730 - 773

[30] Jiang, W. (2007)
"Bayesian Variable Selection for High Dimentional Generalized Linear Models: Convergence Rates of the Fitted Densities". *The Annals of Statistics,* **35**, 1487 - 1511

[31] Jiang, W. (2005)
On the Consistency of Bayesian Variable Selection for High Dimentional Binary Regression and Classification *Neural Computation,* **18**, 2762-2776

[32] Kleijn, B. J. K.(2013).
"Criteria for Posterior Consistency". *arXiv:1308.1263*

[33] Koskinen, J.H., Snijders, T.A.B. (2007).
"Bayesian inference for dynamic social network data", *Journal of Statistical planning and Inference,*. **137**, 3930 - 3938

[34] Kraft, C. (1955).
"Some conditions for consistency and uniform consistency of statistical procedures", 125-142

[35] Le Cam, L. (1960).
"Locally asymptotically normal families of distributions". *University of California Publications in Statistics,* **3**, 37-98

[36] Lee, K.J., Jones, G.L., Caffo, B.S., Bassett, S.S. (2014).
"Spatial Bayesian Variable Selection Model on Functional Magnatic Resonance Imaging Time-series Data". *Bayesian Analysis,* **9**, 699 - 732

[37] Leskovec, J., Krevl, A. (2015).
SNAP Datasets:Stanford. *Large Network Dataset Collection.*

[38] Mur, J., Angulo, A. (2007)
"The Spatial Durbin Model and the Common Factor Tests". *Spatial Economic Analysis,* **1**, 207-226

[39] Murray, I., Ghahramani, Z., MacKay, D. (2012).
"MCMC for doubly-intractable distributions". *arXiv preprint arXiv:1206.6848*

[40] Narasimhan, R., Jayaram, J. (1998).
"Causal linkages in supply chain management: an exploratory study of North American manufacturing firms". *Decision sciences,* **29**, 579-605

[41] Narisetty, N.N., He, X (2014)
"Bayesian variable selection with shrinking and diffusing priors". *Annals of Statistics,* **42**, 789-817

[42] Nowicki, Krzysztof, and Tom A. B. Snijders (2001)
"Estimation and prediction for stochastic blockstructures". *Journal of the American Statistical Association,* **96**. 1077-1087

[43] Park, T., Casella, G. (2008)
"The bayesian lasso". *Journal of the American Statistical Association,* **103**, 681-686

[44] Pesaran, M. H. (2004).
"General diagnostic tests for cross section dependence in panels".

[45] Pattison, P., Wasserman, S (1998)
"Logit models and logistic regressions for social networks: II. Multivariate relations". *British Journal of Mathematical and Statistical Psychology,* **42**, 789-817

[46] Ramcharran, H. (2001)
"Inter-Firm Linkages and Profitability in the Automobile Industry: The implication for Supply Chain Management". *The Journal of Supply Chain Management,* **37**, 11-17

[47] Robins, G., Pattison, P., Kalish, Y., Lusher, D. (2007).
"An introduction to exponential random graph (p*)". *Social Networks,* **29**, 173-191.

[48] Scutari, M. (2013).
"On the Prior and Posterior Distributions Used in Graphical Modelling". *Bayesian Analysis,.* **8**, 505-532

[49] Schwartz, L. (1965).
"On bayes procedures". *Zeitschrift fr Wahrscheinlichkeitstheorie und verwandte Gebiete,.* **4**, 10-26

[50] Smith, M. and Fahrmeir, L. (2007)
"Spatial Bayesian Variable Selection with Application to Functional Magnatic Resonance Imaging". *Journal of American Statistical Association,* **102**, 417-431

[51] Smith, M., Kohn, R. (1996)
"Nonparametric regression using Bayesian variable selection". *Journal of Econometrics,* **75**, 317 - 343

[52] Shalizi, C.S., Rinaldo, A. (2013)
"Consistency Under Sampling of Exponential Random Graph Models". *The Annals of Statistics,.* **41**, 508-535

[53] Strauss, D., Ikeda, M. (1990)
"Pseudolikelihood Estimation for Social Networks". *Journal of the American Statistical Association,.* **85**, 204-212

[54] Tibshirani, R. (1996)
"Regression shrinkage and selection via the lasso". *Journal of the Royal Statistical Society. Series B (Methodological),.* **58**, 267-288

[55] Van de Geer, S. (1993)
"Hellinger-consistency of certain nonparametric maximum likelihood estimators likelihood estimation of exponential family random graph models". *The Annals of Statistics,*. **21**, 14-44

[56] Van Duijn, M.A., Gile, K.J., Handcock, M.S. (2009)
"A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models". *Social Networks,*. **31**, 52-62

[57] Wang, J. (2012)
"Do Firms' Relationship with Principal Customer/ Suppliers Affect Shareholders Income?". *Journal of Corporate Finance,* **18**, 860-878

[58] Weisstein, E.W. (2003)
"Normal Product Distribution". *From MathWorld-A Wolfram Web Resource*

[59] Xu, X. and Ghosh, M. (2015)
"Bayesian Variable Selection and Estimation for Group LASSO". *Bayesian Analysis,* **10**, 909-936

[60] Zou, H. (2006)
"The Adaptive LASSO and It's Oracle Properties". *Journal of American Statistical Association,* **101**, 1418-1429