

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

MARIANA HAUER

**OS MODELOS VAR E VEC ESPACIAIS: UMA ABORDAGEM
BAYESIANA**

Porto Alegre

2007

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

MARIANA HAUER

**OS MODELOS VAR E VEC ESPACIAIS: UMA ABORDAGEM
BAYESIANA**

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como quesito parcial para obtenção do grau de mestre em Economia.

Orientador: Flávio A. Ziegelmann

Porto Alegre

2007

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
Responsável: Biblioteca Gládis W. do Amaral, Faculdade de Ciências Econômicas da
UFRGS

H368m

Hauer, Mariana

Os modelos VAR e VEC espaciais: uma abordagem bayesiana /
Mariana Hauer. – Porto Alegre, 2007.
200 f. : il.

Orientador: Flávio A. Ziegelmann.

Dissertação (Mestrado em Economia) - Universidade Federal do Rio
Grande do Sul, Faculdade de Ciências Econômicas, Programa de Pós-
Graduação em Economia, Porto Alegre, 2007.

1. Econometria espacial. 2. Modelo de previsão. 3. Econometria:
Matemática aplicada. 4. Estatística aplicada. 5. Modelo econométrico.
I. Ziegelmann, Flávio A. II. Universidade Federal do Rio Grande do Sul.
Faculdade de Ciências Econômicas. Programa de Pós-Graduação em
Economia. III. Título.

CDU 330.115
519.24

MARIANA HAUER

**OS MODELOS VAR E VEC ESPACIAIS: UMA ABORDAGEM
BAYESIANA**

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como quesito parcial para obtenção do grau de mestre em Economia.

Aprovada em Porto Alegre, 14 de setembro de 2007.

Prof. Dr. Marcelo S. Portugal

UFRGS

Prof. Dr. Lorí Viali

PUCRS

Prof. Dr. Paulo de Andrade Jacinto

Universidade Federal de Alagoas

Agradecimentos

Em primeiro lugar eu gostaria de agradecer ao Prof. Flávio Ziegelmann que, desde a escolha do tema, me ajudou muito e sempre foi muito solícito, indicando textos interessantes, encaminhando convites de seminários sobre econometria espacial que serviram de motivação, despertando meu interesse nesta área. Gostaria ainda de agradecer-lo por sua competência e seriedade, que foram fundamentais para a consecução deste trabalho.

Gostaria de agradecer também ao Prof. Marcelo Portugal, que não só aceitou participar da minha banca para a apresentação do projeto deste trabalho, como sugeriu um tema a ser abordado na aplicação, indicando textos que me ajudaram a compreender e a me interessar ainda mais pelo assunto abordado. Além dele, gostaria de agradecer ao Prof. Márcio Laurini, do Ibmec SP, pelo incentivo e indicação de livros que cobriam o assunto sobre a econometria Bayesiana.

Sou grata às minhas colegas de mestrado Andreza Palma e Nina Yano, que mesmo à distância sempre estiveram ao meu lado através de trocas de emails que dividiam não só os assuntos acadêmicos como os assuntos cotidianos, me acalmando e me ajudando.

Agradeço também à minha família – minha mãe Maria Cecília, meu pai Georges e meus irmãos Gustavo e Gabriel – pelo apoio incessante e a meus amigos – Renata Pestana, Thais Lutti, Carolina Pinto, Ellem Andreosi e Walter Guimarães – pelo apoio e conforto emocional.

Por fim, devo ainda mencionar o apoio das instituições, sem o qual este trabalho não teria sido possível. Meus agradecimentos a CAPES, pela bolsa referente ao curso de mestrado e à UFGRS, que foi fundamental para o meu crescimento não só acadêmico.

Resumo

O objetivo deste trabalho é apresentar o Modelo Vetorial Autorregressivo (VAR) e uma das suas variações, o Modelo Vetorial de Correções de Erros (VEC), segundo uma abordagem Bayesiana, considerando componentes regionais, que serão inseridos nos modelos apresentados através de informações *a priori* que levam em consideração a localização dos dados. Para formar tais informações *a priori* são utilizados conceitos referentes à econometria espacial, como por exemplo, as relações de contigüidade e as implicações que estas trazem. Como exemplo ilustrativo, o modelo em questão será aplicado a um conjunto de dados regionais, coletados por estados brasileiros. Este conjunto de dados consiste em observações da variável produção industrial para oito estados, no período de janeiro de 1991 a setembro de 2006. Em função da escolha do modelo adequado, a questão central foi descobrir em que medida a incorporação destas informações *a priori* no modelo VEC Bayesiano é coerente quando estimamos modelos que consideram informações localizacionais.

Palavras-chave: Econometria espacial. Modelos VAR e VEC. Produção industrial.

Abstract

The main goal of this work is to present the Vector Autoregressive Model (VAR) and one of its variations, the Vector Error Correction Model (VEC), according to a Bayesian variant, considering regional components that will be inserted in the models presented through prior information, which takes in consideration the data localization. To form such prior information, spatial econometrics is used, as for example the contiguity relations and the implications that these bring to the modeling. As illustrative example, the model in question will be applied to a regional data set, collected for Brazilian states. This data set consists of industrial production for eight states, in the period between January 1991 and September 2006. The central question is to uncover whether the incorporation of these prior informations in the Bayesian VEC Model is coherent when we use models that consider contiguity information.

Keywords: Spatial Econometric. VAR and VEC Models. Industrial Production.

Sumário

1	Introdução.....	9
2	A Econometria Espacial: alguns conceitos e modelos	11
2.1	Dependência Espacial.....	13
2.2	Heterogeneidade Espacial.....	16
2.3	Quantificando localização nos modelos	18
2.4	Modelos de Regressão Espacial.....	27
3	A Abordagem Bayesiana.....	38
3.1	Análise Econométrica.....	39
3.2	Uma noção intuitiva para a abordagem Bayesiana.....	40
3.3	O Teorema de Bayes.....	43
3.4	Estimação por simulação: MCMC.....	61
4	O Modelo VAR Espacial com uma abordagem Bayesiana.....	68
4.1	O modelo VAR Bayesiano com informação <i>a priori</i> de Minnesota	70
4.2	O Modelo VAR Bayesiano com informação <i>a priori</i> espacial.....	73
4.3	Uma extensão do Modelo VAR Bayesiano: o Modelo VEC Bayesiano.....	79
4.4	O Modelo Vetorial Bayesiano de Correção de Erros – VECM Bayesiano ..	81
4.5	A Função Verossimilhança para os Modelos VAR e VEC Bayesianos.....	84
5	Uma aplicação para o Modelo VEC Bayesiano: o caso da Produção Industrial no Brasil.....	86
5.1	A escolha do modelo	88
5.2	A escolha do método de estimação.....	90
5.3	Resultados Empíricos	96
5.4	Conclusão	118
6	Considerações Finais	120
	Referências	122

Apêndice A - Apresentação dos resultados dos testes iniciais necessários para escolha do modelo utilizado no exercício proposto no Cap. 5.	128
Apêndice B - Apresentação dos resultados empíricos do modelo VEC Bayesiano utilizando informação <i>a priori</i> espacial estimado por Gibbs Sampling.	132
Apêndice C - Apresentação dos resultados empíricos do modelo VEC Bayesiano utilizando informação <i>a priori</i> de Minnesota modificada e estimado por Gibbs Sampling	165
Apêndice D - Tabelas usadas para o cálculo do erro percentual de previsão dentro da amostra no período de setembro/2003 a setembro/2006.	197

1 Introdução

A atenção para a importância da localização ou para a interação espacial tem ganhado espaço recentemente não só na econometria aplicada como também na econometria teórica. No passado, modelos que incorporavam de forma explícita o espaço (geografia), e, portanto aplicações de econometria espacial, foram utilizadas primeiramente em campos como a ciência regional, a economia urbana e a geografia econômica. Recentemente, métodos de econometria espacial são mais e mais aplicados em investigações empíricas de campos tradicionais da economia incluindo, entre outros, estudos de análise de demanda, economia do trabalho, economia do setor público, finanças públicas regionais, economia agrícola e economia ambiental. No campo teórico, existe um aumento de interesse principalmente na literatura que aborda diferentes metodologias que consideram especificações alternativas dos modelos, estimadores e testes estatísticos com o objetivo de incorporar a característica geográfica dos dados.

O objetivo geral deste trabalho é apresentar o modelo VEC espacial, explicitando os conceitos teóricos referentes à econometria espacial que são básicos para a compreensão do modelo, assim como suas limitações. Por exemplo, a estimação de um modelo VEC espacial pode ser complicada devido à superparametrização. Se a especificação do VEC for um pouco mais geral, pelo fato deste modelo considerar tanto a dimensão do espaço como do tempo, haverá muitos parâmetros a serem estimados. Dependendo do número de observações em questão, pode ser inviável computacionalmente estimar todos os parâmetros. Uma forma encontrada de contornar este problema é tratá-lo através da abordagem Bayesiana que considera informação *a priori*, tornando possível a estimação dos parâmetros do modelo. O objetivo específico deste trabalho é estimar o modelo VEC Espacial Bayesiano, utilizando dois tipos de informação *a priori*, *a priori* espacial e *a priori* de Minnesota, que incorporam as relações de contigüidade entre os dados. Para tanto, serão utilizados dados trimestrais referentes à produção industrial de oito estados brasileiros.

O Capítulo 2 apresenta os conceitos sobre a econometria espacial, incluindo os diferentes tipos de abordagem que tratam problemas que surgem no tratamento de dados regionais e modelos comumente utilizados para tanto.

O Capítulo 3 apresenta os conceitos referentes à econometria Bayesiana, fundamentais para entender o processo de estimação utilizado neste trabalho. Introduce os conceitos básicos da Teoria da Probabilidade e desenvolve-se até a apresentação do processo de estimação iterativo MCMC, utilizado no Capítulo 5.

O Capítulo 4 apresenta os modelos VAR e VEC Bayesianos e as diferentes especificações de informações *a priori* utilizadas, explicando como as relações de contigüidade podem ser incorporadas no modelo através destas informações *a priori*. Para tanto, são apresentados dois tipos e especificações de informação *a priori*, que consideram as relações de vizinhança de formas distintas: a informação *a priori* espacial e a informação *a priori* de Minnesota.

O Capítulo 5 é dedicado a apresentação de resultado utilizando os modelos apresentados no Capítulo 4. Para tanto, os dados utilizados são referentes à produção industrial em oito estados brasileiros: Bahia, Minas Gerais, Espírito Santo, Rio de Janeiro, São Paulo, Paraná, Santa Catarina e Rio Grande do Sul. É feita uma comparação entre os modelos VEC Bayesianos utilizando as duas formas de informação *a priori* e analisado qual modelo melhor se ajusta aos dados assim como o motivo desta configuração.

O Capítulo 6 conclui, mostrando que o VEC Bayesiano utilizando a informação *a priori* de Minnesota é o modelo que melhor prevê os dados sobre produção industrial, comparado ao modelo VEC Bayesiano com informação *a priori* espacial. Isto significa que o modelo que considera tanto a dependência temporal quanto a dependência espacial ponderados da mesma forma explica melhor os dados referentes à produção industrial que, neste caso, é o modelo VEC Bayesiano com informação *a priori* de Minnesota.

2 A Econometria Espacial: alguns conceitos e modelos

Historicamente, a econometria espacial vista como um campo da econometria foi identificada na Europa no início dos anos 70, impulsionada pela necessidade de lidar com dados regionais em modelos econométricos. Paelink e Klassen (1979) definiram o campo em cinco características distintas: (a) a regra da interdependência espacial em modelos espaciais; (b) a assimetria nas relações espaciais; (c) a importância de fatores explicativos localizados em outros espaços; (d) diferenciação entre iterações *ex-ante* e *ex-post*; (e) modelagem explícita do espaço.

Anselin (1988) define a econometria espacial como uma coleção de técnicas que consideram as peculiaridades causadas pelo espaço na análise estatística de modelos regionais. Em outras palavras, a econometria espacial trata dos interesses metodológicos que surgem em considerar diretamente os efeitos espaciais, tal como a autocorrelação e a heterogeneidade espacial. A econometria espacial é uma área da econometria que lida com o tratamento das interações espaciais e estrutura espacial em modelos de regressão para dados em painel e dados em *cross-section*.

No contexto de regressão, os efeitos espaciais são caracterizados por dependência espacial ou, numa linguagem mais conhecida, autocorrelação e heterogeneidade espacial. Esta última é simplesmente a instabilidade estrutural na forma da variância não-constante dos erros (heterocedasticidade) ou de coeficientes variáveis. No entanto, existem três razões importantes em considerar-se a heterogeneidade de uma forma explícita. Primeira, a estrutura por trás da instabilidade é espacial (ou geográfica), de forma que a localização das observações é crucial na determinação da forma da instabilidade. Segunda que, uma vez que a estrutura é espacial, a heterogeneidade pode ocorrer conjuntamente com a autocorrelação espacial e as técnicas convencionais da econometria podem não ser muito apropriadas. Por exemplo, resultados de testes para heterocedasticidade podem ser equivocados no caso que a heterogeneidade se mistura com a autocorrelação espacial. Terceira que, se estivermos trabalhando com uma amostra de dados em *cross-section*, a autocorrelação espacial e a heterogeneidade espacial podem ser equivalentes do ponto de vista observacional. Por exemplo, suponha que estamos interessados numa determinada região e que, de acordo com a econometria espacial, os dados observados desta região influenciam e sofrem influência dos dados

das regiões vizinhas. No caso em que a região de interesse seja, por exemplo, um outlier, isto pode ser interpretado como heterogeneidade espacial, para o caso em que esta região influencia as regiões vizinhas, mas “destoa” das outras regiões, mas também pode ser interpretado como autocorrelação espacial no caso em que esta região influencia suas regiões vizinhas e não influencia as outras regiões. (ANSELIN, 1999).

A autocorrelação espacial surge quando as observações em uma região são sistematicamente dependentes das observações nas regiões vizinhas, enquanto a dependência diminui para regiões mais distantes. A consideração da autocorrelação espacial é relevante para diversos campos da economia, tais como economia regional, economia ambiental, economia urbana, organização industrial e economia internacional. Modelos econométricos que consideram explicitamente as dependências espaciais tiveram sua aplicação na economia inicialmente restrita a poucos campos, especialmente a economia regional. Mais recentemente tem havido um interesse crescente em modelos de dependência espacial aplicados à teoria econômica, talvez pelo crescente avanço da teoria econômica em modelos que consideram as interações entre agentes heterogêneos e também, pelo lado empírico, o aumento de dados disponíveis numa forma não-agregada e também dados que apresentam referência geográfica (GIACOMINI; GRANGER, 2004).

Os trabalhos aplicados em economia regional utilizam dados amostrais coletados com referência à localização, medida como pontos no espaço. O estudo baseado na incorporação do componente geográfico nos dados amostrais é tratado pela econometria espacial. Este novo foco de especificação, estimação e teste da presença de interação espacial pode ser atribuído a dois principais fatores. O primeiro é o interesse nas interações "diretas" dos agentes, através de normas sociais, efeito vizinhança e também relacionado ao aumento do interesse em questões do tipo “as interações individuais podem levar ao comportamento coletivo e a padrões agregados?”. Além da necessidade de uma metodologia que trate dos efeitos geográficos, existe o segundo fator, que é o aumento das bases de dados disponíveis que possibilitam trabalhos com esse enfoque (ANSELIN, 1999).

2.1 Dependência Espacial

O crescente interesse em trabalhar-se com dados com características regionais motivou o aparecimento de novos modelos e técnicas para quantificar os padrões apresentados pelos mapas. Surgiram, então, um grande número de estatísticas espaciais e outras medidas-resumo, principalmente como uma revisão do tratamento clássico da autocorrelação espacial proposto por Cliff and Ord (1979 *apud* ANSELIN, 2000) ¹.

A presença de autocorrelação espacial, e conseqüentemente dependência espacial, em uma coleção de dados amostrais significa que as observações na localização i dependem das observações na localização $j \neq i$, para quaisquer localizações i, j . Também vale notar que as observações na localização i podem depender das observações de mais de uma localização, isto é, de diferentes j 's. Formalmente, podemos estabelecer:

$$y_i = f(y_j), i = 1, \dots, n; j \neq i \quad (2.1)$$

A presença ou ausência de dependência espacial é indicada pelo conceito de autocorrelação espacial, em outras palavras, pela coincidência da similaridade dos valores relacionada à similaridade na localização. Isto é, quando altos valores em uma região tendem a ser associados com altos valores em regiões vizinhas, temos indícios de uma autocorrelação espacial positiva. Mas quando altos valores em uma região estão cercados por baixos valores nas regiões vizinhas, ou vice versa, existe indício de autocorrelação negativa, presente na forma de um *outlier* espacial² (ANSELIN, 2000).

Por que podemos esperar que dados amostrais observados em um ponto no espaço sejam dependentes de valores observados em outras localizações³? Existem duas

¹ Inicialmente, o interesse na identificação de *clusters* e *outliers* nos mapas ocorreram nos campos da epidemiologia e da estatística médica e há uma vasta literatura abrangendo este assunto, que pode ser revisto no trabalho de Marshal, 1991 (ANSELIN, 2000).

² O ponto de referência na análise de autocorrelação espacial é a aleatoriedade espacial, ou falta de referência de qualquer estrutura.

³ A discussão acerca de autocorrelação espacial baseia-se em situações onde os dados apresentam-se na forma de pontos e suas localizações é o foco central de interesse.

principais razões. Primeiro, os dados coletados de observações associadas a unidades espaciais tais como as organizadas por códigos postais, regiões, estados, censos e assim por diante, podem refletir medidas de erros. Isso ocorre porque os meios para coleta de dados não refletem de forma precisa a natureza do processo geracional dos dados da amostra. Como um exemplo, considere o caso das medidas de taxas de desemprego e força de trabalho. Uma vez que os trabalhadores têm mobilidade para mudar de localização para procurar emprego nas áreas vizinhas, as medidas de taxa de desemprego e força de trabalho baseadas na localização onde as pessoas residem, podem exibir dependência espacial. Uma segunda razão é que a dimensão espacial das atividades econômicas ou regionais pode ser realmente um aspecto importante para a modelagem de problemas. A ciência regional é baseada na premissa de que distância e localização são forças importantes em trabalhos de geografia humana e atividades de mercado. Todas essas noções precisam ser formalizadas na teoria da ciência regional, que se baseia na interação espacial e difusão de efeitos e externalidades espaciais.

Como o processo gerador de dados que apresentam o componente de localização difere da visão tradicional? A visão tradicional de uma regressão é que o processo gerador dos dados toma a forma de (2.2) abaixo:

$$y = X\beta + \varepsilon \quad (2.2)$$

onde y representa um vetor de n observações, X é uma matriz $n \times k$ de variáveis explicativas, β é um vetor coluna de k parâmetros e ε é um vetor n de erros estocásticos.

O processo gerador acima, em sua forma mais simples, é tal que a matriz X e os verdadeiros parâmetros β são fixos enquanto o vetor de erros aleatórios ε representa a parte estocástica do vetor y . Dado que a matriz X e os parâmetros β são fixos, a distribuição da amostra no vetor y terá a mesma estrutura de variância-covariância que o vetor ε . Suposições acerca da natureza da estrutura da variância-covariância de ε assumem que a distribuição das observações individuais em y exibe uma variância constante e que a covariância entre as observações é zero.

Normalmente observações a partir de amostras que apresentam dependência espacial não obedecem a essa estrutura. As observações com componente espacial serão altamente correlacionadas, apresentando padrão de decaimento da correlação à medida que comparamos uma observação a outras localizadas mais distantes no espaço. A dependência espacial que surge a partir das interações regionais em dados utilizados pela ciência regional sugere a necessidade de quantificar e modelar a natureza desta dependência.

A autocorrelação espacial pode ser formalmente apresentada pela condição de momento (ANSELIN, 2000):

$$\text{cov}(y_i, y_j) = E(y_i y_j) - E(y_i) \cdot E(y_j) \neq 0 \text{ para } i \neq j \quad (2.3)$$

onde i, j referem-se às localizações individuais e y é a variável aleatória.

Em um conjunto de dados em *cross-section* com n observações, não existem informação suficiente para estimar a matriz de variância-covariância $n \times n$ (2.3) diretamente a partir dos dados. Mesmo a forma assintótica não ajuda uma vez que o número de covariâncias aumenta em n^2 quando o tamanho da amostra só aumenta em n . Em contraste, quando temos amostras repetidas em dados de *cross-section* (por exemplo, dados em painel), pode ser possível explorar outra dimensão e obter estimativas consistentes para a matriz de covariância. Em geral, no entanto, será necessário impor uma estrutura na covariância (ANSELIN, 1999).

Existem três principais frentes na literatura dirigidas a tratar este assunto: uma é baseada na especificação do processo estocástico espacial; a segunda dirige-se à representação paramétrica da estrutura de covariância; e uma terceira onde a covariância é tratada de uma forma geral, não-especificada, através da abordagem não-paramétrica⁴.

⁴ Serão apresentadas na seção 1.5 algumas formas gerais da modelagem direcionada à econometria espacial em que a covariância é tratada por estas diferentes formas.

2.2 Heterogeneidade Espacial

Heterogeneidade espacial é a instabilidade estrutural na forma de variância não-constante dos erros ou coeficientes variáveis (ANSELIN, 1999). Isto é, o termo heterogeneidade espacial refere-se à variação nas relações no espaço (LESAGE, 1999, cap.1, p.7). Nos casos mais gerais, podemos esperar que se mantenha uma relação diferente em todo ponto no espaço. Formalmente, uma relação linear descrevendo esta característica pode ser escrita como:

$$y_i = X_i \beta_i + \varepsilon_i \quad , \quad (2.4)$$

onde a observação i é a observação coletada em $i=1, \dots, n$ pontos no espaço, X_i representa um vetor $1 \times k$ de variáveis explicativas, associadas ao conjunto de parâmetros β_i , y_i é a variável dependente na localização i e ε_i é o vetor de erros estocásticos na relação linear. Uma forma um pouco mais complicada de expressar essa idéia é permitir que a função $f(\cdot)$ da equação (2.1) varie com a observação i , isto é:

$$y_i = f_i(X_i \beta_i + \varepsilon_i) \quad (2.5)$$

Olhando para a formulação descrita em (2.4), não podemos estimar um conjunto de n vetores de parâmetros β_i , dada uma amostra de n observações. Isto porque não temos informação suficiente de dados amostrais para produzir estimativas para todo ponto no espaço - um fenômeno chamado de problema nos graus de liberdade. Para continuar com esta análise, precisamos fornecer uma especificação para a variação no espaço, que seja parcimoniosa, isto é, somente com um número considerável de parâmetros⁵.

⁵ Uma das motivações para o uso de métodos bayesianos na econometria espacial é a sua habilidade em impor restrições, que são de natureza estocástica e não exata, fixa. Métodos bayesianos nos permitem impor restrições variando o grau de incerteza *a priori*. No limite, se a restrição imposta tiver um alto grau de certeza, ela se torna uma restrição fixa.

LESAGE (1999, cap.1, p.8) aponta diferentes questões surgidas a partir das diferentes especificações, tais como:

- (1) quão sensíveis são as inferências a partir de diferentes especificações tratando as variações espaciais?
- (2) determinada especificação é consistente com a informação dos dados amostrais?
- (3) como diferentes especificações para um determinado fenômeno se comportam e quais tipos de inferências elas fornecem?

Anselin (1999) aponta três principais razões para considerar a heterogeneidade espacial de forma explícita, como citado na introdução deste capítulo. Na primeira delas está a estrutura por trás da instabilidade espacial, de forma que a localização dos dados é crucial para determinar como se apresenta esta instabilidade. Por exemplo, um grupo que apresenta heterocedasticidade pode ser modelado com diferentes variâncias nos erros para diferentes subconjuntos geográficos dos dados. Mais formalmente, considere um conjunto S com n unidades geográficas, divididas em R subconjuntos compactos não-sobrepostos S_r (com $r=1, 2, \dots, R$) tal que para qualquer r, s ($r \neq s$), $S_r \cap S_s = 0$, e $\bigcup_{r=1, \dots, R} S_r = S$. A heterocedasticidade espacial referente a cada um dos grupos segue então a forma das variâncias nos erros espacialmente agrupadas para a observação i , $\text{var}(\varepsilon_i) = \sigma_r^2$ quando $i \in S_r$. Da mesma maneira, os coeficientes variáveis da regressão podem ser definidos de tal forma que eles correspondam aos regimes espaciais, ou subconjuntos geográficos S_r dos dados onde o modelo apresenta diferente inclinação, $\beta_i = \beta_r$, para uma observação $i \in S_r$. A segunda razão é que muitas vezes a heterogeneidade espacial ocorre ao mesmo tempo em que a autocorrelação espacial e as técnicas da econometria padrão nem sempre podem ser utilizadas da maneira convencional. A terceira razão é que a heterogeneidade espacial e a autocorrelação espacial podem ser equivalentes do ponto de vista observacional. Este fato requer que ambos os aspectos do problema sejam estruturados com bastante cuidado para que os parâmetros não se tornem não-identificados. Então, um problema não deve ser considerado isolado do outro.

2.3 Quantificando localização nos modelos

Antes de nos preocuparmos em como a dependência e a heterogeneidade espaciais entram nos diferentes modelos, vamos olhar os aspectos da localização dos dados amostrais. Dado que as observações são organizadas pelas localizações, temos duas fontes de informação que devemos considerar.

A localização no espaço cartesiano, representada pelas coordenadas de latitude e longitude, é uma fonte de informação. Esta informação nos permite calcular as distâncias entre as observações localizadas em pontos distintos do espaço. Observações próximas podem refletir um maior grau de dependência espacial, comparada àquelas mais distantes, sugerindo que a força da dependência espacial entre as observações deve diminuir com o aumento da distância entre as mesmas.

A distância também pode ser importante em modelos que consideram a heterogeneidade espacial. Caso a relação que estivermos modelando varie no espaço, as observações próximas podem exibir um padrão similar de variação e aquelas que estão mais distantes, um padrão distinto.

Outra fonte de informação sobre a localização que devemos considerar são as relações de contigüidade, que refletem as posições relativas da localização das observações. Medidas de contigüidade baseiam-se no conhecimento do tamanho e do formato das unidades descritas em um mapa. A partir disso, podemos determinar quais unidades são vizinhas (que têm bordas que se tocam) ou quais unidades estão próximas uma das outras. Relativo à dependência espacial, unidades vizinhas devem apresentar um maior grau de dependência espacial comparadas a unidades não-vizinhas. Para heterogeneidade espacial, as relações devem ser similares para as unidades vizinhas. Essas duas fontes de informação não necessariamente são diferentes.

Vamos ilustrar como ambos os tipos de fontes de informação localizacional podem ser usadas na modelagem econométrica espacial.

2.3.1 Quantificando contigüidade espacial

Analogamente à análise de séries de tempo, os processos estocásticos espaciais podem ser caracterizados como processos autorregressivos espaciais (SAR) e médias-móveis espaciais (SMA), ainda que existam diferenças entre as metodologias utilizadas em séries de tempo ou em *cross-section*. No entanto, em contraste com a noção de deslocamento ao longo do eixo do tempo, não existe um conceito correspondente no domínio do espaço, principalmente porque as observações são distribuídas de uma forma irregular geograficamente (ANSELIN, 1999).

A Figura 2.1 mostra um exemplo hipotético apresentado por LeSage (1999, cap.1, p.11), que ilustra como cinco regiões podem se apresentar em um mapa. Desejamos construir uma matriz binária W de dimensão 5×5 , contendo 25 elementos que tomam valores de 0 ou 1 com o objetivo de capturar a noção de conectividade (ou vizinhança de primeira ordem) entre as cinco regiões descritas no mapa. Cada relação de contigüidade relativa a cada uma das cinco regiões é registrada nas linhas e colunas da matriz W . Por exemplo, o elemento da linha 1 e coluna 2 da matriz registra a presença (representada por 1) ou ausência (representada por 0) de uma relação de contigüidade entre a região 1 e a região 2. Desta forma, uma matriz construída a partir das relações de contigüidade deve ser simétrica uma vez que se a região 1 é contígua à região 2, então a região 2 deve ser contígua à região 1.

Existem muitas formas de construir uma matriz que apresente informações de contigüidade sobre as regiões. Abaixo enumeramos algumas formas alternativas para definir-se uma matriz binária W que reflita as relações de contigüidade entre as cinco regiões apresentadas na Figura 2.1. Comece com uma matriz nula (de zeros) e então considere as seguintes formas para definir a presença de relações de contigüidade.

Contigüidade linear: Defina $W_{ij} = 1$ para as regiões que dividem a mesma margem diretamente à direita ou à esquerda da região de interesse. Para a linha 1, onde apresentamos as relações associadas à região 1, teremos todos $W_{1j} = 0, j=1, \dots, 5$. Por outro lado, para a linha 5, onde apresentamos as relações associadas à região 5, teremos $W_{53} = 1$ e todos os outros elementos da linha iguais a zero.

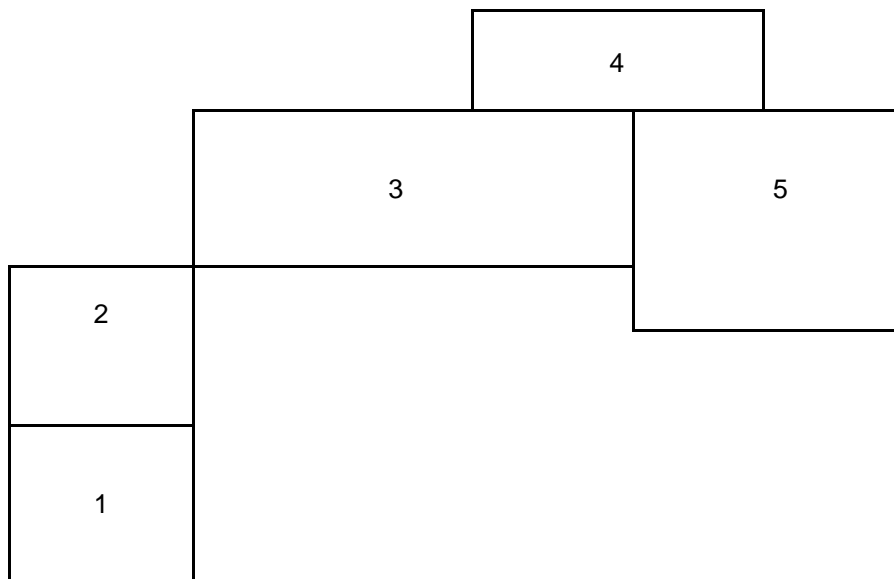


Figura 2.1 – Exemplo de uma estrutura de regiões vizinhas que serão base para a formação dos tipos de matrizes de contigüidade apresentados nesta seção.

Fonte: LeSage, 1999, p. 11.

Contigüidade rook: Defina $W_{ij} = 1$ para as regiões que dividem a mesma margem diretamente, em qualquer direção, com a região de interesse. Para a linha 1, onde apresentamos as relações associadas à região 1, teremos $W_{12} = 1$ e todos os outros elementos desta linha iguais a zero. Como outro exemplo, a linha 3 apresentará $W_{34} = 1$ e $W_{35} = 1$ e todos os outros elementos da linha iguais a zero.

Contigüidade linear de segunda ordem: Defina $W_{ij} = 1$ para as duas regiões que estão diretamente à direita e à esquerda da região de interesse. Esta definição irá produzir a mesma matriz que a contigüidade linear apresenta para o nosso exemplo.

Contigüidade rook de segunda ordem: Para as duas regiões que estão diretamente à direita, esquerda, acima ou abaixo da região de interesse, defina $W_{ij} = 1$. Esta definição também produzirá a mesma matriz W que a contigüidade *rook* apresenta para o exemplo da Figura 2.1.

Contigüidade queen: Defina $W_{ij} = 1$ para as regiões que apresentam um lado ou um vértice em comum à região de interesse. Por exemplo, para a região 3, teremos $W_{34} = 1$ e $W_{35} = 1$ e todos os outros elementos da linha 3 iguais a zero.

Além destas, é claro, existem outras formas de definir as relações de contigüidade. Outras formas, por exemplo, se baseiam no comprimento da margem entre as regiões ou até mesmo na distância de uma região à outra. A forma de selecionar uma definição ao invés de outra dependerá da natureza do problema em questão (LESAGE, 1999, cap.1, p.13).

É importante notar que os elementos da matriz de contigüidade são não-estocásticos e exógenos ao modelo, visto que são baseados em arranjos geográficos das observações (ANSELIN, 1999).

Como uma ilustração concreta, vamos nos basear na definição de contigüidade *rook* para definir a matriz contigüidade de primeira ordem para o exemplo da Figura 2.1. Esta definição é freqüentemente utilizada em trabalhos aplicados e a motivação para isto é que muitas vezes simplesmente desejamos localizar todas as regiões do mapa que apresentam margens em comum. A matriz W em (2.6) mostra as relações de contigüidade de primeira ordem para as cinco regiões da Figura 1.1:

$$W = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix} \quad (2.6)$$

Notamos que W é simétrica e, por convenção, a matriz sempre terá zeros na diagonal principal. Uma modificação freqüentemente utilizada é transformar a matriz W para que ela tenha a soma das suas linhas igual a 1. Uma versão padronizada da matriz W a partir da Equação (2.6) é mostrada a seguir:

$$C = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix} \quad (2.7)$$

A motivação para a padronização pode ser vista considerando o que acontece se multiplicarmos a matriz C por um vetor de observações de alguma variável associada às cinco regiões, que chamaremos de y . Esta matriz produz $y^* = Cy$, que representa uma nova variável igual a média das observações das regiões contíguas (ou vizinhas):

$$\begin{pmatrix} y_1^* \\ y_2^* \\ y_3^* \\ y_4^* \\ y_5^* \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix} \quad (2.8)$$

$$\begin{pmatrix} y_1^* \\ y_2^* \\ y_3^* \\ y_4^* \\ y_5^* \end{pmatrix} = \begin{pmatrix} y_2 \\ y_1 \\ 0.5y_4 + 0.5y_5 \\ 0.5y_3 + 0.5y_5 \\ 0.5y_3 + 0.5y_4 \end{pmatrix}$$

Esta é uma forma de quantificar a noção de que $y_i = f(y_j)$, $j \neq i$, apresentada em (2.1).

A expressão abaixo mostra uma relação linear que usa a variável y^* , a partir de (2.8), como uma variável explicativa a partir de y em observações amostrais espaciais em *cross-section*.

$$y = \rho Cy + \varepsilon \quad . \quad (2.9)$$

O escalar ρ representa um parâmetro da regressão a ser estimado e ε representa os erros estocásticos na relação. O parâmetro ρ reflete a dependência espacial dos dados amostrais, medindo a influência média das observações contíguas ou vizinhas nas observações no vetor y . Além disso, podemos calcular a proporção da variação total em y explicada pela dependência espacial utilizando $\tilde{\rho}Cy$, onde $\tilde{\rho}$ é o valor estimado de ρ .

2.3.2 Quantificando a posição espacial

Outra forma de modelagem econométrica espacial faz uso direto das coordenadas de latitude-longitude associadas às observações espaciais. Muitos métodos tentam lidar com a heterogeneidade espacial usando regressões localmente lineares, que utilizam sub-amostras com o objetivo de fornecer estimativas. Em um modelo que considera que as relações variam no espaço, isto é, que considera heterogeneidade na amostra, um modelo localmente linear fornece uma forma parcimoniosa para estimar múltiplas relações que variam de acordo com a localização espacial das observações (LESAGE, 1999, cap.1, p.16).

Casetti (1992) introduz uma aproximação que consiste em um método denominado "expansão espacial". O modelo é mostrado em (2.10), onde y é um vetor $nx1$ de variáveis dependentes, associadas às observações espaciais e X é uma matriz $nxnk$ em que os termos x_i representam $kx1$ vetores de variáveis explicativas, como mostrado em (2.11).

O modelo estabelece que os parâmetros variem como uma função das coordenadas de latitude e longitude. Os únicos parâmetros que precisam ser estimados são os $2k$ parâmetros em β_0 , que denotamos β_x, β_y . Note que o vetor de parâmetros β em (2.11) representa um vetor $nkx1$ neste modelo contendo os parâmetros estimados para todas k variáveis explicativas em cada observação.

$$y = X\beta + \varepsilon$$

$$\beta = ZJ\beta_0 \tag{2.10}$$

onde

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}; X = \begin{pmatrix} x_1' & 0 & \dots & 0 \\ 0 & x_2' & & \\ \dots & & \dots & \\ 0 & & & x_n' \end{pmatrix}; \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_n \end{pmatrix}; \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix} \quad (2.11)$$

$$Z = \begin{pmatrix} Z_{x1} \otimes I_k & Z_{y1} \otimes I_k & 0 & \dots \\ 0 & \dots & \vdots & \\ \dots & & Z_{xn} \otimes I_k & Z_{yn} \otimes I_k \end{pmatrix}; J = \begin{pmatrix} I_k & 0 \\ 0 & I_k \\ \vdots & \vdots \\ 0 & I_k \end{pmatrix}; \beta_0 = \begin{pmatrix} \beta_x \\ \beta_y \end{pmatrix}$$

Este modelo pode ser estimado por Mínimos Quadrados para obter estimativas dos $2k$ parâmetros em β_x e β_y . Dadas essas estimativas, a estimativa para s pontos individuais no espaço pode ser derivada usando a segunda equação de (2.10). Este processo é chamado de "processo de expansão". Para ver isto, substitua a segunda equação de (2.10) na primeira, de forma que:

$$y = XZJ\beta_0 + \varepsilon \quad (2.12)$$

Em (2.12), X , Z e J representam a informação disponível dos dados amostrais ou as observações amostrais e somente os $2k$ parâmetros de β_0 precisam ser estimados. Este modelo deve capturar a heterogeneidade espacial, permitindo uma variação nas relações entre as observações vizinhas medidas pelas coordenadas de latitude e longitude. A relação de regressão muda de acordo com a variação na localização para acomodar as estimativas localmente lineares através dos grupos de observações próximas umas das outras.

Outra forma de modelar a variação no espaço é baseada em regressões ponderadas pela localização, com o objetivo de produzir estimativas para todo ponto no espaço, utilizando para isto sub-amostras dos dados a partir das observações vizinhas (ou próximas). Os modelos que se baseiam nesta idéia utilizam os métodos de regressão localmente lineares, também chamados regressão ponderada geograficamente (*do inglês*

geographically weighted regression - GWR). Dado y um vetor $n \times 1$ da variável dependente, com as observações coletadas em n pontos no espaço, X uma matriz $n \times k$ de variáveis explicativas e ε um vetor $n \times 1$ de erros normalmente distribuídos com variância constante. W_i representa uma matriz diagonal $n \times n$ dos pesos baseados na distância para a observação i , que reflete a distância entre a observação i e todas as outras observações. Desta forma, podemos escrever o modelo GWR como:

$$W_i y = W_i X \beta_i + \varepsilon_i \quad (2.13)$$

O subscrito i em β_i indica que este vetor $k \times 1$ de parâmetros é associado à observação i . O modelo GWR produz n desses vetores de parâmetros estimados, um para cada observação. Estas estimativas são calculadas usando regressões de mínimos quadrados nas sub-amostras das observações, como mostra a equação (2.14):

$$\hat{\beta}_i = (X' W_i^2 X)^{-1} (X' W_i^2 y) \quad (2.14)$$

Vale destacar que $W_i y$ denota um vetor n de observações ponderadas pelas distâncias, usadas para calcular estimativas para a observação i e não um único elemento de um vetor. Da mesma forma, $W_i X$ representa uma matriz de dados ponderada pelas distâncias e ε_i representa um vetor n .

Os pesos baseados nas distâncias são especificados como uma função de decaimento da distância entre a observação i e todas as outras observações. De forma geral, temos:

$$W_i = f(\theta, d_i) \quad (2.15)$$

O vetor d_i contém as distâncias entre a observação i e todas as outras observações da amostra. O parâmetro θ produz o padrão de decaimento da influência entre as observações de acordo com a distância. Mudando este parâmetro, teremos estimativas

que variam mais ou menos rapidamente no espaço. Note que somente um parâmetro foi introduzido no modelo, de tal forma que, junto com a informação das distâncias, ele produzirá um conjunto de parâmetros estimados para todo ponto na amostra de dados espaciais.

Vale notar aqui que um modelo não apropriado, que falha em capturar heterogeneidade espacial, terá resíduos que exibem dependência espacial. Por isso a preocupação em utilizar modelos dos tipos descritos acima quando trabalhamos com dados distribuídos geograficamente.

2.3.3 Defasagens espaciais

A idéia de operador defasagem espacial está fortemente relacionada ao conceito de contigüidade espacial. As defasagens espaciais são análogas ao operador defasagens B da análise em séries temporais. Este operador desloca as observações no tempo, onde $B^p y_t = y_{t-p}$ representa a p -ésima ordem de defasagem. Analogamente ao domínio do tempo, o operador de defasagem espacial implica em um deslocamento no domínio do espaço. No entanto, estes sofrem algumas complicações, surgidas quando tentamos fazer a analogia com o domínio do tempo. Isto acontece porque, no domínio do espaço, as observações geralmente estão distribuídas de uma forma irregular.

O operador de defasagem espacial funciona no contexto de produzir uma média ponderada das observações vizinhas. Vimos na seção 2.3.1 que o conceito de vizinhança - ou contigüidade - depende da definição usada. Por analogia à análise de séries de tempo, parece razoável montarmos nossa matriz binária W para as relações de contigüidade de primeira ordem, contendo valores de 0 ou 1 para criar nosso operador defasagem p . No entanto, levar em consideração só as relações de contigüidade de primeira ordem na forma de uma matriz W binária pode ser problemático durante a estimação por máxima verossimilhança (ANSELIN; SMIRNOV, 1994 *apud* LESAGE, 1999).

2.4 Modelos de Regressão Espacial

Nesta seção serão apresentadas as formas mais comuns de tratar o componente espacial nos modelos econométricos. Para isto, esta seção está dividida em três subseções, onde a primeira trata de modelos de defasagens espaciais, a segunda considera os modelos que tratam da autocorrelação espacial e a terceira trata dos modelos de regressão espacial, cujos modelos são bastante parecidos com os modelos tratados pela econometria convencional, mas, no entanto, são modificados com o intuito de considerar a dependência espacial como um regressor.

2.4.1 Defasagens Espaciais

O operador de defasagem espacial⁶ é usado no lugar da noção de deslocamento temporal utilizado nos métodos convencionais para análise em séries de tempo. Nesta configuração de modelo, a definição de um conjunto de vizinhança para cada localização é essencial. Como apresentado na Seção 2.3, este conjunto é obtido através da especificação da matriz de contigüidade. Especificando cada localização i como uma linha, e os vizinhos de i como, correspondendo aos elementos w_{ij} , diferentes de zero em uma matriz fixa de pesos espaciais, de dimensão $n \times n$, chamada matriz contigüidade⁷ W . Então, uma defasagem espacial para y na localização i toma a forma:

$$(Wy)_i = \sum_{j=1, \dots, N} w_{ij} \cdot y_j \quad (2.16)$$

ou na forma matricial, como

⁶ O operador de defasagem espacial é uma média ponderada das variáveis aleatórias a partir das localizações vizinhas, como definido em Anselin (1999).

⁷ A matriz contigüidade também pode ser chamada de matriz peso espacial.

$$Wy \quad (2.18)$$

onde y é um vetor $n \times 1$ das observações na variável aleatória. Desde que para cada i a matriz de elementos w_{ij} são diferentes de zero para aqueles $j \in S_i$ (onde S_i é o conjunto vizinhança), somente os y_j são incluídos na defasagem. Como já apresentado, os elementos da matriz contigüidade são geralmente padronizados pelas linhas, tal que para cada i , $\sum_j w_{ij} = 1$. Conseqüentemente, a defasagem espacial pode ser interpretada como uma média ponderada das observações vizinhas (com w_{ij} agindo como os pesos)⁸.

2.4.2 Expressões formais para autocorrelação espacial

Nesta subseção, são apresentadas três formas distintas de considerar a autocorrelação espacial nos modelos de regressão. A primeira é através dos processos estocásticos, que é a forma mais comum e mais utilizada. A segunda é a representação direta, também bastante conhecida, em que se leva em consideração a representação parcimoniosa da matriz de variância-covariância e a terceira é a aproximação não-paramétrica, que considera a autocorrelação espacial de forma indireta, isto é, torna não explícito a consideração do processo espacial.

2.4.2.1 Processos estocásticos Espaciais

A forma mais comum de expressar a autocorrelação espacial é através da especificação de um processo estocástico espacial, isto é, uma relação funcional entre uma variável aleatória numa determinada localização e a mesma variável aleatória em outras

⁸ É importante notar que os elementos da matriz peso (ou matriz contigüidade) são não estocásticos e exógenos ao modelo. Geralmente eles são baseados na distribuição geográfica dos dados. A noção apresentada aqui, retomada da seção 2.3, é bastante geral e algumas das diferentes formas para a composição da matriz contigüidade já foram apresentadas.

localizações. A estrutura da covariância então se apresenta a partir da natureza do processo. Por exemplo, dada uma matriz W de contiguidade de dimensão $n \times n$, um vetor y $n \times 1$ de variáveis aleatórias, e um vetor ε $n \times 1$ de erros aleatórios i.i.d, um processo autorregressivo espacial simultâneo é definido como

$$(\mathbf{y} - \mu\mathbf{i}) = \rho\mathbf{W}(\mathbf{y} - \mu\mathbf{i}) + \varepsilon \quad \text{ou} \quad (\mathbf{y} - \mu\mathbf{i}) = (\mathbf{I} - \rho\mathbf{W})^{-1}\varepsilon \quad , \quad (2.19)$$

enquanto um processo de médias móveis espacial como

$$\mathbf{y} = \lambda\mathbf{W}\varepsilon + \varepsilon \quad \text{ou} \quad \mathbf{y} = (\mathbf{I} + \lambda\mathbf{W})\varepsilon \quad , \quad (2.20)$$

onde \mathbf{I} é uma matriz identidade $n \times n$, \mathbf{i} é um vetor $n \times 1$ de 1's, μ é a média comum das variáveis aleatórias y_i , os termos i.i.d, de média zero, ε_i têm variância comum σ^2 , e ρ e λ são, respectivamente, os parâmetros autorregressivo e de médias móveis. Como na análise em séries de tempo, também é possível escrever um processo autorregressivo espacial como um processo de médias-móvel espacial infinito, uma vez que os elementos de \mathbf{W} são geralmente menores que um e, além disso, na maioria dos casos $|\rho| < 1$, tal que:

$$(\mathbf{y} - \mu\mathbf{i}) = \mathbf{I}\varepsilon + \rho\mathbf{W}\varepsilon + \rho^2\mathbf{W}^2\varepsilon + \dots \quad . \quad (2.21)$$

As restrições impostas pela estrutura de pesos, além da forma específica do processo espacial, determinam a matriz de variâncias-covariâncias para \mathbf{y} como uma função de dois parâmetros: a variância σ^2 e o coeficiente espacial, ρ ou λ . Para a estrutura SAR em (2.19), assumindo $E(\mathbf{y} - \mu\mathbf{i}) = 0$, isto fica:

$$\text{cov}[(\mathbf{y} - \mu\mathbf{i}), (\mathbf{y} - \mu\mathbf{i})] = E[(\mathbf{y} - \mu\mathbf{i})(\mathbf{y} - \mu\mathbf{i})'] = \sigma^2 [(\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \rho\mathbf{W})]^{-1} \quad . \quad (2.22)$$

Esta é uma matriz completa, o que implica que um choque em qualquer localização afeta todas as localizações através do efeito multiplicador espacial⁹. Em contraste, para um processo SMA, assumindo $E(\mathbf{y}) = 0$, a covariância será:

$$\text{cov}(\mathbf{y}, \mathbf{y}) = E(\mathbf{y}\mathbf{y}') = \sigma^2(\mathbf{I} + \rho\mathbf{W})(\mathbf{I} + \rho\mathbf{W})' = \sigma^2[\mathbf{I} + \rho(\mathbf{W} + \mathbf{W}') + \rho^2\mathbf{W}\mathbf{W}'], \quad (2.23)$$

o que resulta em uma interação local entre uma determinada localização e seus vizinhos de primeira e segunda ordem (através de \mathbf{W} e $\mathbf{W}\mathbf{W}'$, respectivamente), mas não gera um multiplicador espacial como no modelo SAR.

A principal distinção entre os processos AR e MA no espaço, comparados ao domínio do tempo, é que mesmo que os termos referentes ao erro ε_i sejam i.i.d, os elementos das diagonais de (2.22) e (2.23) não serão constantes. Mais ainda, a heterocedasticidade depende da estrutura de vizinhança inserida na matriz contigüidade \mathbf{W} . Conseqüentemente, o processo em \mathbf{y} não é estacionário em covariância.

2.4.2.2 Representação direta

Outra forma bastante usada para a especificação formal da autocorrelação espacial é apresentar os elementos da matriz de variâncias-covariâncias numa forma parcimoniosa, como uma função direta de um número pequeno de parâmetros e uma ou mais variáveis exógenas. Geralmente, isto envolve uma função inversa a partir de uma distância métrica, por exemplo:

$$\text{cov}(\varepsilon_i, \varepsilon_j) = \sigma^2 \mathbf{f}(\mathbf{d}_{ij}, \varphi) \quad , \quad (2.24)$$

onde ε_i e ε_j são os termos dos erros da regressão, σ^2 é a variância dos erros, \mathbf{d}_{ij} é a

⁹ Também chamado de efeito de interação global (ver ANSELIN, 1999).

distância que separa as localizações i e j , e \mathbf{f} é a função de decaimento da distância tal que $\frac{\partial \mathbf{f}}{\partial \mathbf{d}} < 0$ e $|\mathbf{f}(\mathbf{d}_{ij}, \varphi)| \leq 1$, com $\varphi \in \Phi$ como um vetor $p \times 1$ de parâmetros em um subconjunto aberto Φ de \mathbf{R}^p . Usando (2.24) para os elementos individuais, a matriz completa de covariâncias dos erros segue como

$$E(\varepsilon\varepsilon') = \sigma^2 \Omega(\mathbf{d}_{ij}, \varphi) \quad , \quad (2.25)$$

onde, considerando o fator escalar σ^2 , a matriz $\Omega(\mathbf{d}_{ij}, \varphi)$ precisa ser uma matriz positiva definida de correlação espacial, com $\omega_{ii} = 1$ e $|\omega_{ij}| \leq 1, \forall \mathbf{i}, \mathbf{j}$ ¹⁰. Note que o modelo de representação direta não induz à heterocedasticidade.

2.4.2.3 Aproximação não-paramétrica

Uma especificação alternativa recentemente utilizada da matriz de covariância espacial é estimar seus elementos de uma forma não-paramétrica, de maneira que não seja necessário um processo espacial explícito ou uma forma funcional que descreva o decaimento da distância. Isto é comum no caso de dados de painel, quando a dimensão do tempo é consideravelmente maior que a dimensão da *cross-section* e a covariância espacial é estimada a partir da covariância dos resíduos da amostra¹¹.

2.4.3 Tipos de Modelos de Regressão Espacial¹²

Nesta subseção são apresentados os modelos de regressão espaciais mais comuns na

¹⁰ $\omega_{ii} = 1$ assegura que, selecionando uma forma funcional para \mathbf{f} tal que $\mathbf{f}(\mathbf{d}_{ij}, \varphi) = 1$ para $\mathbf{d}_{ij} = 0$.

¹¹ Para mais detalhes e referências de aplicações deste método, ver Anselin (1999, p. 9).

¹² Esta seção é baseada na seção 3, apresentada em Anselin (1999).

literatura. Estes modelos se parecem muito com os modelos lineares convencionais, mas, no entanto, são modificados com o intuito de considerar os efeitos espaciais.

2.4.3.1 Modelo de defasagem espacial e modelo de erro espacial

Em modelos de regressão linear convencional, a dependência espacial pode ser incorporada de duas formas distintas: como um regressor adicional na forma de uma variável dependente espacialmente defasada ($\mathbf{W}y$), ou na estrutura dos erros - $E(\varepsilon_i \varepsilon_j) \neq 0$. A primeira é chamada de modelo de defasagem espacial e é apropriada quando o foco de interesse está em avaliar a existência e a força da interação espacial. Isto é interpretado como uma dependência espacial substantiva de maneira em que está diretamente relacionada ao modelo espacial. A dependência espacial nos termos dos erros da regressão, ou modelo de erro espacial, pode tomar qualquer configuração exposta na seção 2.4.2 quando o interesse está na correção do viés provocado pela autocorrelação espacial, provocada pelos dados espaciais.

Formalmente, um modelo de defasagem espacial, ou um modelo regressivo misto, pode ser apresentado como:

$$\mathbf{y} = \rho \mathbf{W}y + \mathbf{X}\beta + \varepsilon \quad , \quad (2.26)$$

onde ρ é um vetor de coeficientes autorregressivos espaciais, ε é o vetor dos erros¹³, \mathbf{X} é a matriz de variáveis explicativas e β é o vetor dos coeficientes que não autorregressivos espaciais. Ao contrário do que acontece com a reprodução deste modelo no contexto de séries temporais, o termo que inclui as defasagens espaciais, $\mathbf{W}y$ é correlacionado com os erros, mesmo quando estes últimos são i.i.d. Isto pode ser visto a partir da forma reduzida de (2.26):

¹³ Para facilitar a exposição, é assumido que os erros são i.i.d. No entanto, várias formas de heterocedasticidade podem ser incorporadas.

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\beta + (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon} \quad , \quad (2.27)$$

em que cada inversa pode ser expandida em uma série infinita, da mesma forma que em (2.21), incluindo tanto as variáveis explicativas como os termos dos erros em todas as localizações. Conseqüentemente, o termo de defasagem espacial precisa ser tratado como uma variável endógena e, portanto os métodos de estimação precisam considerar esta endogeneidade¹⁴.

O modelo de erro espacial é um caso especial de regressão em que os elementos fora da diagonal principal da matriz de covariância expressam a estrutura de dependência espacial. Como descrita na Seção 2.4.2, a estrutura espacial pode assumir diferentes formas e resulta em uma matriz de variância-covariância dos erros na seguinte forma¹⁵:

$$\mathbf{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Omega}(\boldsymbol{\theta}) \quad (2.28)$$

onde $\boldsymbol{\theta}$ é um vetor de parâmetros, assim como os coeficientes num processo de erros SAR ou SMA. Quando os erros seguem um processo SAR, o modelo resultante pode ser apresentado como uma especificação de defasagem espacial. O modelo em que os erros seguem um processo SAR é

$$\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\varepsilon} \text{ e } \boldsymbol{\varepsilon} = \lambda\mathbf{W}\boldsymbol{\varepsilon} + \mathbf{u} \quad , \quad (2.29)$$

uma vez que $\boldsymbol{\varepsilon} = (\mathbf{I} - \lambda\mathbf{W})^{-1}\mathbf{u}$ e, portanto $\mathbf{y} = \mathbf{X}\beta + (\mathbf{I} - \lambda\mathbf{W})^{-1}\mathbf{u}$, (2.29) é equivalente a

$$\mathbf{y} = \lambda\mathbf{W}\mathbf{y} + \mathbf{X}\beta - \lambda\mathbf{W}\mathbf{X}\beta + \boldsymbol{\varepsilon} \quad , \quad (2.30)$$

que é um modelo de defasagem espacial com um conjunto adicional de variáveis exógenas espacialmente defasadas ($\mathbf{W}\mathbf{X}$) e um conjunto de k restrições não lineares nos

¹⁴ O método OLS levará a estimativas viesadas e inconsistentes neste caso devido ao viés de simultaneidade.

¹⁵ Isto não ocorre na aproximação não-paramétrica.

coeficientes (o produto entre o coeficiente autorregressivo espacial e os coeficientes da regressão, β , devem ser iguais ao negativo dos coeficientes de \mathbf{WX}). A similaridade entre o modelo de erro espacial (2.30) e o modelo de defasagem espacial “puro” (2.26), irá complicar os testes de especificação na prática, uma vez que estes testes modificados para a abordagem dos modelos de defasagens espaciais também considerarão os erros espaciais.

2.4.3.2 Dependência espacial em Modelos de dados em painel

Quando as observações estão disponíveis tanto no espaço quanto no tempo, a dimensão adicional permite a estimação da matriz completa de covariância, usando esta dimensão adicional para ajudar a obter as propriedades assintóticas. No entanto, para um caso em que os dados se apresentam somente em *cross-section*, existe informação insuficiente nas observações para estimar a matriz de covariância completa¹⁶ $\text{cov}(y_{it}y_{js}) \neq 0$, (com $i \neq j$ e $t \neq s$), sem impor qualquer estrutura.

Quando ambas as dependências, temporal e espacial, podem ser parametrizadas, algumas especificações podem ser consideradas, permitindo diferentes combinações destas duas formas de parametrização. Para facilitar a exposição, assumamos que as observações estão organizadas por período do tempo. Focando a atenção para a dependência defasada, e com $f(z)$ como uma forma geral para os regressores (que podem ser defasados no tempo e/ou no espaço), podemos ter quatro modelos distintos:

(a) Espaço recursivo puro: em que a dependência está ligada às localizações vizinhas em diferentes períodos de tempo, ou:

$$\mathbf{y}_{it} = \gamma[\mathbf{W}\mathbf{y}_{t-1}]_i + \mathbf{f}(\mathbf{z}) + \varepsilon_{it} \quad (2.31)$$

onde $[\mathbf{W}\mathbf{y}_{t-1}]_i$ é o *i-ésimo* elemento do vetor de defasagem espacial aplicado às

¹⁶ Por exemplo, n localizações apresentadas num espaço de tempo t , existirá nt observações para estimar uma matriz de covariância de dimensão $(nt)^2$.

observações da variável dependente no período de tempo anterior (usando uma matriz contigüidade $n \times n$ para as unidades em *cross-section*).

(b) Espaço-tempo recursivo: em que a dependência está ligada a mesma localização assim como às localizações vizinhas em diferentes períodos de tempo, ou:

$$\mathbf{y}_{it} = \lambda \mathbf{y}_{it-1} + \gamma [\mathbf{W}\mathbf{y}_{t-1}]_i + \mathbf{f}(\mathbf{z}) + \varepsilon_{it} \quad (2.32)$$

(c) Espaço-tempo simultâneo: com variáveis dependentes tanto na forma tempo-discreta como na forma espacialmente defasada, ou:

$$\mathbf{y}_{it} = \lambda \mathbf{y}_{it-1} + \rho [\mathbf{W}\mathbf{y}_t]_i + \mathbf{f}(\mathbf{z}) + \varepsilon_{it} \quad (2.33)$$

onde $[\mathbf{W}\mathbf{y}_t]_i$ é o i -ésimo elemento do vetor espacialmente defasado no mesmo período de tempo.

(d) Espaço-tempo dinâmico: que considera todas as formas de dependência:

$$\mathbf{y}_{it} = \lambda \mathbf{y}_{it-1} + \rho [\mathbf{W}\mathbf{y}_t]_i + \gamma [\mathbf{W}\mathbf{y}_{t-1}]_i + \mathbf{f}(\mathbf{z}) + \varepsilon_{it} \quad (2.34)$$

Com o objetivo de estimar os parâmetros do modelo espaço-tempo simultâneo, é necessário assumir assintoticidade na dimensão da *cross-section* enquanto que para o modelo espaço-tempo dinâmico esta suposição é necessária em ambas as dimensões.

Um importante tópico a ser considerado quando os efeitos espaciais são incorporados em modelos para dados em painel é a extensão dos modelos em que são permitidos efeitos fixos. Uma vez que a estimação dos modelos de processo espacial precisa das suposições de assintoticidade ($n \rightarrow \infty$), os efeitos fixos (isto é, uma variável *dummy* para cada localização) podem sofrer com problemas de estimadores não consistentes. Então, estes efeitos fixos serão incompatíveis com os processos espaciais e, no lugar destes, devem ser consideradas especificações de efeitos aleatórios.

2.4.3.3 O Modelo VAR Bayesiano com informação *a priori* espacial

O Modelo VAR apresenta um conjunto de relações entre valores passados defasados de todas as variáveis no modelo e o valor corrente de cada variável no modelo. Por exemplo, se y_{it} representa emprego na região i no instante t , a estrutura do modelo VAR permite que a variação do emprego em cada região possa ser explicada pela variação passada do emprego naquela mesma região, y_{it-k} , $k=1, 2, \dots, m$, e também pela variação passada do emprego em outras regiões, y_{jt-k} , $k=1, 2, \dots, m$, $i \neq j$. Isto é atrativo uma vez que regiões têm diferenças em atividades de ciclos de negócios que sugerem relações defasadas em emprego num conjunto de tipos estabelecidos pela estrutura do modelo VAR.

De forma geral, as n variáveis refletem as séries de tempo de n regiões e a estrutura do VAR estabelece um conjunto de relações entre os valores passados defasados de todas as localizações no modelo e o valor corrente de cada localização. Este modelo pode ser estimado por mínimos quadrados ordinários.

Vale observar que o modelo descrito acima pode ser visto como o modelo VAR convencional. No entanto, este modelo leva em consideração a influência espacial uma vez que tratam de emprego em n localizações distintas. Este modelo é um modelo espacial simplificado uma vez que considera somente uma característica para diferentes regiões, no caso y , que denota emprego.

No modelo VAR espacial, os autovalores são determinados tanto pelas defasagens temporais quanto pelas defasagens espaciais. A estacionaridade depende tanto da estrutura de defasagem espacial como também da estrutura de defasagem temporal.

Em alguns casos, a estimação do modelo VAR espacial pode ser superparametrizado. Isto ocorre porque, no modelo VAR espacial, são estimados muitos parâmetros. Por exemplo, num modelo VAR com 8 variáveis e 6 defasagens produz 49 variáveis independentes (8×6 mais uma constante), em cada uma das oito equações do modelo e, portanto, 392 coeficientes a serem estimados, onde uma grande maioria destes serão estatisticamente não significativo. É necessária uma amostra com um grande número de observações, envolvendo uma série de tempo longa, para estimar o modelo, e esta nem sempre está disponível. LeSage e Krivelyova (1999) contornam este problema

implementando um modelo VAR com imposição de restrições *a priori* que estabelecem que os parâmetros correspondentes a defasagens de ordens mais elevadas e regiões não-vizinhas são iguais a zero. Em experimentos de previsão detalhados, esta aproximação baseada na *a priori* espacial gera previsões mais precisas quando comparada à aproximação do VAR bayesiano baseada na chamada *a priori* de Minnesota (DOAN; LITTELMAN; SIMS, 1984).

Um princípio por trás da modelagem na ciência regional é que a localização no espaço é importante. LeSage e Krivelyova (1999) incorporam esse princípio no modelo VAR na forma de informação *a priori*. Esta informação *a priori* é aplicada aos coeficientes do modelo VAR. A descrição detalhada do modelo VAR Bayesiano utilizando informação espacial *a priori* e informação *a priori* de Minnesota será apresentada no Capítulo 4 do presente trabalho e será foco de interesse de aplicação, apresentada no Capítulo 5.

3 A Abordagem Bayesiana

A Estatística é a ciência que estuda a incerteza. A visão Bayesiana interpreta "probabilidade" da mesma forma que a linguagem coloquial, ou seja, como uma medida de incerteza, ou grau de crença, associada à ocorrência de um determinado evento incerto, dada uma informação disponível e algumas suposições aceitáveis. O grau de crença pode ser operacionalizado em termos de probabilidades. A Estatística Bayesiana descreve como um indivíduo deveria agir para evitar inconsistências indesejáveis quando está diante destas incertezas. A maximização da utilidade esperada (ou minimização da perda) fornece uma base para a tomada de decisão racional e o Teorema de Bayes¹⁷ descreve como as crenças se atualizam na medida em que os dados são obtidos. A Econometria Bayesiana consiste dos métodos e modelos da Estatística Bayesiana, aplicados aos fenômenos econômicos (POIRIER; TOBIAS, 2004). Então, uma das principais vantagens da abordagem Bayesiana é que ela é baseada em poucas regras simples de probabilidade. Nas questões abordadas pela econometria, como estimar parâmetros de um modelo, comparar diferentes modelos ou mesmo obter previsões, envolvem as mesmas regras de probabilidade (KOOP, 2004, cap.1, p.1).

A essência do método Bayesiano pode ser descrita quando pensamos nas opiniões expressas como probabilidades. Essas probabilidades *a priori* mudam através da aplicação do Teorema de Bayes, à medida que os dados são coletados, resultando nas probabilidades *a posteriori* (WITHERS, 2002).

Em problemas de tomada de decisão, o uso das probabilidades refletindo as crenças pessoais *a priori* distingue a abordagem bayesiana das demais. O conceito de probabilidade reflete a noção intuitiva apresentada nos problemas. Esta medida de probabilidade geralmente usada no processo de decisão depende do indivíduo e de seu

¹⁷ Em 1763, Bayes publicou um ensaio dirigido a resolver problemas que envolviam probabilidades (*An essay towards solving a problem in the doctrine of chances*) em que descreve o que é conhecido como Teorema de Bayes, (Whithers, 2002), que será apresentado nas seções seguintes e que deu início ao interesse e desenvolvimento de pesquisa nessa área. O Teorema de Bayes fornece uma técnica de rever as probabilidades à luz de novas evidências.

grau de crença, isto é, do que ele julga mais ou menos provável dentro de um determinado contexto e baseado nas informações disponíveis a ele.

A interpretação subjetiva da probabilidade é baseada na avaliação pessoal individual de uma situação. Conseqüentemente, a probabilidade é uma propriedade da percepção individual da realidade, e não uma propriedade da realidade por ela mesma. Para aqueles que utilizam o método subjetivo, não existem "probabilidades desconhecidas verdadeiras" no mundo a ser analisado. Ao invés disso, a probabilidade está "nos olhos do contemplador" (POIRIER; TOBIAS, 2004).

3.1 Análise Econométrica

Uma análise econométrica é o confronto de um modelo econômico com a evidência. Um modelo econômico descreve uma relação entre variáveis econômicas. Geralmente, qualquer modelo contém quantidades observáveis e quantidades não diretamente observáveis¹⁸. Estas últimas são chamadas parâmetros e serão denotadas genericamente por θ ¹⁹. Eles geralmente estão restritos a pertencer a um conjunto denotado por Θ . Qualquer valor particular para θ define uma estrutura particular para o modelo e um conjunto de estruturas particulares é dito indexado ao parâmetro θ . A evidência é fornecida pelos dados observados em um contexto econômico.

O objetivo de uma análise econométrica é responder a duas questões. A primeira questão é a consistência do modelo com a evidência, chamada de crítica do modelo. Tal preocupação equivale a questionar se alguma das estruturas definidas pelo modelo é consistente com a evidência. A segunda questão supõe a existência de alguma estrutura consistente com a evidência, questionando quais são as probabilidades das diferentes estruturas definidas pelo modelo. Uma vez que esta questão seja respondida, o modelo

¹⁸ Por exemplo, considere o modelo $C = \alpha + \beta Y$, onde C (consumo) é linearmente relacionado com Y (renda). Neste caso, as quantidades observáveis (dados) são dadas por C e Y , e a relação entre elas determina os valores para α e β , que são as quantidades não observáveis, também chamadas parâmetros do modelo.

¹⁹ Em nosso caso, $\theta = (\alpha, \beta)$, e o conjunto Θ pode ser tomado como o espaço bidimensional euclidiano.

pode então ser usado com o objetivo de tomar decisões econômicas (LANCASTER, 2004, cap.1, p.2).

A prática dos econométricos é abordar estas duas questões na ordem inversa. Começamos presumindo que nosso modelo é consistente com os dados e investigamos qual é a estrutura mais provável à luz das evidências. Na econometria tradicional isto consiste em formar um bom estimador $\hat{\theta}_0 \in \Theta$, uma estrutura particular que é assumida, de alguma forma, verdadeira. Na análise Bayesiana, este passo consiste em usar os dados para formar uma distribuição de probabilidade *a posteriori* para Θ .

3.2 Uma noção intuitiva para a abordagem Bayesiana

Para iniciar a apresentação da noção intuitiva da abordagem Bayesiana, considere duas variáveis aleatórias, A e B. Pelas regras de probabilidade, temos

$$p(A,B) = p(A|B)p(B) \quad (3.1)$$

onde $p(A,B)$ é a probabilidade conjunta de A e B ocorrerem, $p(A|B)$ é a probabilidade condicional de A ocorrer dado que B tenha ocorrido (i.e, a probabilidade condicional de A dado B) e $p(B)$ é a probabilidade marginal de B. Alternativamente, podemos inverter a ordem de A e B na expressão e encontrar a especificação abaixo para a probabilidade conjunta entre A e B:

$$p(A,B) = p(B|A)p(A) \quad (3.2)$$

Igualando as duas expressões para $p(A, B)$ e reorganizando a expressão, temos:

$$p(B|A) = \frac{p(A,B)p(B)}{p(A)} \quad (3.3)$$

Quando a teoria da probabilidade é aplicada, como ela é na econometria, precisamos decidir como interpretar a probabilidade de um evento. A econometria se concentra na utilização de dados sobre um determinado fato em questão. No entanto, nestas circunstâncias, trabalha-se com modelos que dependem de parâmetros. Vamos tomar como exemplo um modelo de regressão. Em um modelo de regressão o foco de interesse está na estimação dos coeficientes da regressão. Neste caso os coeficientes são os parâmetros em estudo (KOOP, 2004, cap.1, p.2). Considere y sendo um vetor ou uma matriz de dados e θ sendo um vetor ou uma matriz contendo os parâmetros para o modelo que visa explicar y . Estamos interessados em aprender sobre θ , tomando como base os dados, y . A econometria Bayesiana utiliza o Teorema de Bayes²⁰ para fazer isto. Em outras palavras, substituindo B por θ e A por y em (3.3), temos:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} . \quad (3.4)$$

Nesta forma, o Teorema de Bayes mostra como uma crença a respeito dos parâmetros θ , medida como $p(\theta)$, é trocada por outra crença a respeito de θ , medida por $p(\theta|y)$. Assim, temos inicialmente uma afirmação sobre os parâmetros θ de um modelo econômico. Posteriormente, um conjunto de dados y é obtido, θ permitindo-nos, com base na evidência fornecida pelos dados, a atualização de nossa afirmação sobre θ .

Desta forma, podemos interpretar $p(\theta)$ como a distribuição de probabilidades de θ antes que os dados tenham sido observados e $p(\theta|y)$ é a sua distribuição depois dos dados terem sido observados. Quando pensamos desta forma, chamamos $p(\theta)$ de probabilidade *a priori* sobre θ e $p(\theta|y)$ de probabilidade *a posteriori* sobre θ . O Teorema de Bayes pode então ser interpretado como a forma de mostrar a maneira de revisar as crenças sobre os parâmetros a luz das evidências (LANCASTER, 2004, cap.1, p.4).

²⁰ O Teorema de Bayes será apresentado formalmente na seção 2.3 deste mesmo capítulo, de acordo com a forma apresentada por James (1981, cap. 1), *Probabilidade: Um Curso em Nível Intermediário*, cap.1.

A forma mais usual de pensar sobre a diferença entre o parâmetro θ e os dados, y , é que o parâmetro representa uma quantidade desconhecida, tanto antes quanto depois que os dados foram observados, embora suas crenças sobre estes geralmente (mas não necessariamente) possam ser mudadas pela evidência; enquanto os dados são desconhecidos antes de serem coletados, mas conhecidos depois²¹ (LANCASTER, 2004, cap.1, p.8).

A visão Bayesiana é baseada numa visão subjetiva de probabilidade, que argumenta que a incerteza sobre qualquer coisa pode ser expressa usando as regras de probabilidade. Desta forma, θ é tratado como variável aleatória (KOOP, 2004, cap.1, p.2). Por exemplo, suponha X, Y um par de variáveis aleatórias definidas no espaço amostral Ω , que assumem densidade de probabilidade conjunta $p_{X,Y}(x,y)$, com densidades marginais, $f_X(x)$, $f_Y(y)$ e densidades condicionais $p_{X|Y}(x|y)$ e $p_{Y|X}(y|x)$ – usando aqui notação p para a função densidade de probabilidade. Então, o teorema é

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)p_X(x)}{p_Y(y)} \quad (3.5)$$

Nesta notação, os subscritos indicam as variáveis aleatórias e os argumentos indicam os valores particulares delas. Ocultando os subscritos, a equação se torna

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} \quad (3.6)$$

e, quando usada para inferência sobre os parâmetros dado o conjunto de dados, é convencional escrever a equação como a forma apresentada em (3.4). Note que os parâmetros e os dados são tratados de uma forma simétrica antes que os dados tenham sido observados e estão associados a uma distribuição de probabilidade conjunta (LANCASTER, 2004, cap.1, p.7).

²¹ Os parâmetros também podem se referir as funções, e também como constantes, em uma configuração onde é proposto que $y=g(x)$, onde y e x são duas variáveis econômicas. Se $g(\cdot)$ não é dada nem conhecida e, portanto não é um dado, então esta função é um parâmetro.

A partir da equação (3.4), temos que $p(\theta|y)$ é foco do objetivo dos econométristas. Enquanto o interesse baseia-se em θ , podemos ignorar o termo $p(y)$ da equação (3.4), uma vez que ele é independente de θ . Podemos, então, escrever:

$$p(\theta|y) \propto p(y|\theta) \cdot p(\theta) \quad (3.7)$$

O termo $p(\theta|y)$ é chamado de densidade *a posteriori*, enquanto a função densidade de probabilidade (p.d.f.) para os dados, condicional aos parâmetros do modelo, $p(y|\theta)$, é denominada função verossimilhança e $p(\theta)$ é a densidade *a priori*. O lado direito da equação será, portanto, a distribuição de probabilidade conjunta entre os dados (variáveis aleatórias observáveis) e os parâmetros (variáveis aleatórias não-observáveis). Lemos esta equação como “a distribuição *a posteriori* é proporcional a verossimilhança vezes *a priori*” (KOOB, 2004, cap.1, p.2).

A densidade *a priori*, $p(\theta)$, não depende dos dados. Ela contém qualquer informação disponível sobre θ , desde que esta informação não seja proveniente da amostra de dados. Em outras palavras, ela resume o que sabemos sobre os parâmetros antes de tomarmos uma amostra de dados. A função de verossimilhança, $p(y|\theta)$, é a densidade condicional dos dados, dados os parâmetros do modelo. É frequentemente chamada de processo geracional dos dados. A densidade *a posteriori*, $p(\theta|y)$, é o nosso foco de interesse. Ela resume toda a informação sobre os parâmetros θ depois de observar os dados (KOOB, 2004, cap.1, p.3).

Na Seção 3.4 será apresentado cada um dos componentes do Teorema de Bayes numa forma mais completa, mas, antes disto, será apresentado formalmente o Teorema de Bayes, com base nos conceitos de probabilidade, apresentados em James (1981).

3.3 O Teorema de Bayes

Para chegar à fórmula apresentada na Equação (3.4), é importante ter em mente algumas definições sobre probabilidade, uma vez que o Teorema de Bayes é derivado a partir

destas definições. Para tanto, nesta seção serão apresentadas as definições necessárias para então chegar ao Teorema de Bayes na sua forma tradicional.

3.3.1 Conceitos de Probabilidade

Um modelo matemático para um experimento, ou um modelo probabilístico, é constituído de (JAMES, 1981, Cap.1, p.13)

- (a) Um conjunto não-vazio Ω , de resultados possíveis, o *espaço amostral*.
- (b) Uma σ -álgebra \mathcal{A} de *eventos aleatórios*.
- (c) Uma *probabilidade* P definida em \mathcal{A} .

Para entender tal constituição, vamos olhar para algumas definições fundamentais da teoria da probabilidade, e assim entender do que consiste um modelo probabilístico.

DEFINIÇÃO 3.1. Um evento A ao qual atribuímos uma probabilidade será chamado de *evento aleatório* (JAMES, 1981, cap.1, p.5).

Indicando como \mathcal{A} a classe de eventos aleatórios, as seguintes propriedades são válidas para \mathcal{A} :

A.1. $\Omega \in \mathcal{A}$ e $P(\Omega) = 1$, onde Ω é o espaço amostral do experimento.

A.2. Se $A \in \mathcal{A}$, então $A^c \in \mathcal{A}$ e, então, $P(A^c) = 1 - P(A)$.

A.3. Se $A \in \mathcal{A}$ e $B \in \mathcal{A}$, então $A \cup B \in \mathcal{A}$ (ou seja, se atribuímos uma probabilidade a A e outra probabilidade a B , então atribuímos uma probabilidade a “ A ou B ”).

A.3'. Se $A_n \in \mathcal{A}$ para $n = 1, 2, 3, \dots$, então $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Assumindo que A.1, A.2 e A.3 sejam satisfeitas, temos então que \mathcal{A} é uma álgebra de eventos, o que leva a seguinte definição:

DEFINIÇÃO 3.2. Seja Ω um conjunto não-vazio. Uma classe \mathcal{A} de subconjuntos de Ω satisfazendo A.1, A.2 e A.3 é chamada *álgebra* de subconjuntos de Ω (JAMES, 1981, cap.1, p.5).

Proposição 2.1²². Seja \mathcal{A} uma álgebra de subconjuntos de Ω . Então valem as seguintes propriedades:

A.4. $\emptyset \in \mathcal{A}$

A.5. $\forall n, \forall A_1, \dots, A_n \in \mathcal{A}$, temos $\bigcup_{i=1}^n A_i \in \mathcal{A}$ e $\bigcap_{i=1}^n A_i \in \mathcal{A}$.

Esta proposição significa que uma álgebra é fechada para um número finito de aplicações das operações \cup , \cap e c .

DEFINIÇÃO 3.3. Uma classe \mathcal{A} de subconjuntos de um conjunto não-vazio Ω satisfazendo A.1, A.2 e A.3' é chamada *σ -álgebra* de subconjuntos de Ω (JAMES, 1981, cap.1, p.6)²³.

Sem nos preocupar em como definir probabilidade para cada experimento, vamos supor somente que elas existem em certa σ -álgebra \mathcal{A} de eventos, chamados eventos aleatórios. Vamos supor também que a todo $A \in \mathcal{A}$ seja associado um número real $P(A)$, chamado probabilidade de A, de forma que os axiomas a seguir sejam satisfeitos (JAMES, 1981, cap.1, pp. 9-11).

AXIOMA 1. $P(A) \geq 0$.

AXIOMA 2. $P(\Omega) = 1$.

AXIOMA 3. (Aditividade finita). Se $A_1, \dots, A_n \in \mathcal{A}$ são disjuntos (2 a 2), então $P(\bigcup_{k=1}^n A_k) = \sum_{k=1}^n P(A_k)$.

AXIOMA 3'. (σ -aditividade). Se $A_1, A_2, \dots \in \mathcal{A}$ são disjuntos (i.e., mutuamente

²² A prova desta proposição pode ser encontrada em James, B., cap.1, p.5 e p.6.

²³ Para exemplos de σ -álgebras envolvendo eventos aleatórios e definição de σ -álgebra de Borel, ver James (1981, cap.1, p.7 e p.8).

exclusivos), então $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$.

DEFINIÇÃO 3.4. Uma função P definida numa σ -álgebra \mathcal{A} e satisfazendo os Axiomas 1, 2 e 3' chama-se uma *medida de probabilidade em \mathcal{A}* ou simplesmente uma *probabilidade em \mathcal{A}* (JAMES, 1981, cap.1, p.10).

Considerando os Axiomas 1, 2 e 3, o Axioma 3' é equivalente ao:

AXIOMA 4. (Continuidade no vazio). Se a seqüência $(A_n)_{n \geq 1}$, onde $A_n \in \mathcal{A} \forall n$, decrescer para o vazio, então $P(A_n) \rightarrow 0$ quando $n \rightarrow \infty$.

Considere P uma probabilidade em uma σ -álgebra \mathcal{A} . Suponha que todo A nas propriedades apresentadas abaixo pertença a \mathcal{A} . Portanto, as seguintes propriedades de probabilidade²⁴ são conseqüências dos axiomas:

P.1. $P(A^c) = 1 - P(A)$.

P.2. $0 \leq P(A) \leq 1$.

P.3. $A_1 \subset A_2 \implies P(A_1) \leq P(A_2)$.

P.4. $P(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i)$.

P.5. $P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$.

P.6. (Continuidade de Probabilidade). Se $A_n \uparrow A$, então $P(A_n) \uparrow P(A)$. Se $A_n \downarrow A$, então $P(A_n) \downarrow P(A)$.

Com as propriedades acima definidas, agora é possível definir probabilidade condicional e, a partir desta definição, chegar ao Teorema de Bayes.

DEFINIÇÃO 3.5. Seja (Ω, \mathcal{A}, P) um espaço de probabilidade²⁵. Se $B \in \mathcal{A}$ e $P(B) > 0$,

²⁴ Para mais detalhes, ver James, B., cap.1, p.13.

²⁵ Um espaço de probabilidade é um trio (Ω, \mathcal{A}, P) onde: (a) Ω é um conjunto não-vazio, (b) \mathcal{A} é uma σ -álgebra de subconjuntos de Ω e (c) P é uma probabilidade em \mathcal{A} . (James, B., cap.1, p. 13).

a probabilidade condicional de A dado B é definida por (JAMES, 1981, cap. 1, p. 14):

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad A \in \mathcal{A} \quad (3.8)$$

A partir desta definição, temos então que $P(A \cap B) = P(B)P(A|B)$. Esta igualdade pode ser generalizada, de forma que, sendo A , B e C eventos aleatórios, temos $P(A \cap B \cap C) = P(A)P(B|A) \cdot P(C|A \cap B)$. Por indução, o seguinte teorema pode ser estabelecido:

TEOREMA 3.1. (Teorema da Multiplicação ou Teorema da Probabilidade Composta) (JAMES, 1981, cap. 1, p. 16).

Seja (Ω, \mathcal{A}, P) um espaço de probabilidade. Então:

- (i) $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B), \forall A, B \in \mathcal{A},$
 $P(A_1 \cap A_2 \cap \dots \cap A_n) =$
 (ii) $P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap \dots \cap A_{n-1})$
 $, \forall A_1, \dots, A_n \in \mathcal{A}, \forall n = 2, 3, \dots$

Agora, supondo que A_1, A_2, \dots sejam eventos aleatórios mutuamente exclusivos e exaustivos²⁶. Então os A_i formam uma partição do espaço amostral Ω . Vamos supor também que a seqüência A_1, A_2, \dots seja finita ou enumerável, o que significa, por exemplo, que A e A^c formam uma partição, $\forall A \in \mathcal{A}$.

Para todo evento $B \in \mathcal{A}$, temos $B = \cup_i (A_i \cap B)$. Como os A_i são disjuntos, então os $B \cap A_i$ são disjuntos e $P(B) = \sum_i P(A_i)P(B|A_i)$. Então, podemos estabelecer o seguinte teorema:

TEOREMA 3.2. (Teorema da Probabilidade Total (ou Absoluta)) (JAMES, 1981, cap.

²⁶ Se os A_i são disjuntos, então isto significa que eles são mutuamente exclusivos e que $\cup A_i = \Omega$.

1, p. 17).

Se a seqüência (finita ou enumerável) de eventos aleatórios A_1, A_2, \dots formar uma partição de Ω , então

$$P(B) = \sum_i P(A_i)P(B|A_i), \forall B \in \mathcal{A}. \quad (3.9)$$

Usando este teorema, podemos calcular a probabilidade de A_i dado a ocorrência de B:

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} \Rightarrow P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_j P(A_j)P(B|A_j)} \quad (3.10)$$

Esta é a fórmula de Bayes. Ela é útil quando conhecemos as probabilidades dos A_i e a probabilidade condicional de B dado A_i , mas não conhecemos diretamente a probabilidade de B.

3.3.2 Os componentes do Teorema de Bayes

Vamos olhar mais uma vez para a equação (3.4), que apresenta o Teorema de Bayes. Voltando, à esquerda temos que $p(\theta|y)$ é a distribuição *a posteriori*, o numerador no lado direito contém a verossimilhança, $p(y|\theta)$, e a *priori*, $p(\theta)$. O denominador à direita, $p(y)$, é chamado de distribuição marginal dos dados ou distribuição preditora. Como este último não envolve θ ele pode ser desprezado para os propósitos de inferência e geralmente vemos a fórmula de Bayes escrita na forma da equação (3.7) reproduzida abaixo:

$$p(\theta|y) \propto p(y|\theta) \cdot p(\theta) \quad (3.7)$$

Começamos, portanto, olhando cada um dos elementos que formam esta última equação.

3.3.2.1 A Verossimilhança $p(y|\theta)$

A expressão para a distribuição dos dados que serão observados, restrita aos parâmetros, $p(y|\theta)$, tem dois nomes. Pode ser chamada de função densidade de probabilidade de Y dado θ , quando pensamos nela como a densidade de probabilidade de Y desenvolvida no ponto y , condicional ao parâmetro que toma o valor de θ , ou pode ser chamada de função Verossimilhança (função de θ) quando pensamos em y como o dado real coletado, algumas vezes denotado como y^{obs} . Neste caso, costuma-se denotar a Verossimilhança como $l(\theta; y)$. A função de verossimilhança não é, em geral, uma distribuição de probabilidade para θ dado y . A Verossimilhança é uma função de θ com os valores dos dados servindo como parâmetros conhecidos daquela função.

Junto à escolha de uma função Verossimilhança está a escolha de uma família de distribuições de probabilidade, uma para cada $\theta \in \Theta$. A teoria da probabilidade oferece muita destas distribuições. Existe desde distribuições simples, apropriadas aos dados que podem ser estimados como realizações independentes condicionalmente a partir de distribuições elementares de probabilidade com um pequeno número de parâmetros, até modelos de probabilidade para variáveis aleatórias envolvendo muitos parâmetros, com padrões complexos de dependência (LANCASTER, 2004, cap.1, p. 11).

A abordagem clássica trata θ como fixo e desconhecido, e os estimadores de θ baseados na amostra, chamados de $\hat{\theta}$, são tratados como aleatórios. A função de Verossimilhança é então usada para escolher um estimador que a maximize. Em contraste, a inferência Bayesiana trata $\hat{\theta}$ como fixo, condicional aos dados observados y , e trata θ como aleatório. O foco de interesse da abordagem Bayesiana nesta questão está em fazer afirmações sobre a probabilidade *a posteriori* de θ (WITHERS, 2002).

3.3.2.2 O Princípio da Verossimilhança

Em estatística, o princípio da Verossimilhança é um princípio controverso da inferência

estatística que afirma que toda a informação de uma amostra está contida na função Verossimilhança. Uma função Verossimilhança surge a partir de uma distribuição de probabilidade condicional considerada como uma função do seu segundo argumento, mantendo o primeiro argumento fixo. Por exemplo, considere um modelo que fornece a função de densidade de probabilidade de uma variável aleatória observável X como função de um parâmetro θ . Então, para um valor específico x de X , a função $L(\theta|x) = P(X = x|\theta)$ é a função verossimilhança de θ : ela fornece a medida de quão “provável” é qualquer valor particular de θ , se sabemos que X assume o valor de x . Duas funções de verossimilhança são equivalentes se uma é um múltiplo escalar da outra que, de acordo com o princípio da verossimilhança, toda a informação a partir dos dados relevantes para a inferência sobre os valores de θ será encontrada em classes equivalentes (LIKELIHOOD PRINCIPLE, 2007).

Então, o **Princípio da Verossimilhança** estabelece que as verossimilhanças que são proporcionais podem levar às mesmas inferências (dada uma mesma *priori*). Note que os dados que possivelmente viessem a ser observados podem ser diferentes. O que importa à inferência Bayesiana é os dados que efetivamente foram observados; os dados que pudessem ter sido observados, mas não foram, são irrelevantes (LANCASTER, 2004, cap. 1, p. 23).

Exemplo 3.1. Supondo:

- X é o número de sucessos em doze experimentos independentes de Bernoulli com probabilidade de sucesso θ em cada experimento e,
- Y é o número de experimentos independentes de Bernoulli necessários para obter três sucessos, com probabilidade de sucesso θ em cada experimento.

Então, a observação que $X=3$ induz a função verossimilhança:

$$L(\theta|X = 3) = \binom{12}{3} \theta^3 (1 - \theta)^9 = 220 \theta^3 (1 - \theta)^9$$

E a observação que $Y=12$ induz à função verossimilhança:

$$L(\theta|Y = 12) = \binom{11}{2} \theta^3 (1 - \theta)^9 = 55 \theta^3 (1 - \theta)^9$$

Elas são equivalentes porque cada uma é um múltiplo escalar da outra. O princípio da

verossimilhança então diz que as inferências extraídas para um valor de θ podem ser as mesmas em ambos os casos.

A diferença entre observar $X=3$ e $Y=12$ está somente na forma do experimento: em um caso é preciso repetir o experimento doze vezes; no outro, é preciso repetir o experimento até que três sucessos sejam observados. O resultado é o mesmo em ambos os casos. Portanto, o princípio da verossimilhança algumas vezes é estabelecido dizendo que *a inferência deve depender somente do resultado do experimento, e não da forma do experimento* (LIKELIHOOD PRINCIPLE, 2007).

3.3.2.3 Identificação

É perfeitamente possível para uma função Verossimilhança indicar não um valor particular de θ , mas vários elementos de um conjunto de pontos em Θ , considerando valores iguais para a verossimilhança. Por exemplo, no espaço paramétrico formado por dois parâmetros onde θ_1 e θ_2 pertencentes a Θ , e se $P(y|\theta_1) = P(y|\theta_2)$, então não há mudança nas crenças sobre θ .

Não há problema algum neste fato; ele só pode ser desapontador. Você dependerá da distribuição *a priori* para diferenciar os valores de θ . E se a distribuição *a priori* associa iguais probabilidades a estes valores, então eles serão igualmente prováveis *a posteriori*. Podemos imaginar uma verossimilhança derivada a partir de uma distribuição condicional $p(y|\theta)$, $y \in \Omega$, $\theta \in \Theta$ tal que, para um conjunto de $\theta \in \Theta$ temos que $p(y|\theta)$ é igual a uma constante para todo $y \in \Omega$. Isto significa que, não importa quais realizações de Y tenham ocorrido, existe um conjunto de valores de θ em que todos darão os mesmos valores para a verossimilhança. Neste caso dizemos que θ é não-identificado (LANCASTER, 2004, cap. 1, p. 25).

Então, o valor de um parâmetro θ_a é **identificado** se não existe outro valor θ_b tal que $p(y|\theta_a) = p(y|\theta_b)$. Um modelo é identificado se todos os valores dos parâmetros são identificados e, neste caso, o parâmetro θ é dito identificado (LANCASTER, 2004, cap.

1, p. 26). Então, podemos definir identificação como:

DEFINIÇÃO 3.6. Dado $p(y|\theta)$ sendo a função verossimilhança que depende de um vetor k de parâmetros, θ , definido em um conjunto Θ . Então θ é identificado no ponto $\theta = \theta_0 \in \Theta$ se e somente se, para todo $\theta \in \Theta$, $\theta \neq \theta_0$ implica $p(y|\theta) \neq p(y|\theta_0)$ para todos os possíveis valores de y . Se θ é identificado em todos os pontos $\theta_0 \in \Theta$, então dizemos que θ é identificado (Koop, 2004, cap. 12, p. 291).

A falta de identificação não causa problemas para a inferência Bayesiana. Apesar de tudo, uma vez que $p(\theta)$ e $p(y|\theta)$ sejam p.d.f.s válidas, então a *posteriori* $p(\theta|y) \propto p(\theta)p(y|\theta)$ será uma p.d.f válida e a inferência Bayesiana pode ser realizada normalmente. No entanto, existem alguns assuntos que devem ser mencionados. Considere um caso onde o modelo é não-identificado, onde $p(y|\theta) = c$ para todo $\theta \in \Theta_0$ onde $\Theta_0 \subseteq \Theta$. Neste caso a regra de Bayes implica $p(\theta|y) \propto p(\theta)$ para todo $\theta \in \Theta_0$. Em palavras, a *priori* é igual à *posteriori* na região não-identificada e não aprendemos nada sobre os parâmetros nesta região. Se a região não-identificada é ilimitada, então o uso de uma *priori* dos tipos imprópria, não-informativa deve ser evitado²⁷.

3.3.2.4 Permutabilidade

O conceito de permutabilidade é fundamental na análise Bayesiana. Uma seqüência finita de eventos (ou variáveis aleatórias) é **permutável** se a distribuição de probabilidade conjunta da seqüência, ou qualquer subsequência, é invariante sob as permutações do subscrito. Uma seqüência infinita é permutável se qualquer subsequência finita é permutável (POIRIER; TOBIAS, 2004).

A permutabilidade implica que as variáveis aleatórias $\{Y_i\}$ tenham todas as mesmas médias e as mesmas variâncias e que as correlações entre todos os pares Y_i, Y_j precisam

²⁷ Para mais detalhes e um exemplo ilustrativo, ver Koop, cap. 12, p. 292.

ser as mesmas para todos os outros pares. Note que seqüências permutáveis não são necessariamente seqüências de variáveis aleatórias independentes e identicamente distribuídas (i.i.d), apesar de que as seqüências de variáveis aleatórias i.i.d são permutáveis.

A relevância desta idéia à questão da escolha da *priori* é um famoso resultado de Finetti. Ele estabelece que se uma seqüência de n variáveis aleatórias binárias é permutável para todo n , então a distribuição de probabilidade conjunta de Y_1, Y_2, \dots, Y_n precisa tomar a forma:

$$p(y_1, y_2, \dots, y_n) = \int \theta^s (1 - \theta)^{n-s} dF(\theta) \quad (3.11)$$

Esta tem a forma da distribuição marginal dos dados derivada a partir da Verossimilhança igual a $\theta^s (1 - \theta)^{n-s}$ e uma função distribuição *a priori* igual a $F(\theta)$ (LANCASTER, 2004, cap.1, p.27). Então, a permutabilidade implica a existência de uma verossimilhança e de uma *a priori*. Isto significa que você não é necessário que se inicie uma modelagem com a afirmação que um conjunto de variáveis aleatórias é independente e identicamente distribuído. Você pode somente estabelecer que suas crenças sobre as variáveis em questão são permutáveis e isto implicará que o modelo toma à forma de uma verossimilhança e uma *a priori*.

3.3.3 A priori $p(\theta)$

Uma probabilidade *a priori* é uma probabilidade marginal, interpretada como uma descrição do que é conhecido sobre uma variável na ausência de qualquer evidência. A probabilidade *a posteriori* é então a probabilidade condicional da variável, levando em consideração a evidência. A probabilidade *a posteriori* é calculada a partir da *priori* e da função verossimilhança via Teorema de Bayes.

Na inferência estatística Bayesiana, uma distribuição de probabilidade *a priori* de certa quantidade q é a distribuição de probabilidade que irá expressar a incerteza sobre q

antes dos dados serem levados em consideração. Significa atribuir incerteza para uma quantidade incerta. Aplicando o Teorema de Bayes, multiplicando *a priori* pela função Verossimilhança e depois tomando a normalização, obtemos a distribuição de probabilidade *a posteriori*, que é a distribuição condicional da quantidade incerta considerando os dados. Uma *priori* é, portanto uma afirmação puramente subjetiva.

3.3.3.1 *Priori* Conjugada

Muitas vezes, para obter a distribuição de probabilidade *a posteriori*, será usada, quando possível, uma distribuição de probabilidade *a priori* conjugada. Na teoria de probabilidade Bayesiana, uma classe de distribuições de probabilidade *a priori* $p(\theta)$ é chamada de conjugada à classe de funções Verossimilhança $p(y|\theta)$ se as distribuições *a posteriori* resultantes $p(\theta|y)$ são da mesma família que $p(\theta)$. Por exemplo, a família Gaussiana é conjugada a ela mesma: se a função Verossimilhança é Gaussiana, escolher uma *priori* Gaussiana irá assegurar que a distribuição *a posteriori* é, também, Gaussiana.

Considere o problema geral de inferência de distribuição para um parâmetro θ , considerando os dados y . A partir do Teorema de Bayes, a distribuição *a posteriori* é calculada a partir da *priori* $p(\theta)$ e da função Verossimilhança $\theta \Rightarrow p(y|\theta)$ como:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta} \quad (3.12)$$

Onde a função Verossimilhança será considerada fixa. A função Verossimilhança geralmente é bem determinada a partir da definição do processo geracional dos dados. É claro que diferentes escolhas da distribuição *a priori* $p(\theta)$ podem fazer com que a integral seja mais ou menos difícil de ser calculada, e o produto $p(y|\theta) \times p(\theta)$ pode ter uma ou outra forma algébrica. Para certas escolhas da *priori*, a *posteriori* tem a mesma forma algébrica que a *priori*. Tal escolha é uma *priori* conjugada (CONJUGATE

PRIOR, 2007).

3.3.3.2 A função *kernel*

A função de probabilidade de uma variável aleatória Y geralmente assume a forma $kg(y)$ onde k é uma constante numérica que assegura que $kg(y)$ integre um. A parte restante, $g(y)$, que envolve y , é chamada de *kernel* da função. O propósito da constante numérica k é o de assegurar que a função de probabilidade integre um. Uma vez que o *kernel* seja conhecido, a constante pode ser encontrada por integração, mas ela tem pouca importância.

Como notamos anteriormente, o Teorema de Bayes é frequentemente estabelecido com a constante omitida:

$$p(\theta|y) \propto p(y|\theta) \cdot p(\theta) \quad (3.7)$$

ou, em palavras, a *posteriori* é proporcional ao produto da verossimilhança com a *priori* (LANCASTER, 2004, cap. 1, p. 31).

3.3.3.3 *Prioris* Informativas

Uma *priori* informativa expressa uma informação definida, específica sobre uma variável. Um exemplo é a distribuição *a priori* para a temperatura da tarde de amanhã. Uma aproximação razoável é tomar a *priori* como uma distribuição Normal, com valor esperado igual à temperatura da tarde de hoje, com variância igual à variância da temperatura atmosférica durante o dia de hoje.

Este exemplo tem uma propriedade em comum com muitas *prioris*, que é o de tomar a *posteriori* a partir de um problema (no exemplo, a temperatura de hoje) e usá-la como a *priori* para outro problema (no exemplo, a temperatura da tarde de amanhã). Uma

evidência pré-existente é parte da *priori* e na medida em que existem mais evidências, a *priori* é determinada em maior parte por essas informações ao invés de qualquer suposição original. Neste caso a *priori* considera em grande parte o que as evidências anteriores sugerem (INFORMATIVE PRIORS, 2007).

3.3.3.4 *Prioris* Não-Informativas

Uma *priori* não-informativa expressa uma informação vaga ou geral sobre uma variável. *Prioris* não-informativas podem expressar informações tais como “a variável é positiva” ou “a variável é menor que um determinado limite” e em algumas pessoas preferem o termo *priori objetiva*.

Nos problemas de estimação de parâmetros, o uso de uma *priori* não-informativa geralmente fornece resultados bastante parecidos com os resultados encontrados na análise estatística convencional uma vez que a Verossimilhança fornece mais informação do que a *priori* não-informativa.

Problemas filosóficos associados com *prioris* não-informativas estão ligados à escolha de uma métrica apropriada, ou escala de medida. Suponha que queremos uma *priori* para a velocidade de corrida de um atleta conhecido por nós. Podemos especificar, digamos, uma distribuição Normal como a *priori* para esta velocidade, mas também, de forma alternativa, podemos especificar uma *priori* Normal para o tempo que este corredor leva para completar cem metros, que é proporcionalmente recíproco à primeira *priori*. Estas *prioris* são bastante diferentes, mas não é claro qual delas é preferível neste contexto.

Similarmente, se o problema consiste em estimar uma proporção desconhecida entre 0 e 1, podemos dizer que todas as proporções são igualmente prováveis, e então usamos uma *priori* uniforme. Alternativamente, podemos dizer que todas as ordens de magnitude para a proporção são igualmente prováveis, o que fornece uma *priori* proporcional ao logaritmo. A *priori* de Jeffreys tenta resolver este problema calculando uma *priori* que expressa uma mesma crença, não importa a métrica usada. A *priori* de Jeffreys para uma proporção desconhecida p é $p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}}$.

Problemas práticos associados com *prioris* não-informativas incluem a necessidade de que a distribuição *a posteriori* seja própria. As *prioris* não-informativas usuais para variáveis contínuas não limitadas são impróprias. Esta necessidade não será um problema se a distribuição *a posteriori* for própria (PRIORI PROBABILITY, 2007).

3.3.3.5 Priori de Jeffreys

Com o objetivo de contornar os problemas de parametrização, Jeffreys procurou uma regra geral para a escolha da *priori* de tal forma que as inferências *a posteriori* fossem obtidas independentemente da parametrização escolhida. Isto é, ele argumenta que crenças *a posteriori* podem ser invariantes à parametrização usada. Jeffreys mostrou que existe uma regra que satisfaz esta condição de invariância (POIRIER; TOBIAS, 2004). Ele toma o argumento geral de escolher uma *priori* não-informativa proporcional a raiz quadrada da informação de Fischer:

$$I_{\varphi} = -E \left(\frac{\partial^2 \log l(\theta|y)}{\partial \theta^2} \right) \quad (3.13)$$

onde a esperança é tomada em relação a $p(y|\theta)$. Esta é a segunda derivada (negativa) do logaritmo da função verossimilhança ponderada sobre as realizações repetidas de y .

Suponha que uma segunda reparametrização se dá em termos de $h(\theta)$, por exemplo, θ pode ser σ e $\gamma = h(\theta)$ pode ser $1/\sigma^2$. Agora, note que:

$$\frac{\partial \log l}{\partial \gamma} = \frac{\partial \log l}{\partial \theta} \cdot \frac{\partial \theta}{\partial \gamma}$$

$$\frac{\partial^2 \log l}{\partial \gamma^2} = \frac{\partial^2 \log l}{\partial \theta^2} \left(\frac{\partial \theta}{\partial \gamma} \right)^2 + \frac{\partial \log l}{\partial \theta} \frac{\partial^2 \theta}{\partial \gamma^2}$$

Então

$$I_{\gamma} = I_{\theta} \left(\frac{\partial \theta}{\partial \gamma} \right)^2 \quad (3.14)$$

onde a última linha vem do fato que $E(\partial \log l / \partial \theta) = 0$ ²⁸. Neste caso, I_{γ} é a informação sobre γ e I_{θ} é a informação sobre θ . Note que (3.14) implica que $I_{\gamma}^{1/2} = I_{\theta}^{1/2} |\partial \theta / \partial \gamma|$. Agora, a distribuição *a posteriori* de θ em termos de γ , e seguindo a regra de Jeffreys, será $l(\theta) I_{\theta}^{1/2}$. Da mesma forma, em termos de γ , sua distribuição *a posteriori* será $l(h(\gamma)) I_{\gamma}^{1/2}$. A partir disto, podemos observar que temos as mesmas crenças *a posteriori* sobre θ nos dois casos de reparametrização.

As crenças *a priori* formadas usando a regra de Jeffreys são geralmente impróprias, isto é, *prioris* que integram infinito ao invés de integrar um (como no caso das *prioris* próprias). O argumento da invariância à reparametrização generaliza-se para o caso em que θ é um vetor de parâmetros. A informação é substituída por uma matriz informação e a regra de Jeffreys toma a forma $|I_{\theta}|^{1/2}$, isto é, a raiz quadrada do determinante da matriz informação (LANCASTER, 2004, cap. 1, p. 35).

3.3.3.6 Prioris Impróprias

Vamos pensar no Teorema de Bayes escrito na forma da equação (3.10), abaixo reproduzida para facilitar a apresentação:

$$P(A_i | B) = \frac{P(A_i)P(B|A_i)}{\sum_j P(A_j)P(B|A_j)} \quad (3.10)$$

²⁸ Tome a identidade $\int p(y|\theta)dy = 1$; calcule a diferencial em relação a θ ; usando $\frac{\partial \log p}{\partial \theta} = \frac{1}{p} \frac{\partial p}{\partial \theta}$, reorganize a equação.

Então, é claro que deve ele permanecer verdadeiro se todas as probabilidades *a priori* $P(A_i)$ e $P(A_j)$ forem multiplicadas por uma constante. O mesmo deve ser verdadeiro no caso de uma variável aleatória contínua. As probabilidades *a posteriori* continuarão somando um mesmo se os valores das *prioris* não. Então, as *prioris* só devem ser especificadas na proporção correta.

Expandindo esta idéia, em muitos casos a soma ou a integral das *prioris* pode não ser finita para obter respostas sensíveis para as probabilidades *a posteriori*. Quando este é o caso, a *priori* é chamada de *priori* imprópria. Alguns estatísticos usam *prioris* impróprias como *prioris* não-informativas. Por exemplo, se elas precisam de uma distribuição *a priori* para média e variância de uma variável aleatória, pode-se assumir $p(m, v) \sim 1/v$ (para $v > 0$), o que sugere que qualquer valor para a variância positiva se torna menos provável na ordem inversa a este valor.

Visto que:

$$\int_{-\infty}^{\infty} dm = \int_0^{\infty} \frac{1}{v} dv = \infty \quad (3.15)$$

Isto será uma *priori* imprópria tanto para a média como para a variância (IMPROPER PRIORS, 2007).

3.3.4 A *posteriori* $p(\theta|y)$

A probabilidade *a posteriori* de um evento aleatório é a probabilidade condicional que é determinada quando a evidência relevante é levada em consideração.

A distribuição de probabilidade *a posteriori* de uma variável aleatória, dado o valor de outra variável aleatória pode ser calculada pelo Teorema de Bayes multiplicando a distribuição de probabilidade *a priori* pela função verossimilhança e então dividindo pela constante de normalização, como segue (POSTERIOR PROBABILITY, 2007):

$$f_{X|Y=y}(x) = \frac{f_X(x)L_{X|Y=y}(x)}{\int_{-\infty}^{\infty} f_X(x)L_{X|Y=y}(x)dx} \quad (3.16)$$

Esta expressão é a função densidade de probabilidade *a posteriori* para uma variável aleatória X , dado o conjunto de dados $Y=y$ onde:

- $f_X(x)$ é a densidade *a priori* de X ;
- $L_{X|Y=y}(x) = f_{Y|X=x}(y)$ é a função verossimilhança como função de x ;
- $\int_{-\infty}^{\infty} f_X(x)L_{X|Y=y}(x)dx$ é a constante de normalização e
- $f_{X|Y=y}(x)$ é a densidade *a posteriori* de X , dado o conjunto de dados $Y=y$.

3.3.4.1 Apresentação da distribuição *a posteriori*

Quais as maneiras mais comuns de apresentar a distribuição *a posteriori*? As três principais formas são:

(1) Momentos: a prática econométrica tradicional geralmente calcula uma estimativa pontual de θ , assim como uma estimativa para o desvio-padrão do estimador. Desta forma, apresenta-se a distribuição informando a média (ou mediana) da distribuição *a posteriori* junto com o desvio-padrão da mesma (LANCASTER, 2004, cap. 1, p. 43).

(2) Região da densidade *a posteriori* superior: O análogo bayesiano a um intervalo de confiança é encontrar, a partir da distribuição *a posteriori* de θ , um intervalo no espaço paramétrico Θ tal que, com probabilidade de 0.95, θ esteja dentro dele. É claro que existem muitas formas de capturar 95% da probabilidade de uma distribuição. A prática Bayesiana padrão é construir um intervalo de tal forma que nenhum ponto (de Θ), ou “contido no intervalo” tenha densidade de probabilidade menor que qualquer ponto fora dele. Isto é chamado de intervalo de (95%) densidade *a posteriori* superior²⁹

²⁹ Conhecido como intervalo hpd na literatura, do inglês *highest posterior density interval*.

(LANCASTER, 2004, cap. 1, p. 43).

(3) Cálculo das Marginais: o cálculo envolvido para formar a distribuição *a posteriori* do objeto de interesse pode ser matematicamente bastante complicado. Se o objeto de interesse for, digamos o terceiro elemento θ_3 do parâmetro de um modelo com k elementos para encontrar sua densidade marginal consistem em calcular a soma:

$$p(\theta_3|y) = \int_{\theta_1} \int_{\theta_2} \int_{\theta_4} \dots \int_{\theta_k} p(\theta|y) d\theta_1 d\theta_2 d\theta_4 \dots d\theta_k \quad (3.17)$$

Isto é, uma integração de $k-1$ dimensões. Este é, em geral, um problema difícil. Felizmente, existem dois artifícios bastante abrangentes para encontrar uma solução: o primeiro é o uso de aproximações para a distribuição *a posteriori* e o segundo é um método de amostragem, (feito por cálculos computacionais), baseado em resultados de Cadeias de Markov e computacionalmente intensivo (LANCASTER, 2004, cap.1, p. 44).

3.4 Estimação por simulação: MCMC

Desde o início dos anos 90 é possível observar um aumento significativo em pesquisas Bayesianas aplicadas. Este fato deve-se ao desenvolvimento de ferramentas computacionais poderosas, que possibilitam a estimação de modelos estatísticos Bayesianos mais complicados. A maioria destes modelos Bayesianos não pode ser estimada usando técnicas convencionais. O objetivo desta seção é, portanto, apresentar, em termos gerais, os elementos básicos dos cálculos Bayesianos.

Considere o problema enfrentado pela econometria Bayesiana de calcular a média *a posteriori* de alguma função de interesse, g :

$$E[g(\theta)|y] = \frac{\int_{\Theta} g(\theta) f(\theta|y) d\theta}{\int_{\Theta} f(\theta|y) d\theta} \quad (3.18)$$

O denominador da expressão acima representa a constante de normalização da *posteriori* conjunta, uma vez que ela é somente conhecida como constante de proporcionalidade do Teorema de Bayes. Na prática, em qualquer problema de complexidade razoável, a expressão acima não pode ser desenvolvida analiticamente. Então, vamos considerar o fato de não existir solução analítica para (3.18) no restante desta sessão.

3.4.1 Métodos não-iterativos

A forma mais simples de olhar para os cálculos diretos da integral em (3.18) é supor que podemos tomar extrações diretas a partir da *posteriori* conjunta $f(\theta|y)$. Denote $\{\theta_i\}_{i=1}^N$ como um conjunto de N extrações, onde $\theta_i \sim f(\theta|y)$. A Lei Forte dos Grandes Números³⁰ garante que:

$$E[g(\theta)|y] = \frac{1}{N} \sum_{i=1}^N g(\theta_i) \quad (3.19)$$

O momento acima é um estimador consistente de $E[g(\theta)|y]$. Podemos notar que a média *a posteriori* e o desvio-padrão *a posteriori* de θ podem ser obtidos desta forma, considerando uma escolha adequada para a função g , e que a precisão da estimativa aumenta na medida em que N aumenta. Este aumento no número N de extrações está nas mãos dos pesquisadores e não está limitado pelo tamanho do espaço amostral em questão. Podemos, então, estimar o momento *a posteriori* de interesse com precisão absoluta simplesmente aumentando o número de extrações tomadas a partir da *posteriori* (POIRIER; TOBIAS, 2004).

A aproximação apresentada em (3.19) é também chamada de integração direta de Monte

³⁰ A Lei Forte dos Grandes Números estabelece que, se X_1, X_2, X_3, \dots é uma seqüência infinita de variáveis aleatórias, que são independentes duas a duas e identicamente distribuídas com $E(X_i) = \mu$ e $E(X_i^2) < \infty$, então $P(\lim_{n \rightarrow \infty} \overline{X_n} = \mu) = 1$, i.e., a média da amostra certamente converge para μ (STRONG LAW OF LARGE NUMBERS, 2007).

Carlo e requer alguns requisitos para que esta aproximação seja válida. Com o objetivo de aplicar tal método, é necessário que seja possível obter amostras i.i.d a partir da distribuição *a posteriori* $f(\theta|y)$. Uma vez que (3.19) só pode ser desenvolvida analiticamente em poucos casos, quase nunca conseguimos obter amostras i.i.d a partir de $f(\theta|y)$ em problemas de complexidade razoável. Para este fim, um método mais aplicável é conhecido como *importance sampling*. Reescrevendo a equação (3.19) com integrais, temos uma noção intuitiva para o estimador de *importance sampling*:

$$E[g(\theta)|y] = \frac{\int_{\Theta} g(\theta)f(\theta|y)d\theta}{\int_{\Theta} f(\theta|y)d\theta} \quad (3.20)$$

$$= \frac{\int_{\Theta} \left[\frac{g(\theta)f(\theta|y)}{I(\theta)} \right] I(\theta)d\theta}{\int_{\Theta} \left[\frac{f(\theta|y)}{I(\theta)} \right] I(\theta)d\theta} \quad (3.21)$$

Onde $I(\theta)$ é chamada de função valor. Escrevendo desta forma, podemos ver que as integrais originais foram transformadas em novos problemas de cálculos de momentos, onde o cálculo da média é agora apresentado em função de $I(\theta)$ ao invés de $f(\theta|y)$. Para termos o estimador por *importance sampling*, desde que possamos tomar extrações a partir da função $I(\theta)$, a integração direta de Monte Carlo pode ser apresentada individualmente tanto no numerador quanto no denominador:

$$\begin{aligned} E[g(\theta)|y] &= \frac{\frac{1}{N} \sum_{i=1}^N [g(\theta_i)f(\theta_i|y)]/I(\theta_i)}{\frac{1}{N} \sum_{i=1}^N f(\theta_i|y)/I(\theta_i)} = \frac{\frac{1}{N} \sum_{i=1}^N g(\theta_i)w(\theta_i)}{\frac{1}{N} \sum_{i=1}^N w(\theta_i)} \\ &= \sum_{i=1}^N g(\theta_i)\hat{w}(\theta_i) \end{aligned} \quad (3.22)$$

A expressão (3.22), que é o estimador de *importance sampling*, é uma média ponderada de $g(\theta_i)$ com $\hat{w}(\theta_i) = w(\theta_i)/\sum_i w(\theta_i)$ denotando o peso normalizado e $w(\theta_i) = f(\theta_i|y)/I(\theta_i)$. Note que para o caso de *importance sampling*, $\theta_i \sim I(\theta)$, que não são extrações a partir de $f(\theta|y)$, como no caso da integração direta de Monte Carlo.

Uma vez que a função valor $I(\theta)$ está sob o controle do pesquisador, esta deve ser uma densidade a partir da qual as extrações podem ser facilmente obtidas (POIRIER; TOBIAS, 2004).

A equação (3.22) pode ser vista como uma forma conveniente de resolver os cálculos dos momentos da *posteriori*. Note que se $I(\theta)$ é uma aproximação pobre para $f(\theta|y)$, então os pesos $\tilde{w}(\theta)$ geralmente será menor para a maioria dos valores θ_i e, então a soma em (3.22) será dominada por poucos termos que receberão uma ponderação maior, resultando numa estimativa instável e pouco precisa. A precisão de uma estimativa calculada por *importance sampling* irá aumentar na medida em que $I(\theta)$ se aproxima da distribuição $f(\theta|y)$. No entanto, se $I(\theta)$ e $f(\theta|y)$ coincidem, então os pesos $\tilde{w}(\theta) = 1/N$, e o estimador em (3.22) se reduz ao do caso do estimador direto de Monte Carlo.

3.4.2 Métodos Iterativos – MCMC (“Markov Chain Monte Carlo”)

Como discutido no início desta sessão, os métodos de simulação iterativos, principalmente o Gibbs Sampler e o algoritmo de Metropolis-Hastings, são ferramentas poderosas que facilitam os cálculos em muitos modelos complexos. Apesar destes dois algoritmos geralmente serem apresentados como instrumentos distintos para simular a *posteriori* conjunta, esta diferença é de alguma forma, artificial. Podemos considerar o Gibbs Sampler como um caso especial do algoritmo de Metropolis-Hastings onde saltos ao longo das distribuições condicionais são aceitos com probabilidade um.

A estratégia geral para métodos iterativos é seguir os passos dos algoritmos para gerar uma série de extrações³¹, digamos, $\theta_0, \theta_1, \theta_2, \dots$ que converge em distribuição para alguma densidade alvo – em nosso caso, a densidade *a posteriori* $f(\theta|y)$. Os algoritmos são construídos e então a *posteriori* $f(\theta|y)$ é a distribuição estacionária única da cadeia de parâmetros. Uma vez que a convergência para a densidade alvo é encontrada, podemos usar estas extrações da mesma forma que as extrações são usadas na

³¹ Algumas vezes também chamada de cadeia de parâmetros.

integração direta de Monte Carlo para calcular a média *a posteriori*, o desvio-padrão *a posteriori*, etc. Na prática, tomamos cuidado em garantir que a cadeia de parâmetros tenha convergência para a densidade alvo, descartando o conjunto inicial de extrações antes de a convergência ocorrer (frequentemente chamado de período de *burn-in*), e então usar a amostra pós-convergência para calcular as densidades desejadas. Diferente dos métodos não-iterativos discutidos anteriormente, as extrações pós-convergência que são obtidas usando os métodos iterativos, provavelmente serão correlacionadas, como por exemplo, a distribuição de θ^t depende do último parâmetro da amostra tomada, θ^{t-1} . Se a correlação entre as extrações é forte, isto pode dificultar a travessia do espaço paramétrico e o erro padrão numérico, associado com os pontos estimados pode ser bem grande.

3.4.2.1 Gibbs Sampler

Dado θ sendo um vetor de parâmetros $k \times 1$, associado a uma distribuição *a posteriori* $f(\theta|y)$ e escrevendo $\theta = [\theta_1 \dots \theta_k]$. O algoritmo de Gibbs Sampling é descrito da seguinte forma:

1. Selecione um vetor inicial de parâmetros $\theta_0 = [\theta_0^1 \theta_0^2 \theta_0^3 \dots \theta_0^k]$. Esta condição inicial pode ser escolhida de forma arbitrária, tomando uma amostra a partir da *priori*, ou talvez possa ser obtida a partir de um método de estimação convencional, como por exemplo, por mínimos quadrados.

(a) Tome uma amostra θ_1^1 a partir da densidade condicional *a posteriori* completa $f(\theta^1 | \theta^2 = \theta_0^2, \theta^3 = \theta_0^3, \dots, \theta^k = \theta_0^k, y)$.

(b) Tome uma amostra θ_1^2 a partir de $f(\theta^2 | \theta^1 = \theta_1^1, \theta^3 = \theta_0^3, \dots, \theta^k = \theta_0^k, y)$.

(c) \vdots

(d) Tome uma amostra θ_1^k a partir de $f(\theta^k | \theta^1 = \theta_1^1, \theta^2 = \theta_1^2, \dots, \theta^{k-1} = \theta_1^{k-1}, y)$.

2. Repita os passos de *a* - *d* para obter $\theta_2 = [\theta_2^1 \ \theta_2^2 \ \dots \ \theta_2^k]$, θ_3 , etc. sempre condicionado aos valores mais recentes da extração dos parâmetros.

Para executar o Gibbs Sampler é necessária a habilidade de tomar extrações a partir das condicionais *a posteriori* do modelo. Ainda que a densidade conjunta *a posteriori* $f(\theta|y)$ frequentemente pode ser intratável, as condicionais completas $\{f(\theta^j | \theta^{-j}, y)\}_{j=1}^k$ (com θ^{-j} denotando todos os parâmetros que não o parâmetro θ^j) provam ser fora das formas padrão em muitos casos. Por esta razão o Gibbs Sampler geralmente é usado para estimar uma variedade de modelos econométricos (POIRIER; TOBIAS, 2004).

3.4.2.2 Algoritmo Metropolis-Hastings

O algoritmo de Metropolis-Hastings é um algoritmo do tipo aceita/ rejeita em que um valor candidato, digamos θ_c , é proposto e então decidimos se estabelecemos θ_{t+1} (o próximo valor da cadeia) igual a θ_c ou se permanecemos no valor corrente da cadeia, θ_t . Formalmente, dado $P(\theta|\theta_t)$ sendo uma densidade aproximada proposta (onde a dependência potencial no valor corrente da cadeia é feita de forma explícita)³², e considere amostras geradas a partir de $P(\theta|\theta_t)$ ao invés de amostras geradas a partir da distribuição $f(\theta|y)$.

Supondo que θ_c é tomado da amostragem a partir de $P(\cdot | \theta_t)$, vamos então estabelecer $\theta_{t+1} = \theta_c$ com probabilidade:

³² O último valor da cadeia θ_t não tem que afetar a amostragem a partir da densidade proposta. Se esta dependência é anulada, $P(\theta|\theta_t) = P(\theta)$ e a cadeia resultante é chamada de cadeia independente. Tal cadeia pode surgir se for para escolher uma densidade para calcular a moda *a posteriori* e a matriz de covariância *a posteriori* e então usada para centralizar e escalonar a densidade proposta.

$$\min \left\{ 1, \frac{f(\theta_0|y) P(\theta_1|\theta_0)}{P(\theta_0|\theta_1) f(\theta_1|y)} \right\} \quad (3.23)$$

E caso contrário, estabeleça $\theta_{t+1} = \theta_t$. No caso de uma densidade proposta simétrica, a probabilidade de aceitação descrita acima se reduz a $f(\theta_0|y)/f(\theta_1|y)$, de onde as extrações a partir das regiões de maior densidade são sempre aceitas pelo algoritmo e as extrações de regiões de menor densidade são ocasionalmente aceitas (POIRIER; TOBIAS, 2004).

4 O Modelo VAR Espacial com uma abordagem Bayesiana

Um modelo autorregressivo univariado é um modelo linear que trata de uma única variável e uma única equação, em que o valor corrente desta variável é explicado por seus próprios valores defasados. Em 1980, Christopher Sims apresentou uma extensão do modelo autorregressivo, a qual foi denominada de modelo autorregressivo vetorial, também conhecido como modelo VAR. Assim, um VAR é um modelo linear com n variáveis, de n equações, em que cada variável é explicada por seus próprios valores defasados, além dos valores passados das outras $n-1$ variáveis restantes. Este modelo fornece uma forma sistemática de capturar a dinâmica macroeconômica em mais de uma série de tempo (STOCK; WATSON, 2001).

Formalmente, o modelo VAR com n variáveis pode ser descrito pela equação abaixo. O termo ε_{it} denota o vetor $n \times 1$ dos termos dos erros³³, C_i representa as constantes e y_{it} , $i = 1, \dots, n$ denota as n variáveis do modelo no instante t . Os parâmetros do modelo, $A_{ij}(l)$, tomam a forma, $\sum_{k=1}^m \alpha_{ijk} l^k$ onde l é o operador defasagem definido por $l^k y_t = y_{t-k}$, e m é a ordem de defasagem escolhida na representação do modelo.

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{bmatrix} = \begin{bmatrix} A_{11}(l) & \dots & A_{1n}(l) \\ \vdots & \ddots & \vdots \\ A_{n1}(l) & \dots & A_{nn}(l) \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{bmatrix} + \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{nt} \end{bmatrix} \quad (4.1)$$

O modelo VAR estabelece um conjunto de relações entre os valores passados defasados de todas as variáveis do modelo e o valor corrente de cada variável do modelo. Como no exemplo aplicado no Capítulo 5, se y_{it} representa a produção industrial no estado i no instante t , a estrutura do modelo VAR permite que a produção industrial em cada estado seja explicada pela produção industrial passada no mesmo estado, y_{it-k} .

³³O vetor de erros satisfaz as seguintes suposições:

$E(\varepsilon_t) = 0$ - todo termo de erro tem média igual a zero;

$E(\varepsilon_t \varepsilon_t') = \Omega$ - a matriz de covariância contemporânea é Ω (uma matriz $n \times n$ positiva definida);

$E(\varepsilon_t \varepsilon_{t-k}') = 0$ para qualquer k diferente de zero – não existe correlação no tempo, em particular, não existe correlação serial entre os termos individuais dos erros.

$k = 1, \dots, m$, e também pela produção industrial passada em outros estados, Y_{jt-k} , $k = 1, \dots, m$ e $j \neq i$. Isto é interessante uma vez que a estrutura do modelo permite olhar para as diferenças regionais, isto é, diferenças entre os estados, e olhar para a forma de dependência temporal, isto é, a estrutura de dependência entre as variáveis.

Apesar da estrutura do modelo ser capaz de extrair informações tanto da dependência temporal como da dependência espacial, o modelo VAR apresenta limitações empíricas. Por exemplo, um modelo com 9 variáveis e 6 defasagens produz 55 variáveis independentes, incluindo a constante, (nove variáveis multiplicadas por seis defasagens mais a constante), em cada uma das 9 equações do modelo, com um total de 495 coeficientes a serem estimados (55 variáveis independentes por equação multiplicadas por 9 equações do modelo). Neste caso, seriam necessárias séries de tempo bastante longas para estimar o modelo VAR e, na ausência destas, a estimação do modelo poderia se tornar problemática, comprometendo os resultados encontrados. Uma forma encontrada de contornar este problema é tratá-lo através da abordagem Bayesiana, que considera informação *a priori*, tornando possível a estimação dos parâmetros do modelo (LESAGE, 1999).

Outra limitação apresentada pelo modelo VAR está relacionada à correlação temporal. Uma vez que as variáveis independentes do modelo representam valores defasados, por exemplo, $Y_{1t-1}, Y_{1t-2}, \dots, Y_{1t-6}$, estas tendem a apresentar correlações altas, resultando, muitas vezes, em parâmetros estimados com pouca precisão. A abordagem Bayesiana, através do uso de informação *a priori*, muitas vezes é capaz de contornar este problema.

Neste capítulo apresentaremos o modelo VAR segundo uma abordagem Bayesiana, utilizando informação *a priori* espacial. A abordagem Bayesiana em modelos VAR se destaca pelo fato de conseguir fornecer previsões macroeconômicas com razoável precisão, uma vez que a metodologia Bayesiana permite a imposição de restrições nos parâmetros do modelo e, portanto, reduz os problemas dimensionais enfrentados pelos modelos VAR. Com isto, é possível observar ganhos na estimação dos parâmetros e, conseqüentemente, previsões mais precisas (FELIX; NUNES, 2002). No Capítulo 5 será apresentada uma aplicação prática deste modelo apresentado aqui, assim como a análise dos resultados encontrados.

Muitos modelos VAR com abordagem Bayesiana baseiam-se numa especificação de informação *a priori* conhecida como *priori* de Minnesota, proposta inicialmente por Doan, Litterman e Sims (1984). Neste capítulo será apresentado um modelo que considera esta informação *a priori* assim como um modelo reparametrizado a partir do modelo primeiramente proposto, que utiliza informação *a priori* espacial.

Outro aspecto que deve ser citado é que modelos VAR não levam em consideração as relações explícitas de longo prazo, chamadas de relações de cointegração, entre as variáveis. Na presença de relações de cointegração entre as variáveis consideradas, um modelo VAR com um mecanismo de correção de erros pode melhorar o desempenho das previsões para as séries em questão. Também neste capítulo apresentaremos um modelo VAR seguindo a abordagem Bayesiana que incorpora o mecanismo de correção de erros e este modelo será chamado de modelo de vetores de correção de erros Bayesiano ou VECM Bayesiano.

4.1 O modelo VAR Bayesiano com informação *a priori* de Minnesota

Modelos de vetores autorregressivos foram primeiramente usados em experimentos de previsão de séries macroeconômicas, onde foi desenvolvido por Doan, Litterman e Sims (1984) um conjunto de *prioris* Bayesianas de médias e variâncias para os parâmetros. A especificação da informação *a priori* desenvolvida neste trabalho foi chamada de *priori* de Minnesota e demonstrou uma melhora nas previsões de séries de tempo macroeconômicas.

As variâncias *a priori* determinadas pela *priori* de Minnesota tratam todas as variáveis de forma similar, exceto as variáveis dependentes defasadas em cada uma de suas equações, determinando uma influência decrescente das variáveis autorregressivas independentes. A *priori* de Minnesota associa média *a priori* igual à zero para os coeficientes em todas as variáveis independentes que não sejam auto defasadas, de forma que exista uma influência decrescente destas variáveis no modelo. Isto é, a informação *a priori* de Minnesota, construída desta forma, agrega a suposição de que as variáveis independentes que não são auto defasadas exercem uma menor influência que as variáveis dependentes auto defasadas em cada equação. Por exemplo, no modelo

VAR apresentado na forma de (4.1), podemos dizer que $y_{1t-1}, y_{2t-2}, \dots, y_{3t-1}, y_{3t-2}, \dots, y_{nt-1}, y_{nt-2}, \dots$ exerce menos influência que $y_{1t-1}, y_{2t-2}, \dots$ com o objetivo de explicar y_{1t} , e esta influência é decrescente na medida em que as variáveis independentes se tornam mais distantes no tempo. A informação *a priori* de Minnesota é aplicada aos coeficientes do modelo VAR, isto é, aos parâmetros desconhecidos na Equação 4.1 (LESAGE; KRIVELYOVA, 1999).

Com o intuito de contornar os problemas de alta correlação que geralmente são observados nos modelos VAR, Doan, Litterman e Sims (1984) incorporam a informação *a priori* Bayesiana. A *a priori* de Minnesota assume que muitas séries macroeconômicas poderiam ser descritas por um passeio aleatório com tendência. De acordo com isso são assumidas distribuições *a priori* normais independentes para os parâmetros do modelo VAR, com suas médias assumidas iguais aos parâmetros de um passeio aleatório, isto quer dizer, suposição de médias iguais a 1 para as variáveis auto defasadas e médias iguais a zero para as outras variáveis. As variâncias das distribuições *a priori* são definidas de acordo com a relação funcional ligando-as a um segundo conjunto de parâmetros menor que o primeiro, denominados hiperparâmetros. A forma com que cada equação é limitada ao redor da média *a priori* é determinada por um conjunto de hiperparâmetros de “overall tightness”, que pode ser diferente para as diferentes equações do modelo. Para controlar o aumento da banda ao redor da média *a priori* para as defasagens que se afastam no tempo e para evitar um número excessivo de hiperparâmetros, é assumido que a largura da banda aumenta com a ordem de defasagem, isto é, a variância aumenta à medida que a variável se torna mais defasada no tempo. Além disso, existem hiperparâmetros que controlam as relações entre as variáveis, que podem diferir entre as equações (FELIX; NUNES, 2003).

As médias e variâncias *a priori* tomam a seguinte forma:

$$\begin{aligned} \beta_i &\sim N(1, \sigma_{\beta_i}^2) \\ \beta_j &\sim N(0, \sigma_{\beta_j}^2) \end{aligned} \quad (4.2)$$

Onde β_i denota os coeficientes associados com a variável dependente defasada em cada equação do modelo VAR e β_j representa qualquer outro coeficiente. As médias *a priori* para as variáveis dependentes defasadas são estabelecidas iguais a um, baseadas na crença de que estas são variáveis explicativas importantes. Reciprocamente, é estabelecida uma média *a priori* igual à zero para todos os outros coeficientes da equação, β_j em (4.2), indicando que estas variáveis são vistas como menos importantes no modelo.

As variâncias *a priori*, $\sigma_{\beta_i}^2$, especificam incerteza sobre as médias *a priori* $E(\beta_i) = 1$ e $\sigma_{\beta_j}^2$ indicam incerteza sobre as médias *a priori* $E(\beta_j) = 0$. Doan, Litterman e Sims (1984) sugerem uma fórmula para gerar os desvios-padrão como uma função de um pequeno número de hiperparâmetros – θ , ϕ e uma matriz de pesos $w(i, j)$ - uma vez que o modelo VAR apresenta um número alto de parâmetros a serem estimados. Desta forma é possível especificar variâncias *a priori* individuais para um grande número de coeficientes no modelo, usando poucos parâmetros que são denominados hiperparâmetros (LESAGE, 1999). Os valores assumidos pelos hiperparâmetros são cruciais em um modelo Var Bayesiano, pois eles determinam de que forma é permitido que os coeficientes do modelo desviem de suas médias *a priori* e de que forma o modelo VAR Bayesiano se aproxima ou se afasta de um modelo não-Bayesiano (irrestrito), (FELIX; NUNES, 2003).

A especificação do desvio-padrão *a priori* de β_j imposta na variável j , na equação i numa ordem de defasagem k é:

$$\sigma_{i,jk} = \theta w(i,j) k^{-\phi} \left(\frac{\hat{\sigma}_{ui}}{\hat{\sigma}_{uj}} \right) \quad (4.3)$$

Onde $\hat{\sigma}_{ui}$ é o desvio-padrão estimado a partir de uma autorregressão univariada a partir da variável i . O termo $\hat{\sigma}_{ui}/\hat{\sigma}_{uj}$ é um fator escalar que ajusta o desvio-padrão *a priori* para os diferentes ajustes individuais das variáveis i e j . O parâmetro θ é chamado de parâmetro de “overall tightness” e controla a restrição imposta pelo desvio-padrão *a priori* de β_j na primeira defasagem da variável dependente. O termo $k^{-\phi}$ é a função de

decaimento da ordem de defasagem, com $0 \leq \phi \leq 1$ refletindo a taxa de decaimento, que significa que um aumento na ordem de defasagem faz com que o desvio-padrão diminua. Isto significa que variáveis mais defasadas no tempo (com uma ordem de defasagem maior) terão um menor desvio-padrão *a priori*, isto é, podemos afirmar com mais certeza que estas têm menor influência no modelo, corroborando a suposição inicial deste modelo que assume que variáveis mais distantes no tempo representam variáveis menos importantes.

A função $w(i, j)$ especifica a forma de ajuste do desvio-padrão da *priori* para a variável j na equação i , relativo à forma de ajuste da auto defasagem da variável i na equação i .

Os hiperparâmetros usados na *priori* de Minnesota têm valores $\theta = 0.1$ e $\phi = 1.0$. A matriz de pesos usada é:

$$W = \begin{bmatrix} 1 & 0.5 & \dots & 0.5 \\ 0.5 & 1 & & 0.5 \\ \vdots & \vdots & \ddots & \vdots \\ 0.5 & 0.5 & \dots & 1 \end{bmatrix} \quad (4.4)$$

A matriz de pesos impõe uma média *a priori* de $E(\beta_i) = 1$ de forma mais “solta”, isto é, de forma imprecisa. Isto porque a variável dependente defasada em cada equação é considerada uma variável importante. A matriz de pesos também impõe uma média *a priori* igual à zero para os coeficientes das outras variáveis em cada equação de forma menos “solta”, isto é, com uma maior precisão. Isto porque os coeficientes β_j são associados com variáveis consideradas menos importantes no modelo.

Como já citado no início desta sessão, podemos observar que a *priori* de Minnesota trata todas as variáveis do modelo VAR, exceto a variável dependente defasada, de uma forma idêntica e prioriza a especificação da variância *a priori*.

4.2 O Modelo VAR Bayesiano com informação *a priori* espacial

Como vimos na seção anterior, a informação *a priori* de Minnesota trata todas as variáveis, exceto as variáveis dependentes auto-defasadas, da mesma forma. Cada defasagem de uma determinada variável que não a variável dependente recebe o mesmo

peso na matriz de pesos $w(i, j)$. No entanto, quando estamos tratando de dados regionais, que apresentam relações de vizinhança, a informação *a priori* baseada em relações de contigüidade pode se tornar mais interessante quando tratamos de um modelo Bayesiano. Incorporar este tipo de informação *a priori* condiz com o conceito fundamental da econometria espacial de que a localização é importante para a análise de dados regionais que apresentam relações de contigüidade.

Lesage e Krivelyova (1999) incorporam a informação *a priori* espacial que restringe os parâmetros que correspondem às defasagens temporais de variáveis que não apresentam relação de vizinhança impondo uma média *a priori* igual à zero para estas variáveis, isto é, para as variáveis que não são vizinhas à região da variável dependente em uma determinada equação. Então, a *priori* apresentada aqui envolve médias e variâncias baseadas em relações de contigüidade.

Como motivação para a média desta *priori* espacial, considere o modelo autorregressivo espacial apresentado por Anselin (1988), mostrado na equação (4.5). O modelo abaixo não é um modelo Bayesiano. No entanto, é com base na idéia intuitiva deste modelo apresentado por Anselin que é formada a idéia para a informação *a priori* espacial na abordagem Bayesiana. Esta apresentação serve para entendermos como se apresenta a noção de dependência espacial e autocorrelação espacial, noções que desejamos considerar quando utilizamos uma informação *a priori* espacial, baseada nas relações de contigüidade.

$$\begin{aligned} \mathbf{y} &= \rho \mathbf{W}_1 \mathbf{y} + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I_n) \end{aligned} \quad (4.5)$$

Modelos autorregressivos espaciais são usados para relacionarem dados amostrais com referências em diferentes pontos no espaço. Na equação (4.5), \mathbf{y} é um vetor $n \times 1$ de variáveis dependentes e \mathbf{W}_1 é a matriz de pesos espaciais, geralmente contendo as relações de contigüidade de primeira ordem, definidas no Capítulo 2.

O modelo apresentado na Equação (4.5) é um modelo autorregressivo espacial clássico, análogo ao modelo autorregressivo em séries de tempo. É importante notar que este modelo é aplicado aos dados em *cross-section* e é usado aqui para somente fornecer

uma noção intuitiva para a especificação da média *a priori* espacial, apresentada adiante neste mesmo capítulo. Isto é, este modelo não será estimado.

Em aplicações do modelo autorregressivo espacial para dados em *cross-section*, é usada a versão padronizada da matriz contigüidade \mathbf{W}_1 . Esta matriz será montada a partir da definição de relações de contigüidade linear de primeira ordem, apresentada na Seção 2.3.1. Lembrando, a padronização da matriz de contigüidade espacial é feita de tal forma que a soma de cada uma das linhas da matriz somem 1. Isto faz com que a variável espacialmente defasada, $\mathbf{W}_1 y$, represente uma média ponderada de valores a partir das observações vizinhas a y . Isto é, olhando para a expressão $y = \rho \mathbf{W}_1 y$ a partir do modelo autorregressivo apresentado em (4.5), o vetor y que contém os dados em *cross-section* das localizações estruturadas de acordo com a matriz \mathbf{W}_1 , produz uma variável explicativa igual à média das observações a partir das localizações vizinhas.

Com base nisto, podemos iniciar a apresentação da informação *a priori* espacial. Uma média *a priori* para os coeficientes do modelo VAR nas variáveis contíguas auto defasadas será igual a $1/c$, onde c é o número de regiões contíguas a cada variável do modelo. Em outras palavras, esta *priori* espacial é baseada no modelo de passeio aleatório com tendência que pondera as regiões contíguas (KAMARIANAKIS, 2003):

$$y_{it} = \alpha + \sum_{j \in C_i} (1/c_j) y_{jt-1} + \epsilon_{it} \quad (4.6)$$

Onde C_i é o conjunto de regiões c_j vizinhas à região i . Esta *priori* espacial é diferente da *priori* de Minnesota, em que esta última considera menos importante a variável dependente defasada y_{it-1} , usando uma média *a priori* igual à zero, descontando completamente a influência autorregressiva dos valores passados desta variável na equação que descreve a variação de y_{it} . Para as séries de tempo, a *priori* de Minnesota enfatiza um modelo de passeio aleatório com tendência, usando médias *a priori* baseadas em um modelo $y_{it} = \alpha + y_{it-1}$. O termo do intercepto reflete a tendência do

modelo e é estimado usando uma *priori* difusa³⁴. Já a *priori* espacial, é baseada num modelo de passeio aleatório com tendência, que pondera as regiões contíguas, como apresentado na Equação (4.6). Da mesma forma que a *priori* de Minnesota, o intercepto denota a tendência e é estimado usando uma *priori* difusa. É usado um modelo de passeio aleatório baseado na média das influências contíguas autorregressivas de primeira ordem no lugar da variável dependente auto-defasada da *priori* de Minnesota (LESAGE; KRIVELYOVA, 1999).

Assim como na *priori* de Minnesota, é usada uma média *a priori* igual à zero para os coeficientes em todas as defasagens que não sejam as defasagens de primeira ordem. Mas para o caso da *priori* espacial, as defasagens não são tomadas em relação ao tempo e sim, em relação ao espaço. Doan, Litterman e Sims (1984) constatam que especificar médias iguais a zero como informação *a priori* para os coeficientes do modelo apresenta bons resultados em lidar com os problemas de colinearidade em modelos de regressão.

As variâncias *a priori* para os parâmetros do modelo diferenciam-se de acordo com a possibilidade dos coeficientes serem ou não associados com as variáveis a partir de regiões contíguas ou não contíguas e com a ordem de defasagem. A motivação intuitiva para isto são as crenças de que: (1) variáveis não contíguas são menos importantes que variáveis contíguas porque existe uma diminuição da influência à medida que aumenta a distância entre as regiões vizinhas; e (2) defasagens de ordens superiores são menos importantes que defasagens de ordens inferiores, pois existe um declínio da influência das variáveis conforme elas se distanciam no tempo. As observações em séries de tempo a partir de um passado mais distante exercem uma menor influência no valor corrente que as observações em séries de tempo a partir de um passado menos distante, distribuídas no espaço que estamos modelando.

Estas duas crenças a respeito da especificação da variância *a priori* são refletidas por (LESAGE; KRIVELYOVA, 1999, p.305):

(1) Os parâmetros associados com as variáveis em séries de tempo não contíguas são determinados por uma menor variância *a priori*, então uma média *a priori* igual à zero é imposta com mais certeza. A intuição em impor uma média *a priori* com uma menor

³⁴ Uma *priori* difusa é uma distribuição *a priori* do parâmetro em questão que assume igual probabilidade para cada valor possível, com o objetivo de representar a noção de que o analista não tem nenhuma indicação sobre o valor do parâmetro a ser estimado ou simplesmente não deseja assumir nenhum padrão de comportamento ou informação para o parâmetro em questão.

variância *a priori* está baseada na crença de que regiões não contíguas são menos importantes num modelo que considera a localização dos dados.

(2) Uma menor variância *a priori* é imposta às variáveis contíguas em séries de tempo auto defasadas. Então, as médias *a priori* fazem com que as médias das séries de tempo vizinhas sejam iguais, e esta imposição também é feita com um maior grau de certeza. Desta forma, as médias *a priori* refletem a crença de que séries de tempo espacialmente contíguas devem apresentar movimentos comuns no tempo.

(3) Aos parâmetros associados às variáveis não contíguas com ordem de defasagem superior a um serão dados uma variância *a priori* que se torna menor à medida que a ordem de defasagem aumenta. Isto significa que uma média *a priori* igual à zero é imposta com mais certeza conforme a ordem de defasagem aumenta. Então, a crença que a influência decai com o tempo e com variáveis não contíguas é refletida nesta especificação.

(4) Parâmetros associados com todas as ordens de defasagens, exceto com as auto defasagens, das variáveis em séries de tempo terão uma maior variância *a priori*. Neste caso, portanto, as médias *a priori* são impostas de forma mais solta. Isto é motivado pelo fato que não temos um alto grau de confiança na especificação da média *a priori* igual à zero para os valores defasados das variáveis espaciais em séries em tempo. Talvez elas possam exercer alguma influência, fazendo com que a média *a priori* igual a zero não seja apropriada. Ainda é imposto um padrão de decaimento nas variáveis contíguas com ordens de defasagens mais distantes no tempo através da diminuição da variância *a priori* à medida que a ordem de defasagem aumenta. Isto reflete a idéia que a influência das variáveis contíguas e não contíguas decai com o tempo.

Uma forma flexível de estabelecer as médias *a priori* espaciais e os desvios-padrão para a variável j na equação i na ordem de defasagem k é mostrada abaixo:

$$\begin{aligned} \pi(a_{ijk}) &= N(1/c_i, \sigma_c), j \in C, k = 1; i, j = 1, \dots, n \\ \pi(a_{ijk}) &= N(0, \tau\sigma_c/k), j \in C, k = 2, \dots, m; i, j = 1, \dots, n \\ \pi(a_{ijk}) &= N(0, \theta\sigma_c/k), j \notin C, k = 1, \dots, m; i, j = 1, \dots, n \end{aligned} \quad (4.7)$$

Onde

$$0 < \sigma_c < 1 \quad (4.8)$$

$$\tau > 1 \quad (4.9)$$

$$0 < \theta < 1 \quad (4.10)$$

Para as variáveis $j=1, \dots, n$ na equação i que são contíguas à região i ($j \in \mathcal{C}$), a média *a priori* para a variável com ordem de defasagem $k=1$ será igual à média do número de regiões contíguas à variável i e zero para as variáveis não contíguas ($j \notin \mathcal{C}$). O desvio-padrão será igual a σ_c para a primeira defasagem, e obedece a restrição apresentada na Equação (4.8), refletindo uma imposição da média *a priori* com mais certeza, motivada pela contigüidade espacial.

Usamos $\tau\sigma_c/k$ para as defasagens de ordem superior a um e impomos um padrão de decaimento linear nesta variância à medida que a ordem de defasagem aumenta. A Equação (4.9) estabelece a restrição necessária, imposta com um menor grau de crença (de forma mais imprecisa), para assegurar que uma média *a priori* igual à zero seja associada aos parâmetros das variáveis contíguas com defasagens maiores que um. Para os parâmetros das variáveis contíguas com defasagem igual a um é imposta uma restrição com um maior grau de crença (de forma menos imprecisa) para assegurar uma média *a priori* igual à $1/c_i$. Isto é, conforme a ordem de defasagem aumenta a variância da média *a priori* igual à zero, associada às variáveis contíguas, diminui, refletindo a noção intuitiva que variáveis mais distantes no tempo são menos importantes ao modelo. A diminuição da variância com o aumento da ordem de defasagem significa que, conforme as variáveis se tornam distantes no tempo, podemos afirmar com mais certeza que a média *a priori* para estas é igual à zero.

A restrição apresentada em (4.10) impõe uma média *a priori* igual à zero para as regiões não contíguas com mais confiança quando comparada a média *a priori* igual à zero para as regiões contíguas. Isto é, com $0 < \theta < 1$, a variância da média *a priori* igual à zero

para as variáveis não contíguas será menor que a variância da média *a priori* igual à zero para as variáveis contíguas, pois $\sigma^2 > 1$ (ver a especificação da segunda e terceira equações de (4.7)).

Portanto, as fórmulas matemáticas apresentadas acima capturam adequadamente todos os aspectos da motivação intuitiva da especificação das variâncias *a priori*.

4.3 Uma extensão do Modelo VAR Bayesiano: o Modelo VEC Bayesiano

Uma extensão importante de Modelos VAR é o Modelo de Vetor de Correção de Erros, que torna possível a consideração das relações de cointegração entre as variáveis. Tanto o modelo VAR irrestrito como os Modelos VAR Bayesianos apresentados anteriormente neste capítulo não consideram tais relações.

Então, uma crítica aos modelos VAR e VAR Bayesiano baseia-se no fato destes não levarem em consideração, de forma explícita, eventuais relações de longo prazo entre as variáveis. A teoria postula que certas variáveis seguem uma trajetória comum ao longo do tempo, ou, pelo menos, que não divergem continuamente no tempo, isto é, são cointegradas. Como argumentado por Felix e Nunes (2002), na presença de cointegração entre as variáveis, um modelo VAR com mecanismo de correção de erros – chamado de Modelo Vetorial de Correção de Erros - pode apresentar melhores resultados de previsão para horizontes mais longos que os modelos VAR e VAR Bayesiano. Se as variáveis em nível implicam uma relação de cointegração, então um VAR em primeira diferença será mal especificado. Então, o VAR estimado com dados cointegrados será mal especificado se os dados forem diferenciados e, se os dados forem usados em nível, eles omitirão restrições importantes. Um modelo Vetorial Bayesiano de Correção de Erros – abreviado na literatura por VECM Bayesiano – pode ser usado, então, para combinar as vantagens do modelo VAR Bayesiano com os benefícios em levar em consideração relações de forma explícita de longo prazo em exercícios de previsão.

A importância do mecanismo de correção de erros na literatura sobre cointegração deriva da demonstração do trabalho de Engle e Granger (1987) que, se duas variáveis

são integradas de primeira ordem e são cointegradas, elas são modeladas como se tivessem sido geradas a partir do Modelo de Correção de Erros. Este trabalho de Engle e Granger enfatiza o uso das relações de equilíbrio de longo prazo presentes na teoria econômica para modelos empíricos de séries de tempo, com o objetivo de explicar a dinâmica de curto prazo dos dados econômicos. Conclui-se que o Modelo de Correção de Erros apresenta melhores resultados de previsão de curto prazo e, conseqüentemente, produzirão melhores resultados também de previsão de longo prazo, uma vez que leva em consideração a relação de equilíbrio entre as variáveis.

Um motivo para a comparação entre os Modelos VAR e VECM está na contribuição para a qualidade de previsão acrescentada pela variável de correção de erros no Modelo de Vetor de Correção de Erros. O fato de o modelo VECM englobar um modelo VAR em primeira diferença que não contém a variável de correção de erros é usado para desenvolver a contribuição marginal da qualidade de previsão para esta variável adicional de equilíbrio de longo prazo. Lesage (1990) demonstra através de experimentos, que a adição desta única variável aos modelos VAR Bayesianos produz uma redução do erro absoluto médio de previsão e também uma redução na dispersão dos erros. A melhora no desempenho de previsão fornecido pelo modelo VECM Bayesiano é uma função crescente do horizonte de previsão, isto é, a precisão das previsões usando horizontes mais longos a partir de um modelo VECM Bayesiano é maior que as dos modelos que competem com ele.

Um modelo VECM Bayesiano pode ser utilizado para combinar as vantagens de um modelo VAR Bayesiano com os benefícios de se considerar explicitamente as relações de longo prazo em exercícios de previsão. Felix e Nunes (2002) estimaram as relações de longo prazo, utilizando tanto o método de Engle e Granger como o método de Johansen. No método de Engle e Granger foi imposto, implicitamente, um único vetor cointegrante, hipótese esta que pode ser contrariada pelo método de Johansen. Em seguida, os termos de correção de erros foram adicionados ao modelo VAR Bayesiano com as variáveis em primeira diferença. Eles concluem que, apesar do suporte teórico, verifica-se empiricamente que eventuais ganhos da imposição de relações de longo prazo pode ser algo contraditório. Isto é, empiricamente, imposições de restrições de longo prazo através de mecanismos de correção de erros podem não melhorar o desempenho das previsões no contexto dos modelos VAR Bayesianos.

Com base nisto, se faz necessária uma comparação entre modelos VAR Bayesianos e VECM Bayesianos com o objetivo de avaliar o desempenho relativo destes. Uma forma de avaliar este desempenho relativo é através da comparação de uma medida de erros de previsão cometidos por cada um dos modelos, num horizonte de previsão suficientemente longo, para que as relações de longo prazo possam influenciar as previsões, no caso destas serem relevantes. A medida utilizada por Felix e Nunes (2002) é o erro quadrático médio, isto é, a média dos erros ao quadrado. Portanto, o Capítulo 4 será destinado não só à estimação dos modelos apresentados neste capítulo como também terá foco na comparação dos modelos aqui propostos.

4.4 O Modelo Vetorial Bayesiano de Correção de Erros – VECM Bayesiano

Considere um vetor y_t , com n séries de tempo que são integradas de primeira ordem, isto é, são $I(1)$. Uma série $I(1)$ é uma série que necessita ser diferenciada uma vez para que esta se torne um processo puramente estacionário não-determinístico com média zero. O vetor z_t é chamado de vetor de cointegração se existe uma matriz α de dimensão $n \times r$ tal que:

$$z_t = \alpha' y_t \quad (4.11)$$

Em outras palavras, dado um vetor $I(1)$ y_t de séries de tempo. Se existe uma combinação linear $\alpha' y_t$ que é $I(0)$, isto é, uma combinação linear estacionária, então as séries de tempo que compõem o vetor y_t são ditas cointegradas, com matriz de cointegração α . Interpretando $\alpha' y_t = 0$ como um equilíbrio de longo prazo, a cointegração implica que os desvios a partir do equilíbrio são estacionários, com variância finita, mesmo que cada uma das séries do vetor y_t seja não estacionária e tenham variância infinita (ENGLE; GRANGER, 1987).

Engle e Granger (1987) estabelecem que, se duas ou mais séries em y_t são cointegradas, então existe uma representação de correção de erros que toma a seguinte forma:

$$\Delta y_t = A(1)\Delta y_t + \gamma z_{t-1} + \varepsilon_t \quad (4.12)$$

Onde γ é uma matriz de coeficientes $n \times r$ de posto r , z_{t-1} é um vetor de dimensão $r \times 1$ baseado em $r \leq n - 1$ relações de erros de equilíbrio, $z_t = \alpha' y_t$, a partir de (4.11), e ε_t são os resíduos estacionários multivariados. O Modelo de Correção de Erros apresentado na Equação (4.12) é simplesmente um Modelo VAR em primeira diferença com r termos defasados de correção de erros (z_{t-1}) incluídos em cada equação do modelo.

A estimação do Modelo de Correção de Erros, assim como a estimação da sua variação Bayesiana, é feita em dois passos. Primeiro, as relações de longo prazo são estimadas usando as metodologias de Engle e Granger ou de Johansen. Na aproximação de Engle e Granger, é possível estimar ou testar uma única relação de cointegração, que pode ser interpretada como uma combinação linear de todos os vetores que cointegram no espaço de cointegração (FELIX; NUNES, 2002). Por exemplo, no caso de um modelo com somente duas séries, y_t e x_t , o procedimento de Engle e Granger pode ser usado para determinar a variável de cointegração que adicionamos ao Modelo VAR em primeira diferença para torná-lo um Modelo de Correção de Erros. O primeiro passo envolve estimar uma regressão $y_t = \theta + \alpha x_t$ para determinar as estimativas de α e z_t . O segundo passo envolve testes em z_t para determinar se este é estacionário, $I(0)$. Se este é o caso, a condição $y_t = \theta + \alpha x_t$ é interpretada como uma relação de equilíbrio entre duas séries.

Já na metodologia de Johansen, depois do teste de posto, ou de dimensão, do espaço de cointegração, é possível que mais de um vetor de cointegração seja estimado. Na maioria dos modelos de correção de erros estão envolvidas mais que duas variáveis e, então, o procedimento de Engle e Granger descrito acima deve ser generalizado. A aproximação de Johansen então é mais indicada. Johansen (1988) fornece esta generalização que toma a forma de um teste de razão Verossimilhança. Este procedimento consiste de um teste estatístico para a determinação do número r de relações de cointegração entre as n variáveis em y_t , assim como um conjunto de r

vetores de cointegração que serão usados para construir as variáveis de correção de erros no Modelo de Vetor de Correção de Erros.

Como motivação para o entendimento do procedimento de Johansen, considere a reparametrização do modelo de correção de erros apresentada abaixo:

$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + \dots + \Gamma_{k-1} \Delta y_{t-k+1} - \Psi y_{t-k} + \varepsilon_t \quad (4.13)$$

Onde $\Psi = (I_n - A_1 - A_2 - \dots - A_k)$. Se a matriz Ψ é uma matriz nula (tem posto igual à zero, portanto), não existem relações de cointegração entre as variáveis em y_t . Se a matriz Ψ tem posto completo, então existem n relações de equilíbrio de longo prazo e, portanto, todas as variáveis do modelo são cointegradas. Para os casos em que a matriz Ψ tem posto $r < n$, temos r relações de cointegração. O procedimento de Johansen fornece dois testes para determinar o número de relações de cointegração linearmente independentes entre as séries do vetor y_t , que denominamos r nesta apresentação. Ambos os testes são baseados na decomposição entre autovalores e autovetores da matriz Ψ , construída a partir das correlações entre Δy_t e y_{t-k} , levando em consideração que os testes são baseados em estimativas estocásticas. Os testes estatísticos são denominados “estatística traço” e “estatística do autovalor máximo” (LESAGE, 1999, cap.5, p.118).

Depois de determinar o número de relações de cointegração, podemos usar estes resultados junto com os autovetores determinados na decomposição da matriz Ψ para formar um conjunto de variáveis de correção de erros. Estas são construídas usando y_{t-1} - o nível de y defasado um período – multiplicado pelos r autovetores associados com as relações de cointegração para formar as r variáveis de cointegração.

A variação Bayesiana do Modelo de Vetor de Correção de Erros pode ser especificada da mesma forma que o Modelo VAR Bayesiano. As especificações das informações *a priori* serão as mesmas, de acordo com as crenças baseadas em determinada situação. A diferença ocorrerá somente na estimação dos diferentes modelos. A estimação do Modelo de Vetor de Correção de Erros Bayesiano ocorre da mesma forma que a estimação do Modelo VAR Bayesiano em primeira diferença e as r variáveis de

correção de erros são adicionadas ao modelo de acordo com a determinação pelo procedimento de Johansen.

4.5 A Função Verossimilhança para os Modelos VAR e VEC Bayesianos

Para completar as especificações dos modelos Bayesianos apresentados aqui, tornando possível a estimação via Gibbs Sampling, é necessário apresentar as formas da função verossimilhança. Através delas, poderemos amostrar das distribuições *a posteriori*.

Com este objetivo, vamos considerar o modelo VAR escrito como:

$$y_t = \sum_{i=1}^p y_{t-i} A_i + x_t C + u_t \quad (4.14)$$

onde y_t é o vetor linha com as m variáveis de interesse observadas no tempo t e x_t é o vetor linha com q variáveis exógenas influenciando y_t . A_i e C são as matrizes de parâmetros, de dimensão $m \times m$ e $q \times m$ respectivamente.

Para a discussão técnica e para chegarmos à função verossimilhança utilizada no processo de estimação, precisamos escrever a Equação (4.14) da seguinte forma:

$$y_t = z_t \Gamma + u_t \quad (4.15)$$

onde $z_t = \{x_t, y_{t-1}, \dots, y_{t-p}\}$ e o $k = q - pm$ pela matriz Γ de ordem m é dada por $\{C', A_1', \dots, A_p'\}'$. Agrupando de forma convencional os vetores linha y_t , z_t e u_t para $t = 1, \dots, T$ em Y , Z , U , temos o modelo de regressão multivariado:

$$Y = Z \Gamma + U \quad (4.16)$$

Considerando o subscrito i denotando o i -ésimo vetor coluna, temos a equação $y_i = Z\gamma_i + u_i$ para a variável i . Os vetores obtidos agrupando as colunas de Y , Γ e U são os vetores y , γ e u e o sistema pode ser escrito como $y = (I \otimes Z)\gamma + u$.

É assumido que $u \sim N(0, \Psi \otimes I)$, isto é, u_i é assumido sendo i.i.d $N(0, \Psi)$. Então, a verossimilhança é dada por:

$$L(\gamma, \Psi) \propto |\Psi|^{-T/2} \exp\{-tr [(-Z\Gamma)' \Psi^{-1}(Y - Z\Gamma)]/2\} \quad (4.17)$$

Depois de algumas manipulações, encontramos que:

$$\begin{aligned} L(\gamma, \Psi) &\propto |\Psi|^{-T/2} \exp\left\{-\frac{1}{2}(\gamma - \hat{\gamma})'(\Psi^{-1} \otimes Z'Z)(\gamma - \hat{\gamma}) - \frac{1}{2}tr \left[\Psi^{-1}(Y - Z\hat{\Gamma})'(Y - Z\hat{\Gamma}) \right]\right\} \\ &= |\Psi|^{-k/2} \exp\left\{-\frac{1}{2}(\gamma - \hat{\gamma})'(\Psi^{-1} \otimes Z'Z)(\gamma - \hat{\gamma})\right\} \times |\Psi|^{-(T-k)/2} \\ &\quad \times \exp\left\{-\frac{1}{2}tr \left[\Psi^{-1}(Y - Z\hat{\Gamma})'(Y - Z\hat{\Gamma}) \right]\right\} \\ &\propto N(\gamma | \hat{\gamma}, \Psi \otimes (Z'Z)^{-1}) \times IW(\Psi | (Y - Z\hat{\Gamma})'(Y - Z\hat{\Gamma}), T - k - m - 1) \end{aligned}$$

Isto é, o produto de uma densidade Wishart inversa para Ψ por uma densidade Normal para γ condicional em Ψ (KADIYALA; KARLSSON, 1997).

5 Uma aplicação para o Modelo VEC Bayesiano: o caso da Produção Industrial no Brasil

Este capítulo tem como objetivo aplicar dois dos modelos apresentados no Capítulo 4, utilizando dados da produção industrial para o Brasil. Os dados utilizados são referentes a oito estados brasileiros. São eles: Bahia, Minas Gerais, Espírito Santo, Rio de Janeiro, São Paulo, Paraná, Santa Catarina e Rio Grande do Sul. As séries usadas neste capítulo compreendem os índices de produção física calculada pelo IBGE para estes estados, cada uma consistindo de 63 observações, sazonalmente ajustada, no período entre janeiro de 1991 e setembro de 2006³⁵. Para esta aplicação, serão usados dados trimestrais.

A Figura 5.1 apresenta o gráfico das séries dos estados em função do tempo e, já a partir deste gráfico, através da simples observação, temos indícios de cointegração entre as séries. Então, vamos atentar para este fato na aplicação dos modelos, uma vez que confirmadas as relações de cointegração, o modelo adequado para tratar destes dados não será o modelo VAR e sim o modelo VEC.

Outro assunto que devemos levar em consideração, é a ordem de defasagem temporal utilizada no modelo, a fim de que o mesmo seja bem especificado. Para a aplicação do Modelo VAR Bayesiano com informação *a priori* espacial ou a informação *a priori* de Minnesota, utilizaremos a mesma forma de determinação da ordem de defasagem que nos modelos clássicos. Uma forma bastante usada para determinar a ordem de defasagem temporal que deve ser considerada num modelo específico é realizar testes estatísticos com modelos com ordens de defasagens maiores e compará-los a modelos com ordens de defasagens menores. Olhamos para modelos com ordens de defasagens maiores como modelos irrestritos e modelos com ordens de defasagens menores como modelos restritos, é construído um teste estatístico de razão verossimilhança para testar a significância em impor restrições.

³⁵ Os dados brutos dos índices foram dessazonalizados pela razão de média móvel multiplicativa, já aplicada pelo IBGE, e o índice trimestral foi obtido através da média simples do índice mensal.

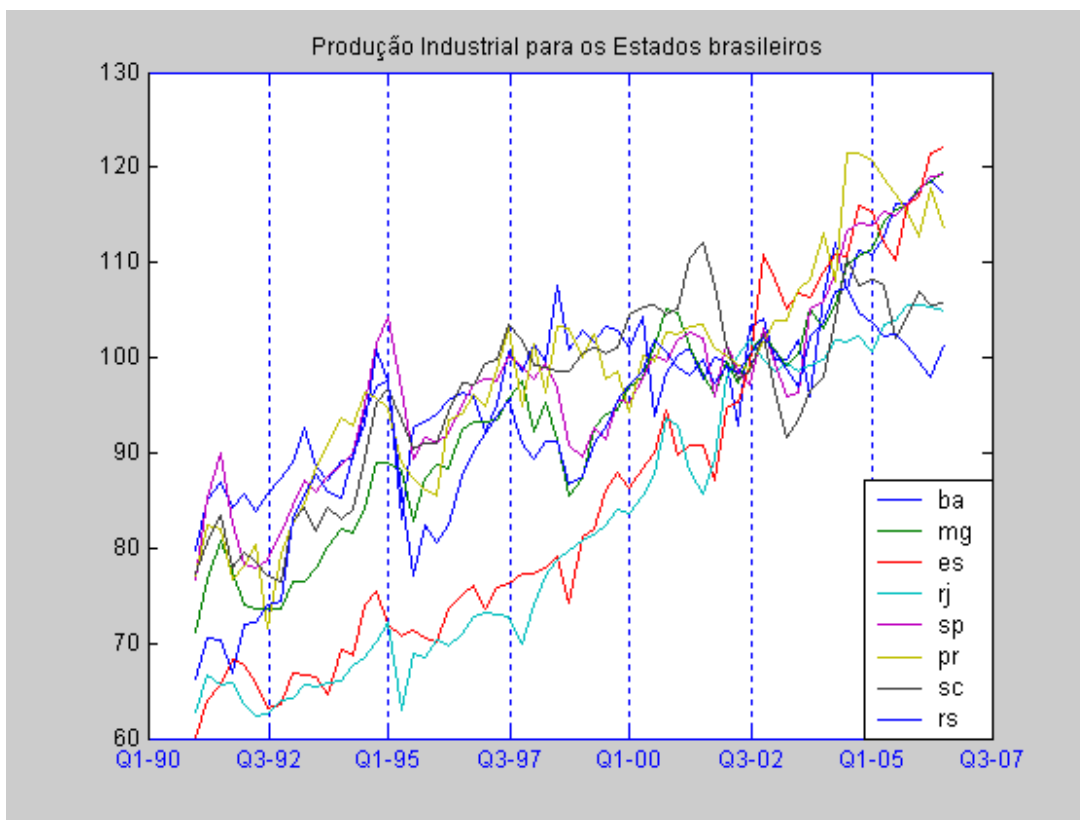


Figura 5.1 – Séries de Produção Industrial trimestrais entre 1991:1 a 2006:3 para os Estados Brasileiros a partir dos dados do IBGE (2007)

Se as restrições são associadas com uma diminuição de qualidade estatisticamente significativa no modelo estimado, concluímos que um modelo de defasagem mais longa é mais apropriado, rejeitando modelos de defasagens menores.

Especificamente, o Teste da Razão de Verossimilhança tem distribuição assintótica qui-quadrado com o número de graus de liberdade igual ao número de restrições impostas:

$$LR = (T - c)(\log|\Sigma_r| - \log|\Sigma_u|) \quad (5.1)$$

onde T é o número de observações, c é o fator de correção para os graus de liberdade proposto por Sims (1980)³⁶, $|\Sigma_r|$ e $|\Sigma_u|$ denotam o determinante da matriz de covariância dos erros a partir do modelo restrito e irrestrito, respectivamente. A partir deste teste de razão-Verossimilhança encontramos o número ótimo de defasagens

³⁶ O fator de correção proposto por Sims, c , é o número de variáveis em cada equação irrestrita do modelo VAR.

utilizadas no modelo proposto. A partir deste critério, obtemos para nossos dados um número de defasagens de ordem 4. Os resultados são apresentados no Anexo 1.

Analisando as propriedades de cointegração dos dados a partir da representação de um VAR linear usando o procedimento de máxima verossimilhança proposto por Johansen, no período total da amostra, o modelo não rejeita a hipótese que existem cinco relações de cointegração para y_t . Os testes são realizados utilizando a estatística-traço e a estatística de autovalor máximo. A estatística-traço fornece que o número de relações de cointegração é igual a cinco, enquanto a estatística de autovalor máximo fornece que o número de relações de cointegração é igual a três, ambas analisadas com um nível de significância de 95%. Os valores referentes aos testes são apresentados também no Apêndice A.

5.1 A escolha do modelo

De acordo com as características apresentadas acima, referente aos dados em que estamos interessados em analisar, podemos então escolher o modelo que parece mais adequado, dentre aqueles apresentados no Cap. 4.

O objetivo é estimar a produção industrial por estado, levando em consideração tanto as relações espaciais quanto as relações temporais. Desta forma, a estrutura de um modelo vetorial parece atender este objetivo. Especificando cada elemento do vetor de variáveis dependentes de um modelo vetorial autorregressivo como a produção industrial de um estado brasileiro, levamos em consideração a estrutura espacial mantida entre os dados, além de considerar também a estrutura temporal. Isto é, olhando inicialmente um modelo VAR, temos o modelo estruturado da seguinte forma:

$$\begin{bmatrix} y_{BA,t} \\ y_{MG,t} \\ y_{ES,t} \\ y_{RJ,t} \\ y_{SP,t} \\ y_{PR,t} \\ y_{SC,t} \\ y_{RS,t} \end{bmatrix} = \begin{bmatrix} A_{BA,BA}(L) & \dots & A_{BA,RS}(L) \\ \vdots & \ddots & \vdots \\ A_{RS,BA}(L) & \dots & A_{RS,RS}(L) \end{bmatrix} \begin{bmatrix} y_{BA,t} \\ y_{MG,t} \\ y_{ES,t} \\ y_{RJ,t} \\ y_{SP,t} \\ y_{PR,t} \\ y_{SC,t} \\ y_{RS,t} \end{bmatrix} + \begin{bmatrix} C_{BA} \\ C_{MG} \\ C_{ES} \\ C_{RJ} \\ C_{SP} \\ C_{PR} \\ C_{SC} \\ C_{RS} \end{bmatrix} + \begin{bmatrix} \varepsilon_{BA,t} \\ \varepsilon_{MG,t} \\ \varepsilon_{ES,t} \\ \varepsilon_{RJ,t} \\ \varepsilon_{SP,t} \\ \varepsilon_{PR,t} \\ \varepsilon_{SC,t} \\ \varepsilon_{RS,t} \end{bmatrix} \quad (5.2)$$

A estrutura espacial é considerada a partir do momento em que cada elemento do vetor de variáveis dependentes corresponde a um estado brasileiro. O modelo apresentado desta forma trata da estimação da série de um estado brasileiro a partir dos passados das séries de outros estados e também da sua própria série observada no passado. Por exemplo, a produção industrial no estado de São Paulo no instante atual, $y_{SP,t}$, é explicada por ela mesma no passado e pelo passado da produção industrial dos outros sete estados. Portanto, esta estruturação do modelo cumpre com nosso objetivo.

Algumas características dos dados em questão devem ser consideradas. Pelo fato de estarmos restritos a um conjunto de dados que trata somente da produção industrial, é esperado que existam muitas relações de cointegração entre as séries. Isto já foi validado pelos testes apresentados na introdução deste capítulo e os resultados são apresentados no Apêndice A. Então, estas devem ser consideradas no modelo. Desta forma, ao invés de usar o modelo VAR, apresentado na equação (5.2), devemos usar o modelo VEC. Além de considerar as relações de cointegração no modelo, o modelo VEC trabalha com os dados em primeira diferença. As séries de produção industrial utilizadas aqui são todas integradas de primeira ordem, isto é, são estacionárias em primeira diferença. Então, as séries serão inseridas no modelo em primeira diferença, o que significa que usaremos as séries na sua forma de taxa de crescimento. O modelo VEC pode ser visto como uma extensão do modelo VAR, como mostra a equação abaixo:

$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + \dots + \Gamma_{k-1} \Delta y_{t-k+1} - \Psi y_{t-k} + \varepsilon_t \quad (5.3)$$

Aqui, Δy_t é o vetor das variáveis dependentes, isto é, o vetor das séries em primeira diferença de produção industrial por estados e $\Psi = (I_n - A_1 - A_2 - \dots - A_k)$ é a matriz construída a partir das correlações entre as variáveis dependentes e as variáveis independentes defasadas, isto é, entre o vetor Δy_t e o vetor y_{t-k} . O posto desta matriz determina o número de relações de cointegração consideradas no modelo e, neste caso, esta matriz tem posto $r \leq n$, onde r é igual a 5 e n é igual a 8 neste caso, isto é, temos 8 variáveis dependentes sendo que, entre elas, existem 5 autovetores de relações de cointegração que entram no modelo para formar o vetor de correção de erros.

Estas relações de cointegração são conhecidas também como relações de longo prazo entre as variáveis uma vez que as séries em questão apresentam um comportamento comum ao longo do tempo. Então, o coeficiente estimado do vetor de cointegração representa a relação de equilíbrio de longo prazo entre as variáveis cointegradas.

Além da escolha do modelo que cumpra com o objetivo de tratar das relações espaciais e temporais propostas aqui, é necessário escolher a ordem de defasagem temporal que mais se apresenta adequada para os dados em questão. A partir do teste de razão Verossimilhança apresentado na introdução, utilizaremos uma ordem de defasagem temporal igual a quatro. Isto significa que, em cada equação do modelo VEC, as variáveis independentes são defasadas até a quarta ordem. Por exemplo, na equação que tem a produção industrial do estado de São Paulo como variável dependente, $y_{SP,t}$, temos como variáveis independentes $y_{SP,t-1}$, $y_{SP,t-2}$, $y_{SP,t-3}$, $y_{SP,t-4}$, além de y_{it-1} , y_{it-2} , y_{it-3} , y_{it-4} , com $i = BA, MG, ES, RJ, PR, SC, RS$.

Portanto, o modelo escolhido será o modelo VEC de ordem quatro, com oito variáveis correspondentes à variação da produção industrial em cada estado brasileiro já citado, e inclui cinco relações de cointegração entre as variáveis.

5.2 A escolha do método de estimação

O modelo escolhido, com oito variáveis com quatro defasagens temporais e cinco relações de cointegração, além da constante incluída, produz 38 variáveis independentes

(8 variáveis multiplicadas por 4 defasagens temporais mais cinco termos a partir das relações de cointegração mais uma constante incluída no modelo) em cada uma das oito equações do modelo, somando um total de 304 coeficientes a serem estimados. Para estimar um modelo VEC com esta estrutura é necessário uma grande amostra de observações, que cobrem um período de tempo longo e estas nem sempre estão disponíveis. Além disto, as variáveis independentes defasadas no tempo podem produzir altas correlações que prejudicam a precisão da estimativa dos parâmetros. Uma forma de contornar este problema é utilizar o método Bayesiano de estimação.

A utilização da informação *a priori* é uma forma de agregar informações que não são consideradas pelo conjunto de dados em si, mas que consideramos importantes para tratar de um problema específico. Isto é, é uma forma de agregar algum tipo de informação que está fora dos dados observados, ao processo de estimação. No contexto do objetivo proposto neste trabalho, podemos utilizar este instrumento para considerar a informação da distribuição espacial dos dados no processo de estimação.

O modelo VEC estruturado na forma apresentada na Seção 5.1 e estimado, por exemplo, por Mínimos Quadrados Ordinários, trata todas as variáveis da mesma forma. Cada defasagem de uma determinada variável recebe o mesmo peso no processo de estimação. Através do método de estimação Bayesiano, podemos agregar o conceito de que a localização é importante através da utilização de uma informação *a priori* que considera as relações de contigüidade no processo de estimação. Isto também pode ser feito num contexto não Bayesiano, utilizando a matriz de contigüidade diretamente no modelo. No entanto, isto aumentaria o número de coeficientes estimados e, estimando um modelo que considera muitas localizações, o processo de estimação não Bayesiano poderia se tornar inviável. Desta forma, agregamos as informações regionais ao processo de estimação Bayesiano. Para tanto, será utilizada a informação *a priori* espacial, apresentada no Capítulo 4, Seção 4.2 e também a informação *a priori* de Minnesota, apresentada no Capítulo 4, Seção 4.1, numa forma modificada. Como já apresentado ali, estas informações *a priori* não tratam todas as variáveis da mesma forma no processo de estimação. Ela pondera as variáveis de forma que variáveis que mantêm uma relação de vizinhança com a variável dependente em questão tenham uma maior importância na estimativa desta variável dependente e variáveis que não mantêm uma relação de vizinhança, uma importância menor. Portanto, no processo de estimação

do modelo VEC Bayesiano, têm-se estimativas ponderadas para os parâmetros do modelo, de acordo com a localização dos dados.

5.2.1 O Modelo

Além da utilização da informação *a priori* espacial, existem outras formas de ponderar as variáveis no processo de estimação. Estas formas para as informações *a priori* dependerão muito do problema tratado pelo modelo e do objetivo do problema aplicado aos dados. Além das diferentes formas de incorporar o mesmo tipo de informação ao processo de estimação, podemos também incorporar o mesmo tipo de informação de diferentes formas. Por exemplo, podemos utilizar uma informação *a priori* que considera as relações de contigüidade de diferentes maneiras além da apresentada no Capítulo 4, Seção 4.2. Podemos utilizar a informação *a priori* de Minnesota numa forma modificada, ponderando os parâmetros do modelo de acordo com a definição da informação *a priori* de Minnesota apresentada no Capítulo 4, Seção 4.1. Esta informação *a priori* de Minnesota modificada será apresentada posteriormente neste mesmo capítulo, mas adiantamos aqui que será uma forma alternativa de considerar as relações de vizinhança do problema de estimação do modelo VEC Bayesiano que, assim como a informação *a priori* espacial, utiliza as relações de contigüidade para ponderar as estimativas para os coeficientes.

Portanto, neste capítulo serão apresentados os resultados do modelo VEC Bayesiano, que utilizam duas formas distintas para agregar informação dos dados regionais. A primeira forma de agregar esta informação é utilizar o modelo VEC Bayesiano com informação *a priori* baseada nas relações de contigüidade de primeira ordem padronizada. Chamaremos este modelo de VEC Bayesiano com informação *a priori* espacial, e suas características de formação foram apresentadas na Seção 4.2. A segunda forma de agregar esta informação é utilizar o modelo VEC Bayesiano com informação *a priori* de Minnesota, modificando a forma em que se apresenta a matriz de ponderação para os coeficientes, apresentada na Equação (4.4), de tal forma que esta matriz de ponderação reflita as relações de vizinhança entre os dados de uma forma alternativa àquela apresentada pela informação *a priori* espacial. Chamaremos este modelo de VEC

Bayesiano com informação *a priori* de Minnesota e suas características principais foram apresentadas na Seção 4.1.

5.2.2 A matriz de contigüidade da informação *a priori* espacial

A especificação das médias e variâncias *a priori* que entram no processo de estimação do VEC Bayesiano com informação *a priori* espacial é apresentada nas Equações (4.7). A matriz de contigüidade de primeira ordem entra na especificação das médias e variâncias *a priori*. Para o caso da produção industrial entre os estados brasileiros considerados aqui, temos a seguinte matriz de contigüidade de primeira ordem:

	BA	MG	ES	RJ	SP	PR	SC	RS
BA	0	1	1	0	0	0	0	0
MG	1	0	1	1	1	0	0	0
ES	1	1	0	1	0	0	0	0
RJ	0	1	1	0	1	0	0	0
SP	0	1	0	1	0	1	0	0
PR	0	0	0	0	1	0	1	0
SC	0	0	0	0	0	1	0	1
RS	0	0	0	0	0	0	1	0

Figura 5.2 – Matriz de contigüidade espacial de primeira ordem não padronizada entre os oito estados brasileiros considerados neste capítulo.

A matriz de contigüidade acima considera as relações de vizinhança entre os estados brasileiros. Por exemplo, o estado de São Paulo tem os estados de Minas Gerais, Rio de Janeiro e Paraná como vizinhos, isto é, estes três estados fazem fronteira com o estado de São Paulo. O estado de Minas Gerais tem os estados da Bahia, Espírito Santo, Rio de Janeiro e São Paulo como vizinhos. E, desta forma, podemos olhar para todos os outros estados. Cada um dos pares de estados que mantém uma relação de vizinhança recebe um valor igual a 1 e pares de estados que não mantém uma relação de vizinhança recebe um valor igual a zero na matriz. Então, a linha correspondente ao estado de São Paulo terá valor igual a 1 nas colunas referentes à Minas Gerais, Rio de Janeiro e Paraná e a

linha correspondente a Minas Gerais terá valor igual a 1 nas colunas referentes à Bahia, Espírito Santo, Rio de Janeiro e São Paulo. Desta forma, esta matriz será uma matriz simétrica e é dado um valor igual a zero à diagonal principal por convenção³⁷. A partir desta matriz de contigüidade de primeira ordem não padronizada, podemos construir uma matriz de contigüidade de primeira ordem padronizada, mostrada abaixo:

	BA	MG	ES	RJ	SP	PR	SC	RS
BA	0	1/2	1/2	0	0	0	0	0
MG	1/4	0	1/4	1/4	1/4	0	0	0
ES	1/3	1/3	0	1/3	0	0	0	0
RJ	0	1/3	1/3	0	1/3	0	0	0
SP	0	1/3	0	1/3	0	1/3	0	0
PR	0	0	0	0	1/2	0	1/2	0
SC	0	0	0	0	0	1/2	0	1/2
RS	0	0	0	0	0	0	1	0

Figura 5.3 – Matriz de contigüidade espacial de primeira ordem padronizada entre os oito estados brasileiros utilizada na estimação do VEC Bayesiano com informação *a priori* espacial.

A matriz de contigüidade espacial de primeira ordem padronizada considera as mesmas relações de vizinhança que a matriz de contigüidade de primeira ordem não padronizada. A diferença aqui é que esta matriz é padronizada de tal forma que a soma de cada uma de suas linhas seja igual a 1. Por exemplo, a linha correspondente ao estado de Minas Gerais terá valor igual a $\frac{1}{4}$ nas colunas referentes a Bahia, Espírito Santo, Rio de Janeiro e São Paulo, de tal forma que a soma desta linha seja igual a 1. Desta forma, com todos os outros estados. Esta será a matriz de contigüidade que entrará na construção das médias e variâncias *a priori* no processo de estimação do modelo VEC Bayesiano com informação *a priori* espacial.

5.2.3 A matriz de pesos da informação *a priori* de Minnesota

A especificação das médias e variâncias *a priori* que entram no processo de estimação do VEC Bayesiano com informação *a priori* de Minnesota é apresentada nas Equações (4.2) e (4.3). Como apresentado na Seção 4.1, *a priori* de Minnesota utiliza uma matriz

³⁷ Ver Capítulo 2, seção 2.4.1, definições de matrizes de contigüidade espacial.

de pesos que especifica a forma que o modelo ajusta as variáveis no processo de estimação, onde esta matriz de pesos define as variáveis dependentes autodefadas como variáveis mais importantes no modelo e as outras variáveis dependentes como menos importantes³⁸ no modelo. Esta matriz de pesos na sua forma original é apresentada abaixo:

	BA	MG	ES	RJ	SP	PR	SC	RS
BA	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5
MG	0.5	1	0.5	0.5	0.5	0.5	0.5	0.5
ES	0.5	0.5	1	0.5	0.5	0.5	0.5	0.5
RJ	0.5	0.5	0.5	1	0.5	0.5	0.5	0.5
SP	0.5	0.5	0.5	0.5	1	0.5	0.5	0.5
PR	0.5	0.5	0.5	0.5	0.5	1	0.5	0.5
SC	0.5	0.5	0.5	0.5	0.5	0.5	1	0.5
RS	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1

Figura 5.4 – Matriz apresentada na equação (4.4), utilizada como a matriz de pesos na definição da informação *a priori* de Minnesota na sua forma original.

A matriz de pesos apresentada na forma da Figura 5.4 significa que a informação *a priori* de Minnesota considera peso igual a 1 para as relações autorregressivas no processo de estimação do modelo e um peso menor, igual a 0.5, para todas as outras variáveis, independentemente delas manterem ou não uma relação de vizinhança umas com as outras. No entanto, podemos mudar a especificação desta matriz de tal forma que as relações de vizinhança sejam consideradas. Então, teremos o mesmo modelo especificado na forma apresentada na Seção 4.1, mas com pesos modificados, de tal forma que estes reflitam as relações de vizinhança. Para tanto, é construída uma matriz que considera, da mesma forma que a matriz original, as relações autorregressivas como as relações mais importantes no processo de estimação, assumindo valores iguais a 1 na diagonal principal. Além disso, também consideramos como mais importantes as relações de vizinhança entre os estados, isto é, os estados vizinhos também terão peso igual a 1 nesta nova matriz construída. Por exemplo, Minas Gerais mantém relação de vizinhança com os estados da Bahia, Espírito Santo, Rio de Janeiro e São Paulo. Então, a linha referente ao estado de Minas Gerais terá valor igual a 1 nas colunas correspondentes a seus estados vizinhos. Outra modificação feita nesta nova definição

³⁸ Ver seção 3.2 para definição exata de como a informação *a priori* de Minnesota considera as variáveis dependentes no processo de estimação na sua forma original.

será que o peso atribuído aos estados que não mantêm uma relação de vizinhança será igual a 0.1, refletindo a crença que estados não vizinhos são menos importantes no processo de estimação, mas ainda assim entram no processo de estimação. Por exemplo, os estados do Paraná, Santa Catarina e Rio Grande do Sul não mantêm uma relação de vizinhança com Minas Gerais e, portanto, as colunas correspondentes a estes estados na linha relacionada a Minas Gerais têm valor igual a 0.1. A matriz construída com base nestas especificações é apresentada abaixo:

	BA	MG	ES	RJ	SP	PR	SC	RS
BA	1	1	1	0.1	0.1	0.1	0.1	0.1
MG	1	1	1	1	1	0.1	0.1	0.1
ES	1	1	1	1	0.1	0.1	0.1	0.1
RJ	0.1	1	1	1	1	0.1	0.1	0.1
SP	0.1	1	0.1	1	1	1	0.1	0.1
PR	0.1	0.1	0.1	0.1	1	1	1	0.1
SC	0.1	0.1	0.1	0.1	0.1	1	1	1
RS	0.1	0.1	0.1	0.1	0.1	0.1	1	1

Figura 5.5 – Matriz de contigüidade espacial entre os oito estados brasileiros utilizada para estimação do modelo VEC Bayesiano com informação *a priori* de Minnesota.

Na próxima seção serão apresentados os resultados empíricos referentes ao modelo VEC Bayesiano, utilizando as duas formas distintas apresentadas aqui de considerar as informações *a priori*.

Dentre os métodos de estimação Bayesianos, será usado o método iterativo MCMC por Gibbs Sampling, descrito no Capítulo 3, seção 3.4.2.1, para estimar o modelo proposto.

5.3 Resultados Empíricos

Os resultados apresentados nesta seção são, portanto, estimados a partir do método de estimação proposto acima. Teremos, então, os resultados estimados para o modelo VEC Bayesiano, a partir do método de estimação MCMC por Gibbs Sampling, utilizando a informação *a priori* espacial, descrita no Capítulo 4, Seção 4.2 e a informação *a priori* de Minnesota, descrita no Capítulo 4, Seção 4.2, na sua forma modificada.

Diferente dos métodos de estimação convencionais, como o método clássico, o método Bayesiano não fornece seus resultados a partir de um teste de hipóteses, como por exemplo, um teste-t, em que é apresentado um valor estimado para os parâmetros do modelo e, de acordo com uma hipótese inicial, se aceita ou se rejeita aquele valor estimado. O método Bayesiano apresenta distribuições de probabilidade *a posteriori* para os coeficientes do modelo. Portanto, para cada coeficiente do nosso modelo, teremos uma distribuição de probabilidade associada - como resultado do processo de estimação.

Como uma forma de apresentar, interpretar e comparar os resultados empíricos para o modelo VEC Bayesiano utilizando as duas definições de informação *a priori*, tomar-se-á como exemplo o estado de Minas Gerais, mas pode-se estender a forma de interpretação e apresentação dos resultados para os outros estados brasileiros considerados aqui. Todos os resultados apresentados aqui para o estado de Minas Gerais são apresentados no Apêndice B e Apêndice C para os outros estados.

Nas subseções seguintes serão apresentados os resultados empíricos encontrados estimando o modelo VEC Bayesiano utilizando as duas formas distintas de considerar as relações de vizinhança com o objetivo de comparar os resultados encontrados e tentar explicar qual configuração do modelo se adequou melhor ao conjunto de dados aqui utilizados.

5.3.1 Resultados empíricos para o modelo VEC Bayesiano utilizando informação *a priori* espacial

A tabela a seguir apresenta os valores encontrados para as modas dos coeficientes de uma das equações do modelo VEC Bayesiano com informação *a priori* espacial, considerada neste capítulo. Esta é a equação que possui a variação da produção industrial de Minas Gerais como variável dependente.

Dependent Variable mg			
R-squared	0.6551		
Rbar-squared	-0.3264		
sige	3.0632		
Nobs, Nvars	63, 8		
ndraws,nomit	5000, 100		
r-value	4		

Variável	Moda do Coef.	Variável	Moda do Coef.
ba lag1	0.084338	sp lag4	0.033797
ba lag2	-0.138878	pr lag1	-0.019667
ba lag3	0.134565	pr lag2	-0.034014
ba lag4	0.087908	pr lag3	0.001047
mg lag1	0.11336	pr lag4	0.003097
mg lag2	0.027469	sc lag1	0.138452
mg lag3	0.012167	sc lag2	0.012058
mg lag4	-0.000074	sc lag3	0.021921
es lag1	0.067015	sc lag4	0.005999
es lag2	0.208154	rs lag1	-0.006597
es lag3	-0.193758	rs lag2	-0.014682
es lag4	-0.035671	rs lag3	-0.012682
rj lag1	0.235933	rs lag4	-0.01193
rj lag2	-0.057234	ec term ba	-0.375793
rj lag3	-0.207611	ec term mg	-0.085017
rj lag4	-0.05355	ec term es	-0.414301
sp lag1	0.241867	ec term rj	0.080884
sp lag2	-0.213851	ec term sp	0.429261
sp lag3	0.090588	constant	12.568592

Tabela 5.1 – Moda dos coeficientes estimada a partir do método Bayesiano MCMC por Gibbs Sampling para o modelo VEC Bayesiano com informação *a priori* espacial, tendo como variável dependente a variação da produção industrial para o estado de Minas Gerais.

Para cada um dos estados considerados aqui, teremos uma tabela como esta, contendo a moda dos coeficientes estimados. Cada um destes coeficientes estimados em cada uma destas tabelas terá uma distribuição de probabilidade. Esta distribuição de probabilidade é chamada de distribuição *a posteriori* e é esta que apresenta o maior conjunto de informações referente a cada um dos coeficientes. Antes de olharmos para as distribuições de probabilidade *a posteriori* encontradas para os coeficientes, vamos

olhar para a Tabela 5.1 e tentar entender o que cada moda da distribuição *a posteriori* para os coeficientes estimados representa.

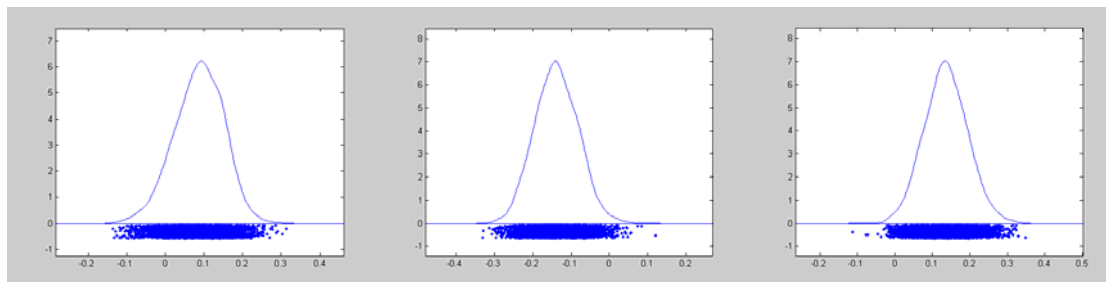
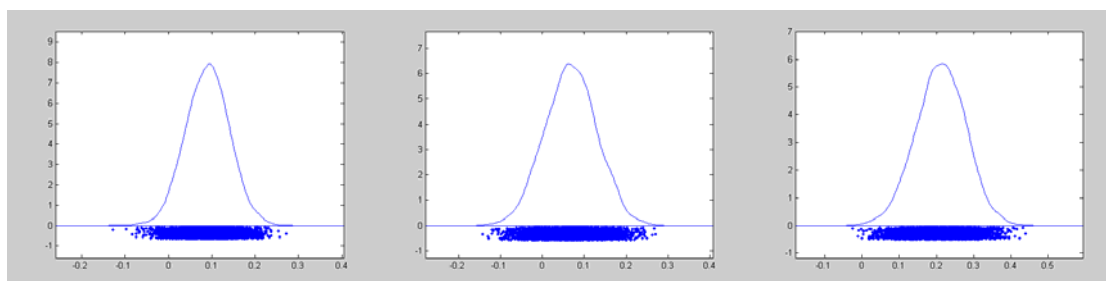
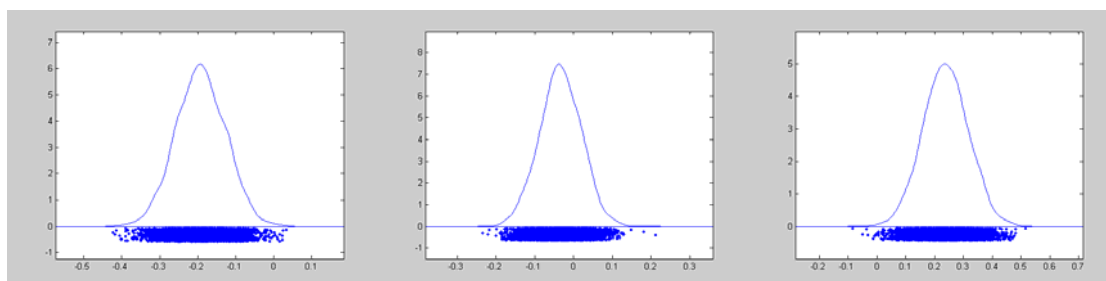
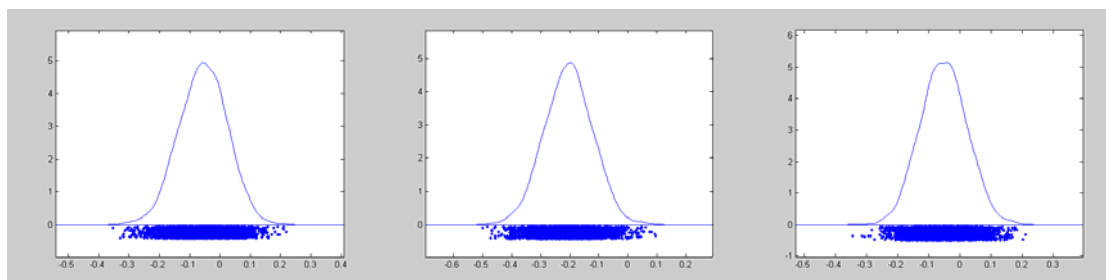
Cada valor assumido para um dos coeficientes que entra na equação de cada uma das equações referente a um estado relaciona a variação da produção industrial no estado com a variação da produção industrial do estado tomado como variável dependente. Por exemplo, vamos considerar o estado de Minas Gerais. O valor assumido pelos coeficientes dos estados brasileiros que entram na equação de Minas Gerais como variáveis explicativas representam quanto da variação da produção industrial em cada um destes estados afeta a variação da produção industrial de Minas Gerais. Isto é, podemos fazer a seguinte pergunta: como a variação da produção industrial de Minas Gerais se comporta, conhecendo a variação da produção industrial dos outros estados, assim como a distribuição espacial dos mesmos? É esta pergunta que os valores dos coeficientes do modelo estruturado desta forma tentam responder.

Devemos tomar muito cuidado quando falamos na moda *a posteriori* do coeficiente estimado. Como já falado, cada coeficiente não terá um valor único como estimativa. Ele terá uma distribuição de probabilidade *a posteriori*. Quando falamos na moda do coeficiente, estamos falando do valor que cada uma destas distribuições *a posteriori* assume como mais provável para cada um dos coeficientes. Isto indica que o valor para esta mediana é o valor em que a densidade *a posteriori* é máxima, mas nada indica que este será o valor assumido realmente por este parâmetro ou mesmo que o valor assumido será próximo à sua moda. Esta informação depende da variância da distribuição *a posteriori*.

Os coeficientes estimados para o modelo VEC Bayesiano, estimados a partir do método de estimação que considera uma informação *a priori* espacial serão, portanto, coeficientes estimados ponderados pela localização. No caso de Minas Gerais, então, as séries referentes aos estados vizinhos – Bahia, Espírito Santo, Rio de Janeiro e São Paulo – são consideradas variáveis mais importantes para estimar a variação da produção industrial de Minas Gerais.

A seguir serão apresentados os gráficos das distribuições de probabilidade *a posteriori* para os coeficientes da equação do modelo VEC Bayesiano com informação *a priori* espacial, que tem a variação da produção industrial de Minas Gerais como variável

dependente. Primeiro, vamos olhar as distribuições *a posteriori* dos coeficientes referentes aos estados que mantém uma relação de vizinhança com Minas Gerais.

(a) $\beta_{BA,t-1}$ (b) $\beta_{BA,t-2}$ (c) $\beta_{BA,t-3}$ (d) $\beta_{BA,t-4}$ (e) $\beta_{ES,t-1}$ (f) $\beta_{ES,t-2}$ (g) $\beta_{ES,t-3}$ (h) $\beta_{ES,t-4}$ (i) $\mu_{RJ,t-1}$ (j) $\beta_{RJ,t-2}$ (l) $\beta_{RJ,t-3}$ (m) $\beta_{RJ,t-4}$

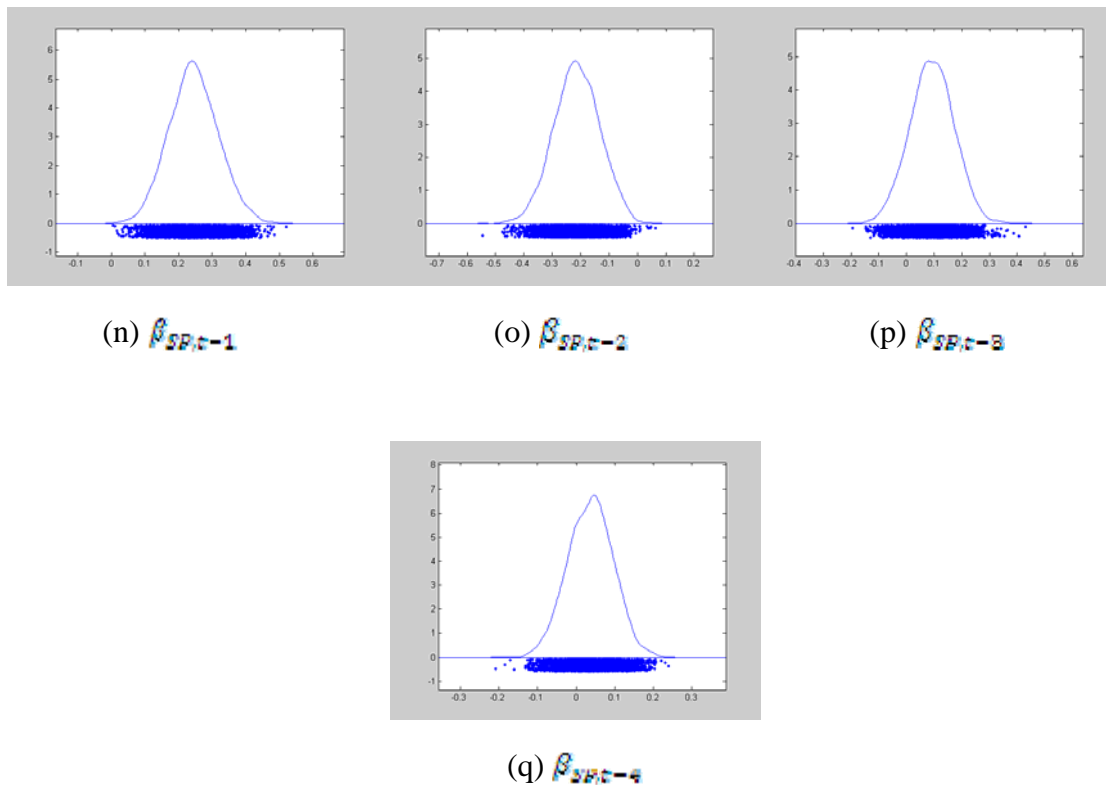


Figura 5.6 – Distribuições de probabilidade *a posteriori* para os coeficientes – aqui chamados de β – para as variáveis explicativas vizinhas ao estado de Minas Gerais, considerando as quatro defasagens temporais incluídas no modelo VEC Bayesiano com informação *a priori* espacial.

As distribuições *a posteriori* apresentadas acima têm moda igual aos valores apresentados na Tabela 5.1 e forma da distribuição mostrada em cada uma das figuras. Estas distribuições foram estimadas por Gibbs Sampling, com número de iterações igual a 5000, onde foram descartadas as primeiras 100 iterações iniciais. Este número é bastante alto de tal forma que seja assegurada a convergência do método.

Além das variáveis relativas aos estados vizinhos a Minas Gerais, as variáveis relativas aos estados não-vizinhos também entram no processo de estimação. No entanto, estas recebem um menor peso, refletindo a crença de que estas variáveis são menos importantes por não manterem uma relação de vizinhança com a série relativa à variável dependente.

Abaixo, são apresentados os gráficos relativos às distribuições *a posteriori* dos coeficientes das variáveis referentes à variação da produção industrial dos estados não-vizinhos a Minas Gerais – Pernambuco, Paraná, Santa Catarina e Rio Grande do Sul.

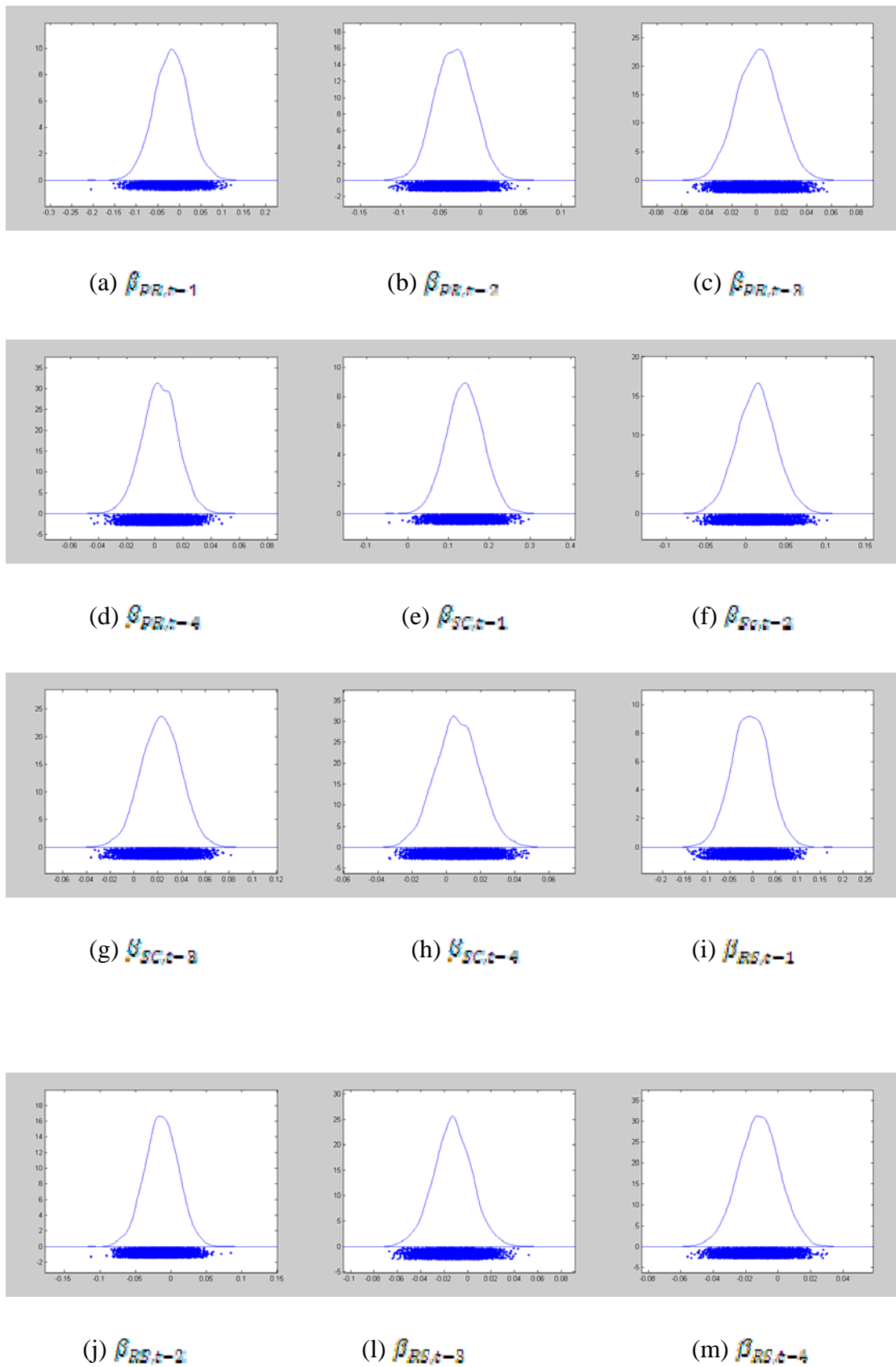
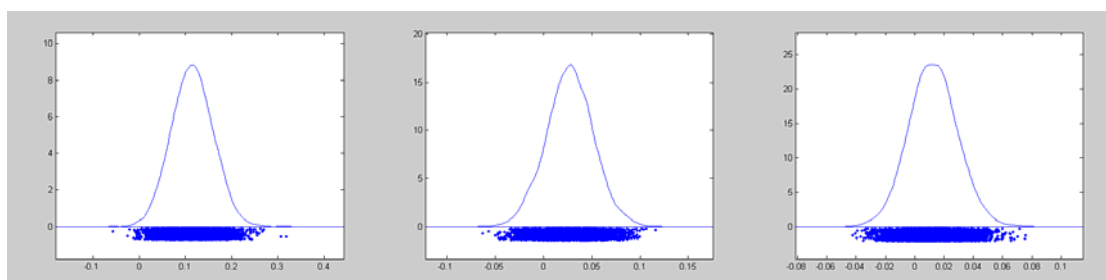


Figura 5.7 – Distribuições de probabilidades *a posteriori* para os coeficientes – aqui chamados de β – para as variáveis explicativas não-vizinhas ao estado de Minas Gerais, considerando as quatro defasagens temporais incluídas no modelo VEC Bayesiano com informação *a priori* espacial.

O ganho do processo de estimação que considera uma informação *a priori* espacial é que, a partir da inserção da matriz contigüidade de primeira ordem padronizada como uma matriz de pesos que entra na especificação da informação *a priori* do método Bayesiano é que temos a partir disto uma informação *a priori* diferente para cada um dos coeficientes. Isto é, será considerada uma informação *a priori* diferente para cada um dos diferentes coeficientes, de acordo com a sua localização. Esta é uma forma de termos um modelo específico e único para cada conjunto de dados regionais. Tratar o mesmo problema tratado aqui com dados referentes a outras localizações usaria o mesmo modelo, mas, no entanto, a informação *a priori* utilizada seria outra, pois a distribuições dos dados seriam diferentes. A matriz contigüidade usada para ponderar a informação *a priori* em outro contexto seria diferente, o que dá flexibilidade ao processo de estimação e cria desta forma, modelos personalizados de acordo com a localização do conjunto de dados considerados.

Outras variáveis que entram no modelo são aquelas referentes aos coeficientes autorregressivos. Isto é, na equação em que a variável explicativa é a variação da produção industrial de Minas Gerais, entram os coeficientes autodefesados relativos à série de Minas Gerais. Além destes, os coeficientes relativos ao vetor de cointegração e à constante. Abaixo, são apresentadas as distribuições *a posteriori* para estes coeficientes.

(a) $\beta_{MG,t-1}$ (b) $\beta_{MG,t-2}$ (c) $\beta_{MG,t-3}$

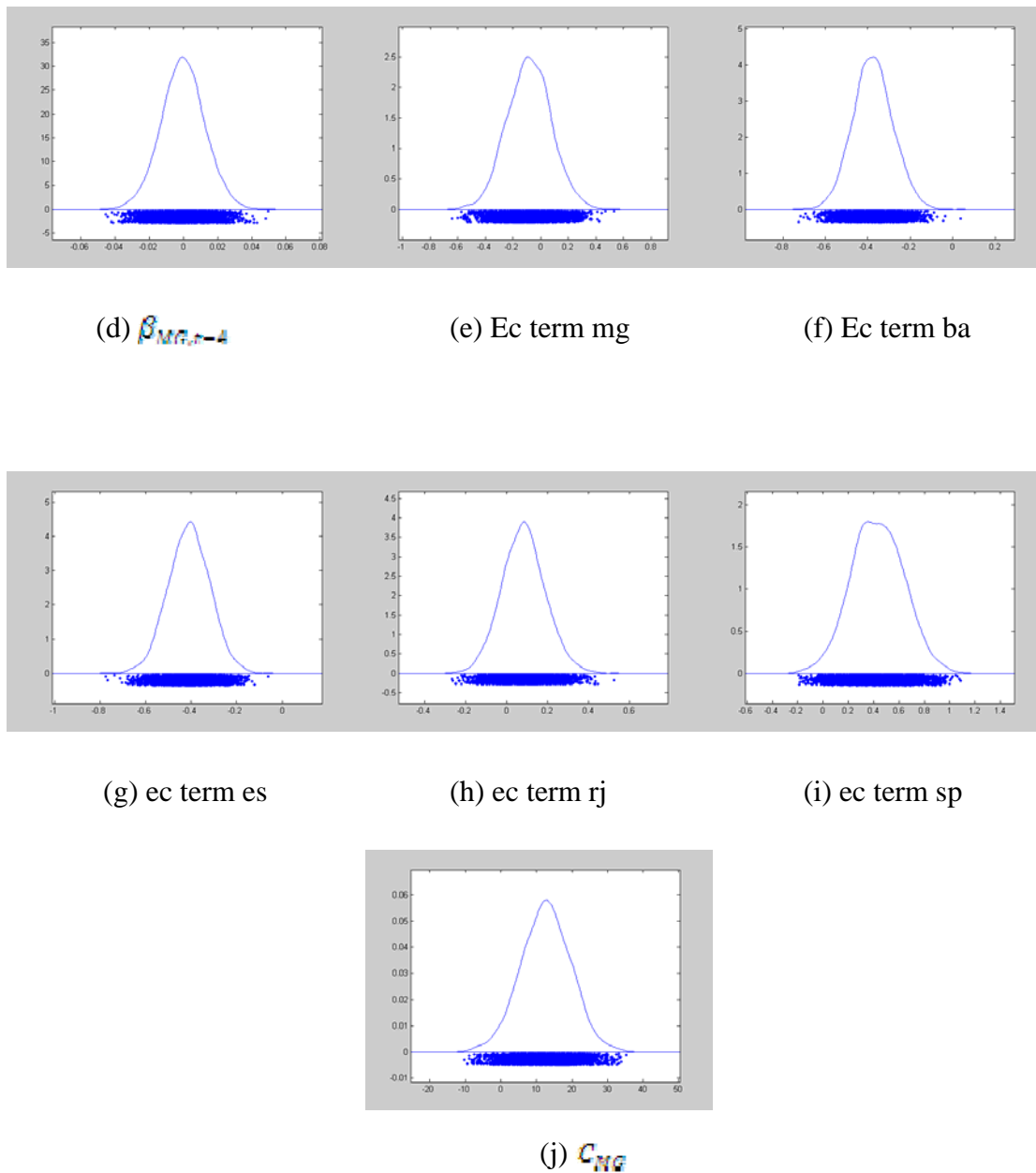


Figura 5.8 – Distribuições de probabilidades *a posteriori* para: os coeficientes das variáveis explicativas autorregressivas relativas ao estado de Minas Gerais, considerando as quatro defasagens temporais incluídas no modelo (a) – (d); para os termos de correções de erros (e) – (i); para a constante incluída nesta equação (j).

Além da distribuição *a posteriori* para os coeficientes, também temos a distribuição *a posteriori* para a variância de cada uma das equações. A figura abaixo mostra o gráfico da distribuição *a posteriori* da variância para a equação do modelo VEC Bayesiano com a variação da produção industrial de Minas Gerais como variável dependente. Cada uma

das oito equações que compõem o modelo terá uma distribuição da variância *a posteriori* que pode ser encontrada no Apêndice B.

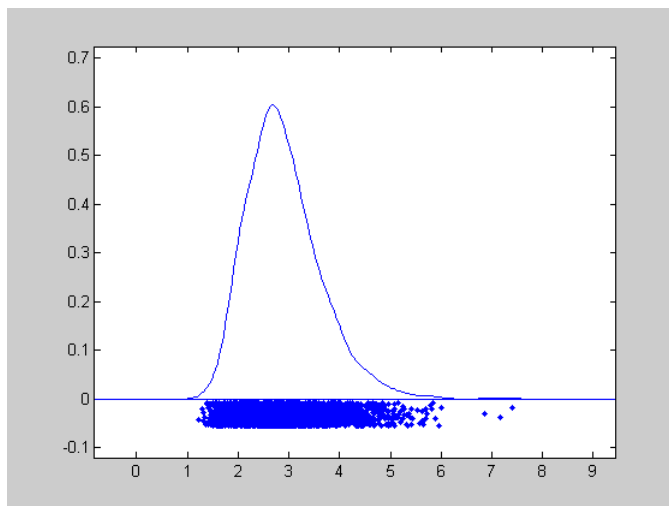


Figura 5.9 – Distribuição de probabilidade *a posteriori* para a variância da equação que tem a variação da produção industrial de Minas Gerais como variável dependente, no modelo estimado usando informação *a priori* espacial.

Então, concluímos aqui a apresentação dos resultados estimados para o modelo VEC Bayesiano com informação *a priori* espacial.

Estimar um modelo VEC Bayesiano considerando as produções industriais para os estados brasileiros pode nos levar a questionar sobre a existência de um padrão de movimento entre as variáveis numa determinada equação. Isto é, as variáveis explicativas que mantêm uma relação de vizinhança com a variável dependente afetarão esta variável dependente sempre da mesma forma, ou de outra forma, entrará na equação da variável explicativa sempre com o mesmo sinal? Observando as modas das distribuições *a posteriori* apresentadas na Tabela 5.1, percebemos que não é possível identificar um padrão de comportamento para as variáveis vizinhas e não-vizinhas para o estado de Minas Gerais. Isto significa que um crescimento na produção industrial de um estado vizinho a Minas Gerais não necessariamente afeta o crescimento da produção industrial de Minas Gerais de forma positiva. Também não é observado nenhum comportamento comum entre as variáveis independentes defasadas, isto é, olhando para

as modas das distribuições *a posteriori* dos coeficientes defasados de mesma ordem temporal, eles também não apresentarão nenhum tipo de comportamento comum, isto é, a produção industrial de alguns estados vizinhos a Minas Gerais afeta de forma positiva o crescimento da produção industrial de Minas Gerais num determinado período do tempo e outros estados vizinhos, de forma negativa, no mesmo período de tempo.

A seguir será apresentado o gráfico da série da variação da produção industrial estimada para o estado de Minas Gerais, assim como a série observada, para podermos comparar nosso modelo com os dados reais. Este gráfico também é estimado para todas as séries de variação da produção industrial para os outros estados brasileiros e também podem ser encontradas no Apêndice B.

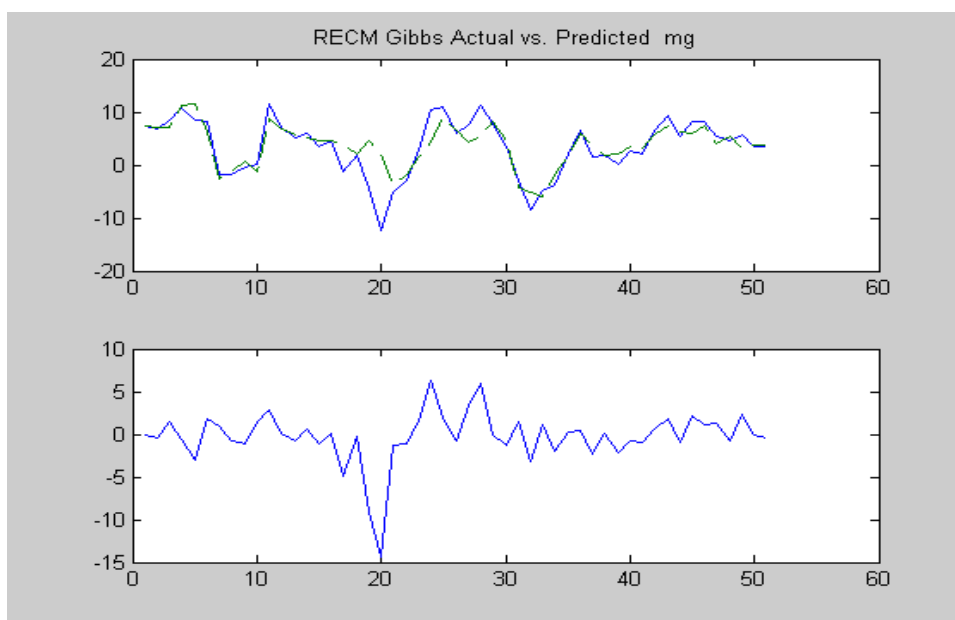


Figura 5.10 – Acima: Série observada para a variação da produção industrial do estado de Minas Gerais *versus* a série estimada – em verde - da variação da produção industrial de Minas Gerais utilizando o modelo VEC Bayesiano com informação *a priori* espacial estimado por Gibbs Sampling. Abaixo: gráfico dos erros estimados.

Pelo gráfico acima, podemos dizer que o modelo estimado se ajusta bem aos dados, mesmo não captando o brusco choque ocorrido em setembro de 1995.

5.3.2 Resultados empíricos para o modelo VEC Bayesiano utilizando informação *a priori* de Minnesota modificada

Da mesma forma que no modelo VEC Bayesiano com informação *a priori* espacial, vamos iniciar a apresentação dos resultados a partir da tabela que mostra os valores das modas dos coeficientes de uma das equações do modelo. A tabela abaixo apresenta estes valores encontrados para a equação que possui a variação da produção industrial de Minas Gerais como variável dependente, aqui com coeficientes estimados utilizando a informação *a priori* de Minnesota modificada.

Dependent Variable mg			
R-squared	0.479		
Rbar-squared	-0.4848		
sige	1.8386		
Nobs, Nvars	63, 8		
ndraws,nomit	5000, 100		
r-value	4		
Variável	Moda do Coef.	Variável	Moda do Coef.
ba lag1	0.051209	sp lag4	0.014819
ba lag2	-0.073361	pr lag1	0.117724
ba lag3	-0.004204	pr lag2	-0.058192
ba lag4	0.033776	pr lag3	0.037861
mg lag1	0.283535	pr lag4	-0.013657
mg lag2	-0.237648	sc lag1	0.100931
mg lag3	0.201865	sc lag2	-0.168671
mg lag4	-0.107405	sc lag3	0.035878
es lag1	-0.024039	sc lag4	-0.081079
es lag2	-0.026129	rs lag1	-0.072849
es lag3	-0.080126	rs lag2	0.055285
es lag4	0.027552	rs lag3	-0.022677
rj lag1	0.356387	rs lag4	-0.02712
rj lag2	-0.067909	ec term ba	0.018815
rj lag3	-0.048303	ec term mg	-0.142289
rj lag4	-0.038802	ec term es	-0.029894
sp lag1	-0.120206	ec term rj	0.006258
sp lag2	-0.025076	ec term sp	0.427333
sp lag3	-0.051765	constant	-6.04196

Tabela 5.2 – Moda dos coeficientes estimada a partir do método Bayesiano MCMC por Gibbs Sampling para o modelo VEC Bayesiano com informação *a priori* de Minnesota modificada, tendo como variável dependente a variação da produção industrial para o estado de Minas Gerais.

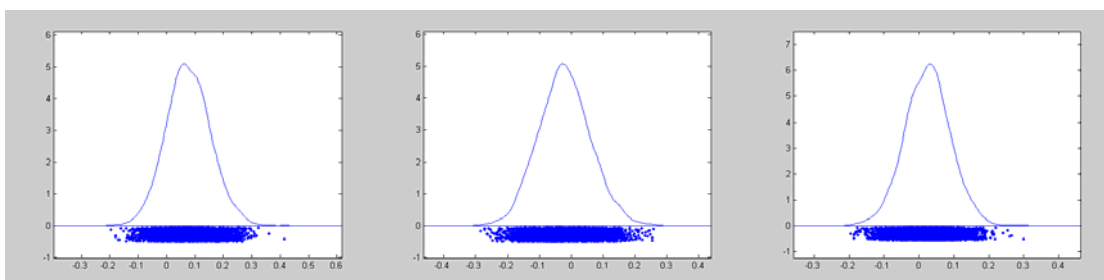
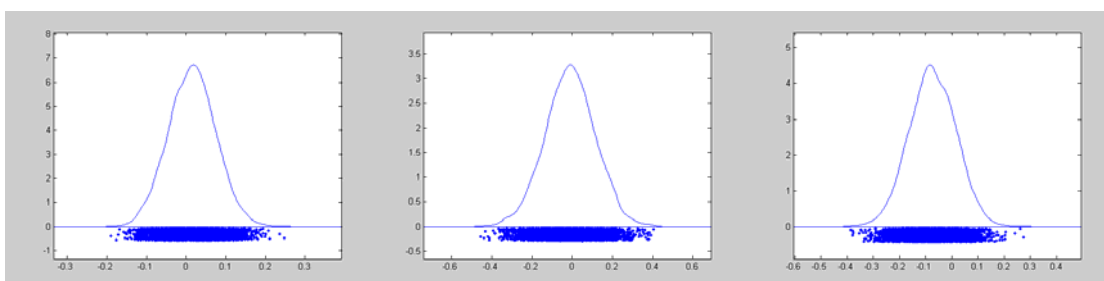
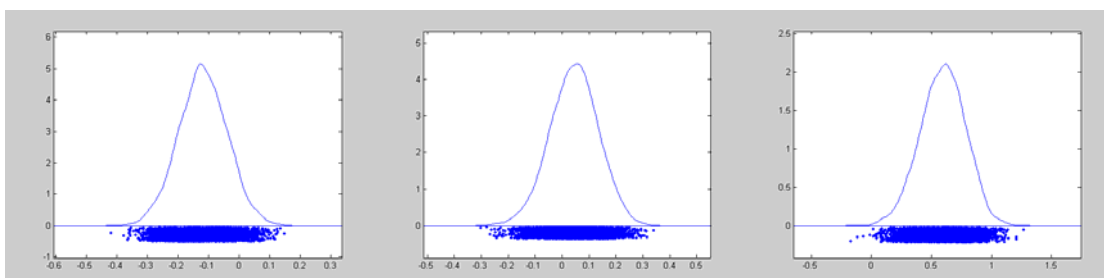
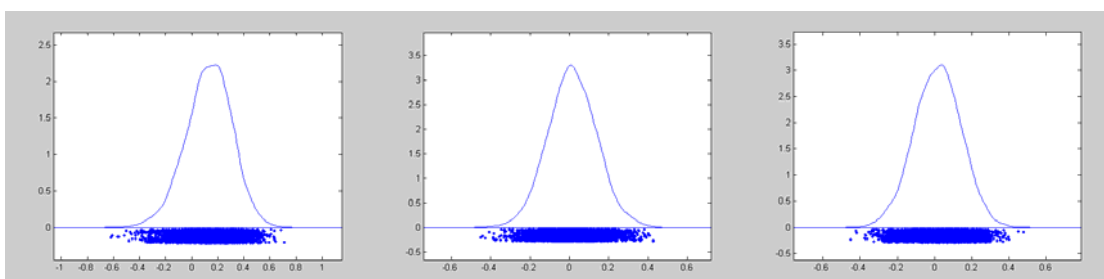
Assim como na Tabela 5.1, a Tabela 5.2 apresenta os valores das modas dos coeficientes estimados. Da mesma forma que no Modelo VEC Bayesiano com informação *a priori* espacial, cada um dos coeficientes estimados para o Modelo VEC Bayesiano com informação *a priori* de Minnesota modificada terá uma distribuição de probabilidade *a posteriori* referente a cada um dos coeficientes.

Aqui, os valores dos coeficientes estimados podem ser interpretados da mesma maneira que no caso do VEC Bayesiano estimado utilizando a informação *a priori* espacial. A diferença está somente nos valores encontrados para as modas das distribuições *a posteriori* e no formato destas distribuições para o Modelo VEC Bayesiano com informação *a priori* espacial e para o Modelo VEC Bayesiano com informação *a priori* de Minnesota modificada.

Os coeficientes estimados a partir do Modelo VEC Bayesiano utilizando o método de estimação que considera informação *a priori* de Minnesota modificada terão estimativas ponderadas, da mesma forma que o método de estimação que considera a informação *a priori* espacial, mas, no entanto, esta forma de ponderação será diferente. A informação *a priori* de Minnesota modificada assume uma média *a priori* igual a 1 para as variáveis que considera mais importantes no modelo, neste caso, as variáveis autodefazadas, e média zero para as outras variáveis, como descrito na seção 4.1. A matriz de pesos que entrará na especificação da informação *a priori* de Minnesota modificada, descrita na Figura 5.5, modifica a crença de que as variáveis autodefazadas são as mais importantes no modelo e que todas as outras variáveis recebem o mesmo peso no processo de estimação. A matriz de pesos considerada aqui modifica esta crença de tal forma que, além de considerar as variáveis autodefazadas de cada equação como as mais importantes, as variáveis relativas aos estados vizinhos também receberão o mesmo peso. Então, serão consideradas com mesma importância as relações autorregressivas do modelo e as relações espaciais do modelo, pois a ambas será associado um peso igual a 1. Às variáveis não-vizinhas será associado também um peso positivo, mas de menor magnitude, e estas outras variáveis serão consideradas igualmente em todas as equações. A elas será atribuído peso igual a 0.1.

A seguir serão apresentados os gráficos das distribuições de probabilidade *a posteriori* para os coeficientes da equação do Modelo VEC Bayesiano com informação *a priori* de Minnesota modificada, que tem a variação da produção industrial de Minas Gerais

como variável dependente. Abaixo, são apresentadas primeiramente as distribuições a *posteriori* dos coeficientes referentes aos estados que mantém uma relação de vizinhança com Minas Gerais.

(a) $\beta_{BA,t-1}$ (b) $\beta_{BA,t-2}$ (c) $\beta_{BA,t-3}$ (d) $\beta_{BA,t-4}$ (e) $\beta_{ES,t-1}$ (f) $\beta_{ES,t-2}$ (g) $\beta_{ES,t-3}$ (h) $\beta_{ES,t-4}$ (i) $\beta_{RJ,t-1}$ (j) $\beta_{RJ,t-2}$ (l) $\beta_{RJ,t-3}$ (m) $\beta_{RJ,t-4}$

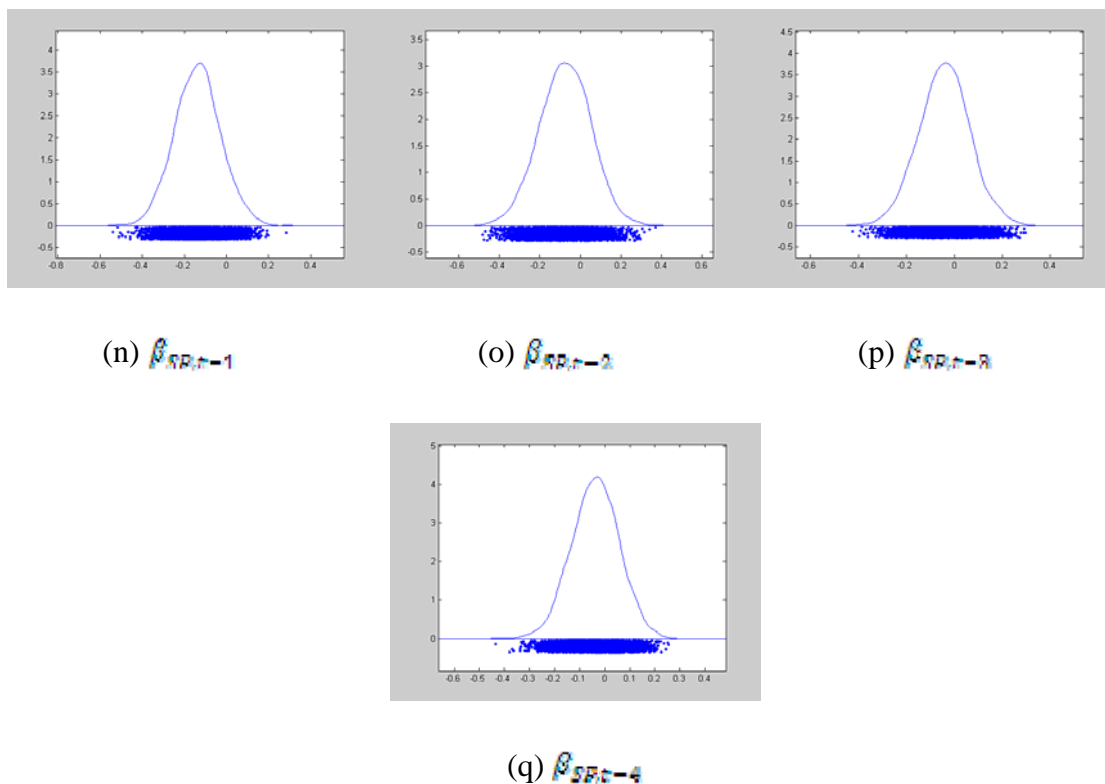


Figura 5.11 – Distribuições de probabilidade *a posteriori* para os coeficientes – aqui chamados de β - para as variáveis explicativas vizinhas ao estado de Minas Gerais, considerando as quatro defasagens temporais incluídas no modelo VEC Bayesiano com informação *a priori* de Minnesota.

As distribuições *a posteriori* apresentadas acima têm moda igual aos valores apresentados na Tabela 5.2 e forma da distribuição mostrada em cada uma das figuras. Estas distribuições também foram estimadas por Gibbs Sampling, com número de iterações igual a 5000, onde foram descartadas as primeiras 100 iterações iniciais.

Além das variáveis relativas aos estados vizinhos a Minas Gerais, as variáveis relativas aos estados não-vizinhos também entram no processo de estimação, no entanto, com um peso menor, como já explicado.

Abaixo, são apresentados os gráficos relativos às distribuições *a posteriori* dos coeficientes das variáveis referentes à variação da produção industrial dos estados não-vizinhos a Minas Gerais – Pernambuco, Paraná, Santa Catarina e Rio Grande do Sul.

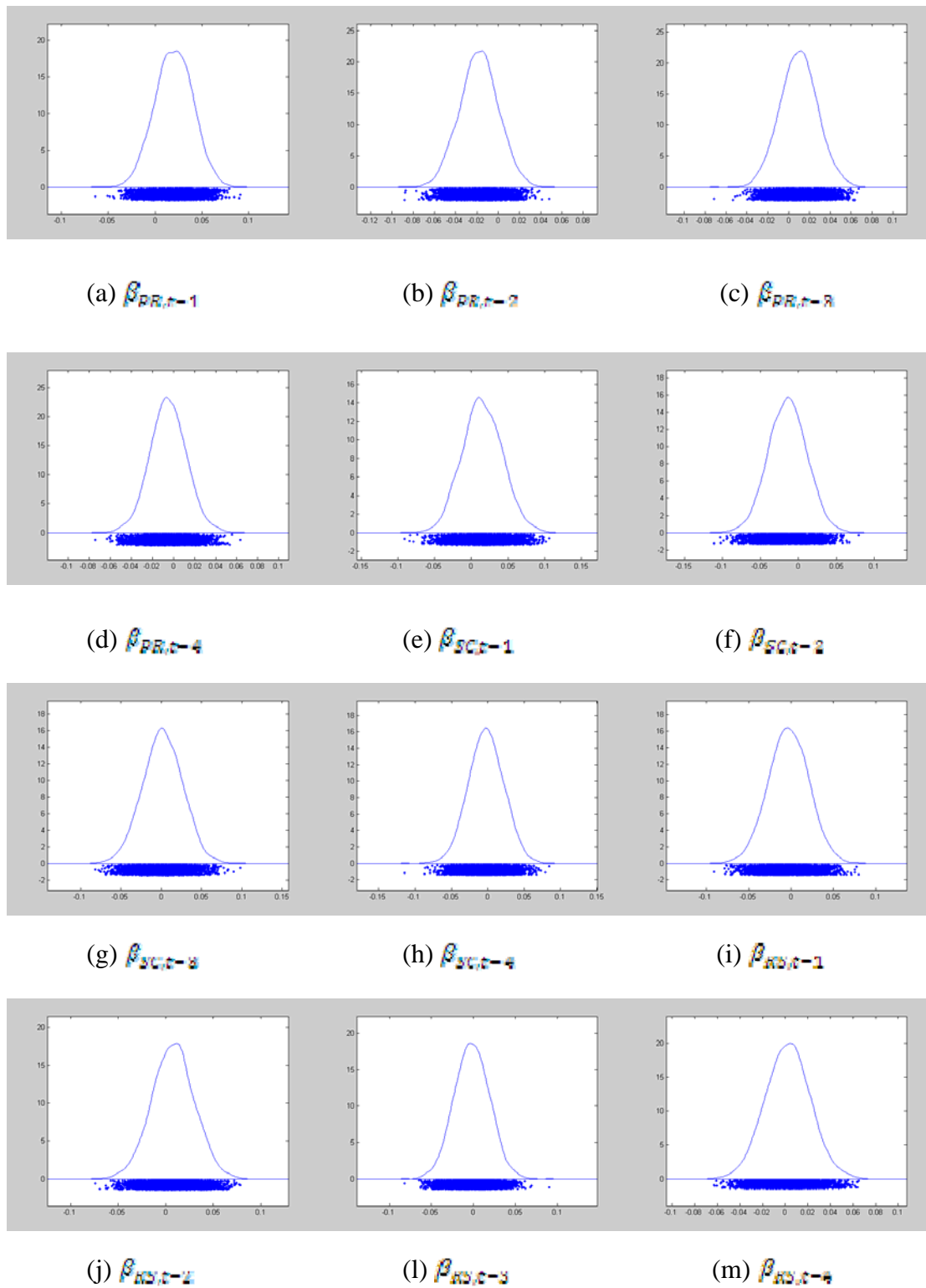


Figura 5.12 – Distribuições de probabilidades *a posteriori* para os coeficientes – aqui chamados de β - para as variáveis explicativas não-vizinhas ao estado de Minas Gerais, considerando as quatro defasagens temporais incluídas no modelo VEC Bayesiano com informação *a priori* de Minnesota.

No caso do processo de estimação que considera a informação *a priori* de Minnesota modificada também teremos a partir disto uma informação *a priori* diferente para cada um dos coeficientes, de acordo com a sua localização, assim como no processo de estimação que considera a informação *a priori* espacial. No entanto, as crenças que entram como informações no processo de estimação são diferentes. No caso da informação *a priori* espacial acreditamos que as relações de contigüidade exercem uma influência maior no processo de estimação e as relações de contigüidade são padronizadas. Por exemplo, os estados vizinhos a Minas Gerais terão, no total, uma ponderação igual a 1. Isto significa que um estado vizinho a Minas Gerais, que tem no total quatro relações de contigüidade, entrará com uma ponderação diferente em outra equação se este mesmo estado for vizinho a outro que tem no total 2 relações de contigüidade. Já na informação *a priori* de Minnesota modificada a ponderação entre os estados vizinhos não sofre modificação, isto é, se dois estados mantêm uma relação de vizinhança, este peso será igual a 1, não importando o número de relações de vizinhança de cada um deles.

Portanto, cada um dos tipos de informação *a priori* agrega a informação localizacional dos dados de uma maneira diferente, refletindo crenças diferentes acerca da interação espacial entre os dados considerados aqui.

Outras variáveis que entram no modelo são aquelas referentes aos coeficientes autorregressivos, isto é, na equação que tem a variável explicativa como a variação da produção industrial de Minas Gerais, entram no modelo os coeficientes autodefesados relativos à série de Minas Gerais. Estes coeficientes são também considerados importantes no modelo e a eles são atribuídos peso igual a 1 na matriz de pesos da informação *a priori* de Minnesota. Além destes, os coeficientes relativos ao vetor de cointegração e à constante. Abaixo, são apresentadas as distribuições *a posteriori* para estes coeficientes.

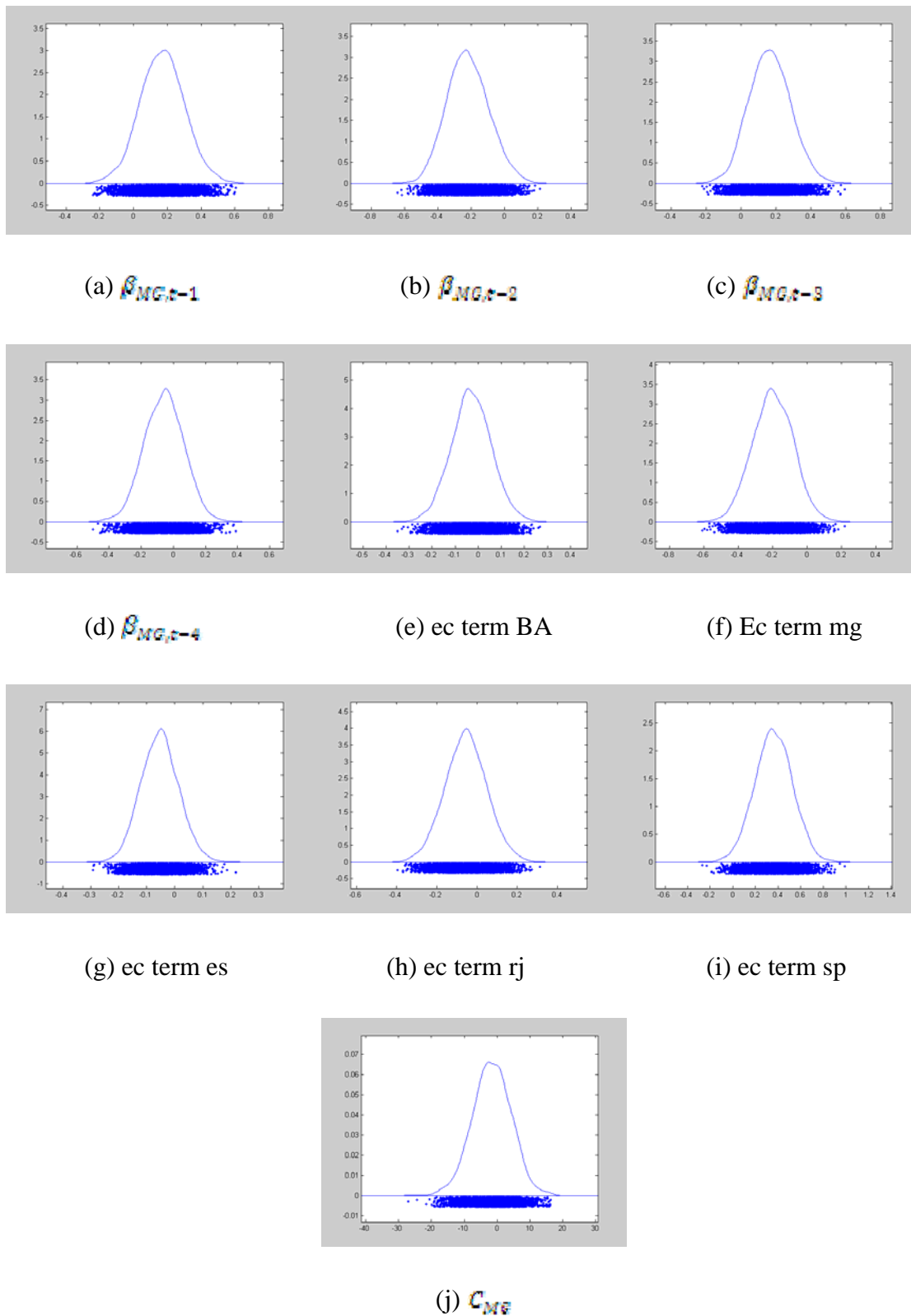


Figura 5.13 – Distribuições de probabilidades *a posteriori* para: os coeficientes das variáveis explicativas autorregressivas relativas ao estado de Minas Gerais, considerando as quatro defasagens temporais incluídas no modelo (a) – (d); para os termos de correções de erros (e) – (i); para a constante incluída nesta equação (j).

Além dos gráficos das distribuições *a posteriori* para os coeficientes, temos também o gráfico da distribuição *a posteriori* da variância para a equação do modelo VEC Bayesiano com a variação da produção industrial de Minas Gerais como variável dependente. Cada uma das oito equações que compõem o modelo terá uma distribuição da variância *a posteriori* que pode ser encontrada no Apêndice C.

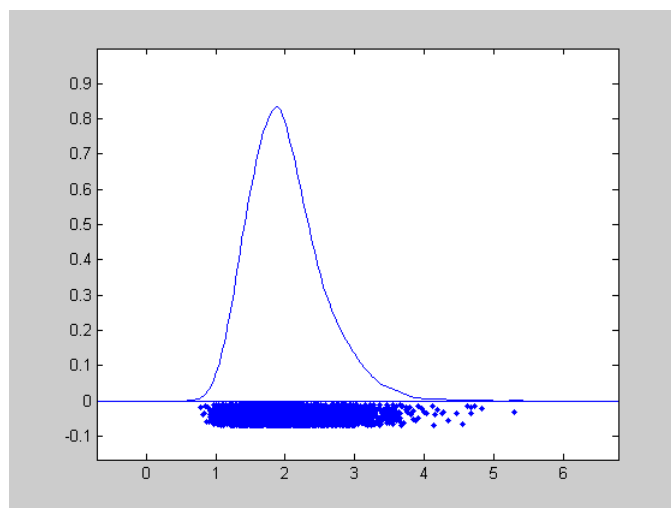


Figura 5.14 – Distribuição de probabilidade *a posteriori* para a variância da equação que tem a variação da produção industrial de Minas Gerais como variável dependente, no modelo estimado usando informação *a priori* espacial.

A seguir será apresentado o gráfico da série da variação da produção industrial estimada para o estado de Minas Gerais, assim como a série observada, para podermos comparar nosso modelo com os dados reais. Este gráfico também é estimado para todas as séries de variação da produção industrial para os outros estados brasileiros e também podem ser encontradas no Apêndice C.

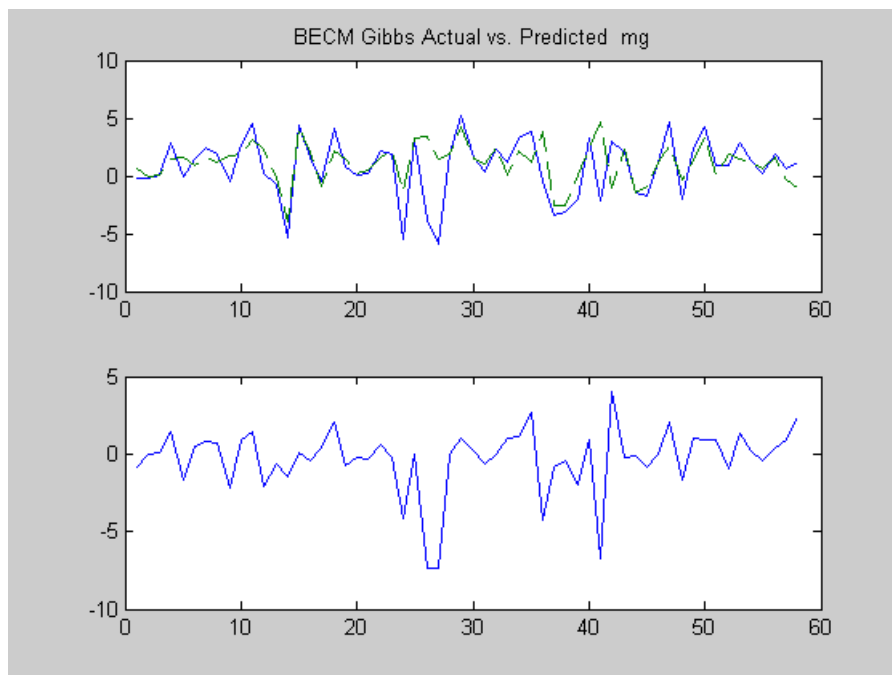


Figura 5.15 – Acima: Série observada para a variação da produção industrial do estado de Minas Gerais *versus* a série estimada – em verde - da variação da produção industrial de Minas Gerais utilizando o modelo VEC Bayesiano com informação *priori* de Minnesota estimado por Gibbs Sampling. Abaixo: gráfico dos erros estimados.

5.3.3 Comparando o Modelo VEC Bayesiano com informação *a priori* espacial e o Modelo VEC Bayesiano com informação *a priori* de Minnesota modificada

O objetivo desta subseção é comparar os dois modelos, o Modelo VEC Bayesiano com informação *a priori* espacial e o Modelo VEC Bayesiano com informação *a priori* de Minnesota modificada, a partir dos resultados apresentados nas subseções anteriores. Para tanto, foi feito um exercício de previsão dentro da amostra para cada um destes modelos e calculados os erros percentuais de previsão, de tal forma que os modelos pudessem ser comparados em relação à precisão da previsão obtida. Os gráficos com as séries originais e as séries estimadas a partir de cada equação do modelo VEC Bayesiano para a especificação de cada tipo de informação *a priori* podem ser vistos na Figura 5.10 para a equação que tem a variação da produção industrial para o estado de Minas Gerais como variável dependente, estimada utilizando *a priori* espacial e na Figura 15.15 para a equação que tem a variação da produção industrial para o estado de

Minas Gerais como variável dependente, estimada utilizando *a priori* de Minnesota modificada. Os gráficos para as séries estimadas, que têm a variação da produção industrial para todos os outros estados que entram nas outras sete equações estimadas no modelo VEC são apresentados no Apêndice B, para o Modelo VEC Bayesiano com informação *a priori* espacial e no Apêndice C, para o modelo VEC Bayesiano com informação *a priori* de Minnesota modificada. Estes gráficos apresentam as séries originais e as séries estimadas a partir de cada um dos modelos.

Com o propósito de comparar os modelos, foi calculado o erro médio percentual para um período da amostra para que fosse possível comparar estas duas formas distintas de considerar informações *a priori* que levam em consideração a localização dos dados. O período da amostra compreende um total de 12 períodos, que se inicia no quarto trimestre de 2003 e termina no terceiro trimestre de 2006. Estes valores foram calculados para cada uma das séries, pois, no Modelo VEC Bayesiano apresentado neste trabalho, cada série da variação da produção industrial neste modelo entra como uma variável dependente no vetor estimado.

O erro percentual médio é calculado pela diferença entre o valor original da série e seu valor previsto, dividido pelo seu valor original e, então, multiplicado por 100. Esta medida nos permite comparar dois modelos distintos uma vez que considera a variação entre os dados observados e previstos e não somente os valores diretos fornecidos pelo modelo. Esta é uma medida que capta a qualidade do ajuste do modelo aos dados e é baseada nela que serão comparados os dois modelos apresentados aqui.

As Tabelas 5.3 e 5.4 apresentam os erros médios percentuais para o Modelo VEC Bayesiano com informação *a priori* espacial e para o Modelo VEC Bayesiano com informação *a priori* de Minnesota modificada. Cada uma das colunas destas tabelas apresenta o erro médio percentual calculado para o respectivo estado, no período que compreende o quarto trimestre de 2003 e o terceiro trimestre de 2006.

As tabelas contendo os valores originais e estimados que são base para o cálculo destas medidas de erro médio percentual são apresentadas no Apêndice D, tanto para os resultados estimados para o Modelo VEC Bayesiano com *a priori* espacial, como para os resultados estimados para o Modelo VEC Bayesiano com *a priori* de Minnesota modificada.

Tabela 5.3 - Medida do erro médio percentual de previsão para o modelo VEC Bayesiano utilizando informação *a priori* espacial

Dates	ba	mg	es	rj	sp	pr	sc	rs
Q4-03	102.383	180.337	123.42	81.143	113.39	181.91	138.791	134.609
Q1-04	82.657	209.357	152.601	112.422	138.767	146.766	124.615	174.989
Q2-04	72.929	180.599	132.887	128.365	112.284	121.758	114.235	152.031
Q3-04	90.573	192.332	139.129	99.923	128.394	127.26	119.889	147.877
Q4-04	82.615	315.475	152.353	101.26	124.658	214.477	167.82	193.34
Q1-05	47.705	473.389	271.185	177.997	248.196	31.005	140.731	359.484
Q2-05	-2.339	117.642	118.082	351.284	37.984	-124.272	86.136	141.213
Q3-05	136.545	306.532	117.087	-43.024	201.144	126.256	167.472	132.771
Q4-05	29.073	361.809	170.461	228.146	61.346	116.505	146.548	238.102
Q1-06	27.558	1038.347	496.663	243.532	557.297	-28.648	154.244	735.924
Q2-06	3.736	-123.564	50.611	1936.357	-32.437	612.681	2.575	37.927
Q3-06	1210.43	4027.242	573.819	748.805	3080.947	757.294	1046.777	850.369

Tabela 5.4 - Medida do erro percentual de previsão para o modelo VEC Bayesiano utilizando informação *a priori* de Minnesota modificada

Dates	ba	mg	es	rj	sp	pr	sc	rs
Q4-03	-8.182	5.632	4.078	0.328	10.93	-0.566	-2.569	4.535
Q1-04	0.355	4.103	4.273	1.861	11.001	4.977	-6.082	5.14
Q2-04	4.262	5.648	7.652	2.278	14.453	-0.109	0.47	8.045
Q3-04	-1.576	9.91	6.143	2.504	17.954	10.551	4.009	7.666
Q4-04	1.196	10.509	11.422	1.749	19.541	9.675	1.652	5.054
Q1-05	-0.777	12.468	7.856	1.877	18.76	8.668	-1.617	2.811
Q2-05	0.923	12.58	11.739	-0.322	22.535	5.957	3.114	2.369
Q3-05	0.386	18.771	-5.293	9.103	17.388	5.093	-17.292	-1.667
Q4-05	5.174	8.074	32.274	-11.055	30.588	0.014	15.561	4.298
Q1-06	-6.759	34.099	-38.917	35.345	5.905	3.219	-50.146	-15.623
Q2-06	20.003	-19.305	119.489	-70.424	64.223	-6.787	98.224	21.808
Q3-06	-42.386	101.031	-220.223	162.651	-60.863	19.103	-237.929	-62.662

Comparando as duas tabelas acima, observamos que todos os valores de medidas para os erros percentuais de previsão para o Modelo VEC Bayesiano utilizando a informação *a priori* de Minnesota modificada apresentam valores bem menores do que os valores de medidas para os erros percentuais de previsão para o Modelo VEC Bayesiano. Portanto, o Modelo VEC Bayesiano utilizando informação *a priori* de Minnesota é o que melhor se ajusta aos dados, fornecendo, portanto, o menor erro de previsão médio

percentual para o período em questão, o que significa que este é o modelo que mais se ajusta aos dados porque fornece o menor erro de previsão.

Então, temos que o modelo VEC Bayesiano utilizando informação *a priori* de Minnesota modificada apresenta os melhores resultados de estimação, o que significa que o modelo que considera os coeficientes com os pesos de acordo com a matriz de ponderação da forma apresentada na Figura 5.5 é o melhor modelo entre os dois apresentados neste estudo.

5.4 Conclusão

O modelo VEC Bayesiano apresentado aqui teve como objetivo estudar a adequabilidade de uma modelagem espacial para estimar as séries para a variação da produção industrial dos diferentes estados brasileiros. Por detectarmos a existência de relações de cointegração, foi escolhido um modelo VEC ao invés de um modelo VAR, de tal forma que fossem levadas em consideração estas relações. Por este conjunto de dados apresentar um componente espacial, isto é, serem dados para diferentes estados brasileiros, foi agregado este componente à estimação através de dois tipos de informação *a priori*. Esta informação introduz no modelo, de forma indireta, a localização dos dados considerados. A forma que estas informações *a priori* são consideradas aqui faz com que tenhamos dois tipos de modelos que fornecem parâmetros estimados de forma ponderada. Esta ponderação se faz pela inserção da matriz contigüidade espacial ou da matriz de pesos da Minnesota modificada à informação *a priori*. Este fato nos leva a ter uma informação *a priori* específica para cada um dos coeficientes, transformando um modelo geral em um modelo que se adéqua aos dados de acordo com a localização dos mesmos.

A diferença entre os dois modelos considerados aqui é que cada um deles assume um tipo de informação *a priori* diferente para o comportamento dos dados. O modelo VEC Bayesiano que utiliza a informação *a priori* espacial atribui pesos diferentes de zero aos estados vizinhos àquele como variável dependente. A informação *a priori* inserida ao modelo desta forma reflete a crença que os estados que mantêm uma relação de vizinhança exercem uma maior influência no processo de estimação. Já o modelo VEC

Bayesiano que utiliza a informação *a priori* de Minnesota, além de atribuir pesos diferentes de zero aos estados vizinhos, atribui pesos também diferentes de zero às variáveis autodefasadas. A *priori* de Minnesota, ao contrário da *priori* espacial, considera as variáveis vizinhas da mesma forma que as variáveis autodefasadas no modelo, atribuindo um peso maior a estas variáveis em relação às variáveis não vizinhas. Na *priori* de Minnesota, tanto as variáveis autodefasadas quanto as variáveis vizinhas entram na mesma especificação da informação *a priori*. A *priori* espacial considera as variáveis autodefasadas de uma forma diferente, isto é, assume uma informação *a priori* para as variáveis autodefasadas diferente da *priori* espacial que pondera as estimativas dos coeficientes vizinhos.

Para a estimação de ambos os modelos, visto conhecermos as distribuições condicionais completas, foi usado um processo de estimação iterativo - o MCMC por Gibbs Sampling.

Com o propósito de comparar os modelos, foi feito um exercício de previsão dentro da amostra, calculando então uma medida de erro de previsão médio percentual para cada um dos modelos. Os resultados apresentados mostram que o modelo VEC Bayesiano utilizando *a priori* de Minnesota é o modelo que fornece os menores erros de previsão, sendo, portanto, o modelo que melhor se ajusta aos dados. Desta forma, podemos concluir que tanto as relações de dependência temporal quanto as relações de dependência espacial podem ser incluídas no modelo através de uma informação *a priori* com a mesma estrutura. Neste caso, uma estrutura de ponderação, que atribui importância maior as variáveis mais próximas no espaço e no tempo. A *priori* de Minnesota usada aqui considera a matriz de pesos construída com base nas relações de contigüidade de primeira ordem tendo maior importância no processo de estimação, assim como as relações autorregressivas. Então, podemos concluir que variáveis mais próximas no tempo e no espaço exercem uma maior influência no processo de estimação, e estas relações não necessariamente precisam ser incluídas no modelo com uma estrutura diferente, como se faz através do modelo VEC Bayesiano utilizando *a priori* espacial. Um modelo mais simples, como o modelo VEC Bayesiano utilizando *a priori* de Minnesota forneceu melhores resultados neste estudo utilizando dados para produção industrial entre os estados brasileiros.

6 Considerações Finais

Com o intuito de apresentar um trabalho bastante completo, antes da apresentação dos modelos VAR e VEC Espaciais seguindo uma abordagem Bayesiana foi apresentada a fundamentação teórica tanto sobre a econometria espacial quanto para os métodos Bayesianos e seus conceitos básicos. O Capítulo 2 foi importante no sentido de localizar o modelo focado dentro da literatura relativamente recente que é a econometria espacial, assim como conhecer outros modelos capazes de tratar dados com características regionais de formas distintas. O Capítulo 3 apresentou os conceitos básicos relativos à econometria Bayesiana e também apresentou idéias intuitivas para entender os métodos Bayesianos.

O Capítulo 4 apresentou os Modelos VAR e VEC Bayesianos, apresentando dois tipos de informação *a priori*: a *priori* de Minnesota e a *priori* espacial. A *priori* de Minnesota associa média *a priori* igual à zero para os coeficientes em todas as variáveis independentes que não sejam auto defasadas, de forma que exista uma influência decrescente destas variáveis no modelo. Isto é, a informação *a priori* de Minnesota, construída desta forma, agrega a suposição de que as variáveis independentes que não são auto defasadas exercem uma menor influência que as variáveis dependentes auto defasadas em cada equação. Já a *priori* espacial é incorporada no modelo de forma que ela restringe os parâmetros que correspondem às defasagens temporais de variáveis que não apresentam relação de vizinhança impondo uma média *a priori* igual à zero para estas variáveis, isto é, para as variáveis que não são vizinhas à região da variável dependente em uma determinada equação. Então, a *priori* espacial envolve médias e variâncias baseadas em relações de contigüidade.

A informação *a priori* de Minnesota foi modificada de tal forma que as variáveis independentes que mantinham uma relação de vizinhança com a variável dependente exercessem uma influência de mesma magnitude que a variável dependente autodefásada. Com o objetivo de observar os resultados dos modelos aplicados a dados reais, foram estimados e analisados os modelos VEC Bayesianos utilizando os dois tipos apresentados de informação *a priori* - a *priori* de Minnesota modificada e a *priori* espacial – utilizando dados da produção brasileira para oito estados.

Estes dois modelos também foram comparados através de um exercício de previsão dentro da amostra e cálculo de uma medida de erro percentual médio e concluímos que o modelo VEC Bayesiano com a informação *a priori* de Minnesota modificada é o modelo que melhor estima os dados. Isto significa dizer que um modelo que atribui um peso maior aos dados de acordo com as regiões, considerando variáveis independentes vizinhas e variáveis dependentes autodefasadas no processo de estimação, é o melhor modelo entre os dois propostos aqui.

Como sugestões para trabalhos futuros poderiam ser estimados os modelos VAR Bayesianos e compará-los aos modelos VEC Bayesianos, utilizando as diferentes configurações de informações *a priori* - que não se limitam às duas apresentadas aqui – e observar qual dos modelos melhor se aplica aos dados além de ser possível analisar e comparar modelos que incluem ou não as relações de longo prazo consideradas pelo VEC Bayesiano aqui estimados. Outra sugestão seria aplicar este modelo a um diferente conjunto de dados e observar se as conclusões apresentadas aqui se mantêm. Uma abordagem interessante seria trabalhar com dados que se apresentem em regiões mais próximas, isto é, com dados que apresentem um maior número de relações de contigüidade entre eles. Por exemplo, dados municipais ou referentes à determinada região metropolitana ou estadual.

Referências

ANSELIN, L. **Spatial Econometrics**. Dallas: Bruton Center. School of Social Sciences. University of Texas, 1999. Disponível em: http://csiss.ncgia.ucsb.edu/learning_resources/content/papers/baltchap.pdf. Acesso em: 23 de maio de 2007.

ANSELIN, L. **Spatial Econometrics: Method and Models**. Dordrecht: Kluwer, 1988.

ANSELIN, L.; FLORAX, R. J. G. M.; REY, S. J. **Advances in Spatial Econometrics: Methodology, Tolls and Applications**. Springer, Verlag, 2004.

ARELLANO, M.; BOND, S. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. **Review of Economic Studies, London**, v. 58, p. 277 – 297, 1991.

BAILLIE, A. D.; ORD, J. K. Prediction from ARMAX Models. **Journal of Econometrics**, v. 12, p.365 – 374, 1980.

BELL, P. K.; BOCKSTAEL, N. E. Applying the Generalized Moments Estimation Approach to Spatial Problems Involving Microlevel Data. **The Reviews of Economics and Statistics**. v. 82. n. 1, 72 – 82. 2000.

BEENSTOCK, M.; FELSENSTEIN, D. Spatial Vector Autoregression. Cidade: Jerusalém. **Hebrew University of Jerusalem**, 2005.

BLUNDELL, R.; BOND, S. Initial Conditions and Moments Restrictions in Dynamic panel data Models. **Econometrica**. v. 87, 115 – 143. 1998.

BYRON, R. P. Polynomial Approximations in Cross Sectional Models. **Journal of Applied Econometrics**, v. 7, n. 3, p. 309 – 322, 1992.

CLIFF, A.D. et al. **Elements of Spatial Structure**. Cambridge: Cambridge University Press, 1975.

CLIFF, A. D.; ORD, J.K. **Spatial Autocorrelation**, London: Pion, 1973.

CONJUGATE PRIOR. In: WIKIPEDIA: The Free Encyclopedia. St. Petersburg, FL: Wikipedia Foundation, 2007. Disponível em:
http://en.wikipedia.org/wiki/Conjugate_prior. Acesso em: 24 out. 2007.

DRISCOLL, J.C.; KRAAY, A. C. Consistent covariance matrix estimation with spatially dependent panel data. **The Review of Economics and Statistics**, v. 80, p. 549 – 560, 1998.

DOAN, T.; LITTERMAN, R. B.; SIMS, C. A. Forecasting and Conditional Projections using Realistic Prior Distributions. **Econometric Reviews**, v. 3, p. 1 – 100, 1984.

ELHORST, J. P. Dynamic Models in space and time. **Geography Analysis**, Groningen, v. 33, p. 119 – 140, 2001.

ELHORST, J. P. Specification and estimation of spatial panel data models. **International Regional Science Review**. v. 26, p. 244 – 268, 2003.

ENGLE, R.F.; GRANGER, C.W.J. Co-Integration and Error Correction: Representation, Estimation, and Testing. **Econometrica**. v. 55, p. 251 – 276, 1987.

FELIX, R. M.; NUNES, L. C. Um Modelo de Previsão Bayesiano para a área do Euro. **Boletim Econômico: Banco de Portugal**, Lisboa, p. 71 – 88, mar. 2002.

Disponível em:

http://www.bportugal.pt/publish/bolecon/docs/2002_1_3_p.pdf.

Acesso em: 1 jun. 2007.

FINGLETON, B. Spurious Spatial Regression: some Monte Carlo results with a spatial unit root and spatial cointegration. **Journal of Regional Science**. v. 39, p. 1 – 19, 1999.

FLORAX, R. J. G. M.; VAN DE LIST, A. J. Spatial Econometric Data Analysis: Moving beyond Traditional Models. **International Regional Science Review**. v. 26, p. 244 – 268, 2003.

GIACOMINI, R.; GRANGER, C. W. J. Aggregation of space-time processes. **Journal of Econometrics**. v. 118, p. 7 – 26, 2004.

HAINING, R. Dynamic Models in space and time. **Geographical Analysis**, v. 33, p. 119 – 140, 1990.

HALL, A. Testing for a unit root in time series with pretest data-based model selection. **Journal of Business and Economic Statistics**. v. 12, p. 461 – 470, 1994.

HENDRY, D. **Dynamic Econometrics**. Oxford: Oxford University Press, 1995.

HSIAO, C. **Analysis of Panel Data**, Cambridge: Cambridge University Press, 1986.

IMPROPER Priors. In: WIKIPEDIA: The Free Encyclopedia. St. Petersburg, FL: Wikipedia Foundation, 2007. Disponível em:

http://en.wikipedia.org/wiki/A_priori_probability#Improper_priors. Acesso em: 24 out. 2007.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, IBGE.

Disponível em: <http://www.ibge.org.br/>

Acesso em: 24 out. 2007.

JAMES, B. **Probabilidade**: um curso em nível intermediário. São Paulo: Instituto de Matemática Pura e Aplicada – CNPq. 1981.

JOHANSEN, S., Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control*, 1988.

KADIYALA, K.R.; KARLSSON, S. Numerical Methods for Estimation and Inference in Bayesian VAR Models. **Journal of Applied Econometrics**, v. 12, n. 2, 99 – 132. 1997.

KAMARIANAKIS, Y. **Spatial-Time Series Modeling: A Review of the proposed methodologies**. Chicago: Regional Economics Applications Laboratory, University of Illinois at Chicago, 2003. (Discussion Papers: Technical Series, 03 – T – 19). Disponível em:

http://www.iacm.forth.gr/regional/papers/spatial_time_series_overview.pdf . Acesso em 23 de maio de 2007.

KAO, C. Spurious regression and residual based tests for cointegration in panel data. **Journal of Econometrics**. v. 90, p. 1 – 44, 1999.

KELEJIAN, H.H.; PRUCHA, I. R. A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. **International Economic Review**. v. 40, p. 509 – 522, 1999.

KELEJIAN, H.H.; PRUCHA, I. R. YUZEFOVICH, Y. **Instrumental Variable Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances: Large and Small Sample Results**. 2003. Disponível em:

http://www.spatialstatistics.info/webposting/kelejian_prucha/KPY_Paper.pdf. Acesso em: 23 maio 2007.

KOOP, G. **Bayesian Econometrics**. London: John Wiley, 2004.

LANCASTER, T. **An Introduction to Modern Bayesian Economics**. Blackwell Publishing, 2004.

LEE, L. F. **Identification and Estimation of Spatial Econometric Models with group interactions, contextual factors and fixed effects**. Cidade: Ohio, Department of Economics – The Ohio State University, 2004. Disponível em:

<http://economics.sbs.ohio-state.edu/lee/wp/group-inter-fix-e-jan04-fin1.pdf>. Acesso em: 23 maio 2007.

LESAGE, J. P.; KRIVELYOVA, A. A Spatial prior for Bayesian vector autoregressive models. **Journal of Regional Science**. v. 39, p. 297 – 317, 1999.

LESAGE, J. P. **Spatial Econometrics**. Department of Econometrics. University of Toledo. 1999. Disponível em:

<http://www.rri.wvu.edu/WebBook/LeSage/spatial/wbook.pdf>. Acesso em 23 de maio de 2007.

LIKELIHOOD Principle. In: WIKIPEDIA: The Free Encyclopedia. St. Petersburg, FL: Wikipedia Foundation, 2007. Disponível em:

http://en.wikipedia.org/wiki/Likelihood_principle. Acesso em: 24 out. 2007.

LUTKEPOHL, H. **Forecasting aggregated vector ARMA process**. Berlim: Springer, Berlim, Heidelberg, 1987.

MORAIS, I.A.; PORTUGAL, M.S. Business Cycle in the Industrial Production of Brazilian States. In: ENCONTRO NACIONAL DE ECONOMIA, 31., 2003, Porto Seguro. Anais do XXXI Encontro Nacional de Economia. Porto Seguro, BA: ANPEC, 2003. CD-ROM.

PAENLINCK, J.; KLASSEN. L. **Spatial Econometrics**. Farnborough: Saxon House, 1979.

POIRIER, D. J.; TOBIAS, J. L. **Bayesian Econometric**: 2004. Disponível em:

http://www.econ.iastate.edu/faculty/tobias/documents/palgrave_pt.pdf. Acesso em: 01 jun. 2007.

POSTERIORI PROBABILITY DISTRIBUTION. In: WIKIPEDIA: The Free Encyclopedia. St. Petersburg, FL: Wikipedia Foundation, 2007. Disponível em:

http://en.wikipedia.org/wiki/Posterior_probability. Acesso em: 24 out. 2007.

PRIORI PROBABILITY DISTRIBUTION. In: WIKIPEDIA: The Free Encyclopedia. St. Petersburg, FL: Wikipedia Foundation, 2007. Disponível em:

http://en.wikipedia.org/wiki/A_priori_probability#Prior_probability_distribution.

Acesso em: 24 out. 2007.

SIMS, C. A. Macroeconomics and Reality. **Econometrica**. v. 48, n. 1, p.1 – 48, 1980.

STOCK, J. H.; WATSON, M.W. Business Cycle Fluctuations in U.S Macroeconomic Time Series. **National Bureau of Economic Research Working Papers**, Cambridge, Mass., n.6528. Abril, 1998. Disponível em:

<http://www.nber.org/papers/w6528>. Acesso em: 1 jun. 2007.

STOCK, J. H.; WATSON, M.W. Vector Autoregressions. **The Journal of Economic Perspective**, v. 15. n. 4, p. 101 – 115, 2001.

STRONG Law of Large Numbers. In: WIKIPEDIA: The Free Encyclopedia. St. Petersburg, FL: Wikipedia Foundation, 2007. Disponível em:

http://en.wikipedia.org/wiki/Strong_law. Acesso em: 24 out. 2007.

WITHERS, S.D. Quantitative methods: Bayesian inference, Bayesian thinking. Progress in **Human Geography**. v. 26. n. 4, p. 553 – 566, 2002.

Apêndice A Apresentação dos resultados dos testes iniciais necessários para escolha do modelo utilizado no exercício proposto no Cap. 5.

Testes de Especificação para o Modelo VAR – Teste de Razão-Verossimilhança

Os resultados apresentados abaixo são adquiridos através do teste Razão-Verossimilhança, em que se compara o modelo irrestrito (modelos com ordem de defasagens mais longas) com o modelo irrestrito (modelos com ordem de defasagens mais curtas). Neste caso, tomamos o modelo com ordem de defasagem máxima igual a 5 (pelo fato de estarmos trabalhando com dados trimestrais) e o modelo com defasagem mínima igual a 1. Para determinar a ordem de defasagem estatisticamente ótima para usarmos no modelo VAR que envolve os nove estados como variáveis explicativas o teste de Razão-Verossimilhança é apresentado abaixo:

nlag = 6 5	LR statistic	163.0356	probability	1.36E-10
nlag = 5 4	LR statistic	76.5025	probability	0.136
nlag = 4 3	LR statistic	123.5519	probability	1.15E-05
nlag = 3 2	LR statistic	64.098	probability	0.473
nlag = 2 1	LR statistic	115.7857	probability	7.96E-05

Considerando um nível de significância de 0.01, concluímos que a ordem de defasagem ótima é 4, pois o teste de Razão-Verossimilhança rejeita a hipótese de que a ordem de defasagem igual a 4 piora a qualidade de estimativa do modelo, quando comparada com a ordem de defasagem igual a 5.

Teste de Cointegração de Johansen

Para a maioria dos modelos de correção de erros, mais que duas variáveis podem manter uma relação de cointegração, então o procedimento de dois passos proposto por Engle e Granger precisa ser generalizado. Johansen fornece esta generalização que toma a forma de um teste de razão-verossimilhança. O procedimento de Johansen fornece um teste estatístico para determinar r , o número de relações de cointegração entre as n variáveis do vetor y_t , assim como um conjunto dos r vetores de cointegração que serão usados para construir as variáveis de correção de erros no modelo vetorial de Correção de Erros.

Dado o valor de r , (o número de relações de cointegração), podemos usar os autovetores fornecidos pela função Johansen junto com os níveis de y_t defasados um período para formar um conjunto de variáveis de correção de erros para o modelo de correção de erros.

NULL:		Trace Statistic	Crit 90%	Crit 95%	Crit 99%
$r \leq 0$	ba	334.143	153.634	159.529	171.09
$r \leq 1$	mg	225.805	120.367	125.619	135.982
$r \leq 2$	es	150.066	91.109	95.754	104.964
$r \leq 3$	rj	85.991	65.82	69.819	77.82
$r \leq 4$	sp	52.386	44.493	47.855	54.682
$r \leq 5$	pr	27.411	27.067	29.796	35.463
$r \leq 6$	sc	8.396	13.429	15.494	19.935
$r \leq 7$	rs	1.434	2.705	3.841	6.635

NULL:		Eigen Statistic	Crit 90%	Crit 95%	Crit 99%
$r \leq 0$	ba	108.339	49.285	52.362	58.663
$r \leq 1$	mg	75.739	43.295	46.23	52.307
$r \leq 2$	es	64.075	37.279	40.076	45.866
$r \leq 3$	rj	33.605	31.238	33.878	39.369
$r \leq 4$	sp	24.975	25.124	27.586	32.717
$r \leq 5$	pr	19.015	18.893	21.131	25.865
$r \leq 6$	sc	6.963	12.297	14.264	18.52
$r \leq 7$	rs	1.434	2.705	3.841	6.635

A tabela acima mostra a estatística traço e a estatística de autovalor máximo, assim como os valores críticos de 90%, 95% e 99%. Para este caso, temos que (usando o nível de significância de 95%), que a estatística traço rejeita $r \leq 0$, pois 334.143 é maior que o valor crítico de 159.529; ela também rejeita $r \leq 1$, $r \leq 2$, $r \leq 3$ e $r \leq 4$ porque as estatísticas-traço para estes valores também excedem os valores críticos associados. Para $r \leq 5$ não podemos rejeitar a Hipótese Nula, então concluímos que $r=5$. Usando a estatística de autovalor máximo, encontramos outro valor para r , temos que $r=3$ utilizando a estatística do autovalor máximo, usando o nível de significância de 95%.

Depois de determinadas o número de relações de cointegração, podemos usar estes resultados, junto com os autovetores formados, para construir um conjunto de variáveis de correção de erros. Estes são construídos usando y_{t-1} , multiplicados pelos r

autovetores associados com as relações de cointegração para formar r variáveis cointegrantes.

Matriz de Contigüidade de Primeira Ordem e Matriz de Contigüidade de Primeira Ordem Padronizada

A matriz abaixo foi construída com base na definição de matriz de contigüidade de primeira ordem, definida no capítulo 1, seção 1.4.1, de acordo com a localização espacial dos oito estados brasileiros utilizados neste estudo. Aos estados que apresentam relação de vizinhança entre si, é dado peso 1 e aos estados que não apresentam relação de vizinhança é dado peso zero. Por exemplo, a Bahia mantém relação de vizinhança com Minas Gerais, então, temos o valor 1 na linha 1, coluna 2. Pelo fato desta ser uma matriz simétrica, temos o valor 1 também na linha 2, coluna 1. As relações de contigüidade completas podem ser vistas logo abaixo.

	BA	MG	ES	RJ	SP	PR	SC	RS
BA	0	1	1	0	0	0	0	0
MG	1	0	1	1	1	0	0	0
ES	1	1	0	1	0	0	0	0
RJ	0	1	1	0	1	0	0	0
SP	0	1	0	1	0	1	0	0
PR	0	0	0	0	1	0	1	0
SC	0	0	0	0	0	1	0	1
RS	0	0	0	0	0	0	1	0

Já a matriz de contigüidade de primeira ordem padronizada segue a mesma relação de contigüidade da matriz apresentada anteriormente. No entanto, ela é padronizada de tal forma que a soma de cada uma das linhas da matriz seja igual a 1. A definição de matriz de contigüidade de primeira ordem padronizada também é apresentada no capítulo 1, seção 1.4.1. Como exemplo, vamos olhar para o estado de Minas Gerais, linha 3 da matriz de contigüidade apresentada acima. Este estado mantém relação de vizinhança com os estados da Bahia, Espírito Santo, Rio de Janeiro e São Paulo. Na matriz acima, as linhas e colunas referentes a estas relações de contigüidade são iguais a 1. Já na coluna abaixo, a soma da linha referente a Minas Gerais será igual a 1, isto é, $\frac{1}{4}$ para cada relação de contigüidade. Abaixo é apresentada a matriz de contigüidade de primeira ordem padronizada.

Apêndice B Apresentação dos resultados empíricos do modelo VEC Bayesiano utilizando informação *a priori* espacial estimado por Gibbs Sampling

PRIOR hyperparameters

sig = 0.10

tau = 6.00

theta = 0.50

Weight matrix =

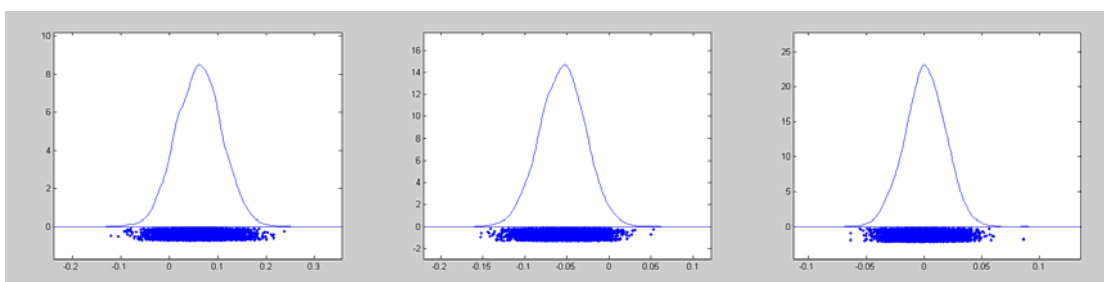
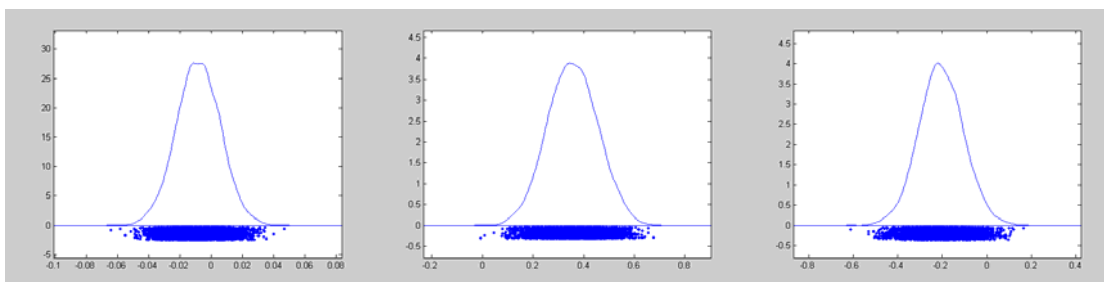
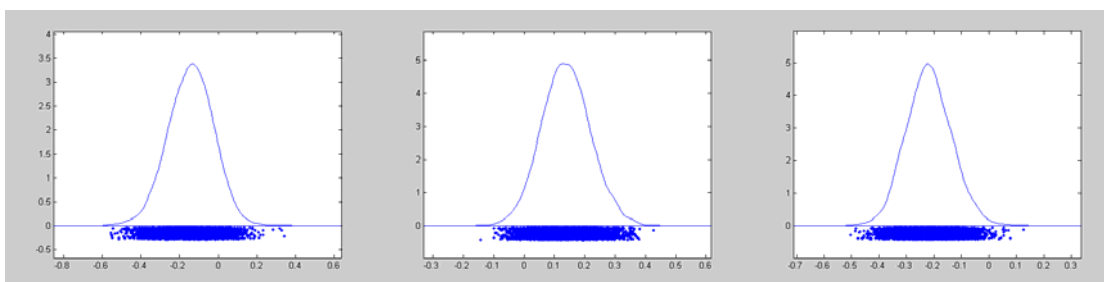
Variable	ba	mg	es	rj	sp	pr	sc	rs
ba	0	0.5	0.5	0	0	0	0	0
mg	0.25	0	0.25	0.25	0.25	0	0	0
es	0.333	0.333	0	0.333	0	0	0	0
rj	0	0.333	0.333	0	0.333	0	0	0
sp	0	0.333	0	0.333	0	0.333	0	0
pr	0	0	0	0	0.5	0	0.5	0
sc	0	0	0	0	0	0.5	0	0.5
rs	0	0	0	0	0	0	1	0

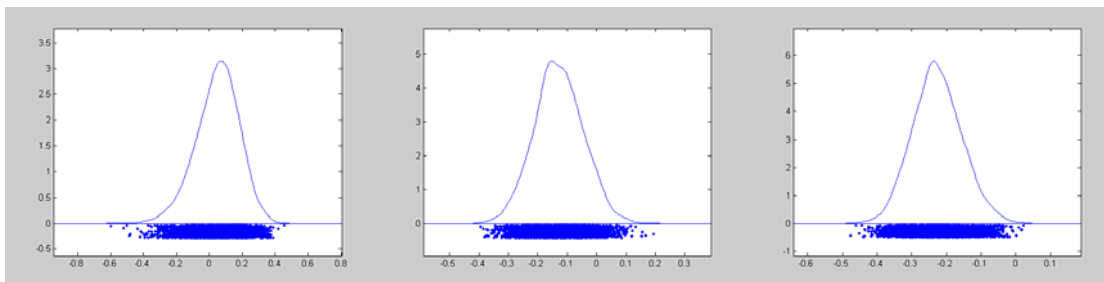
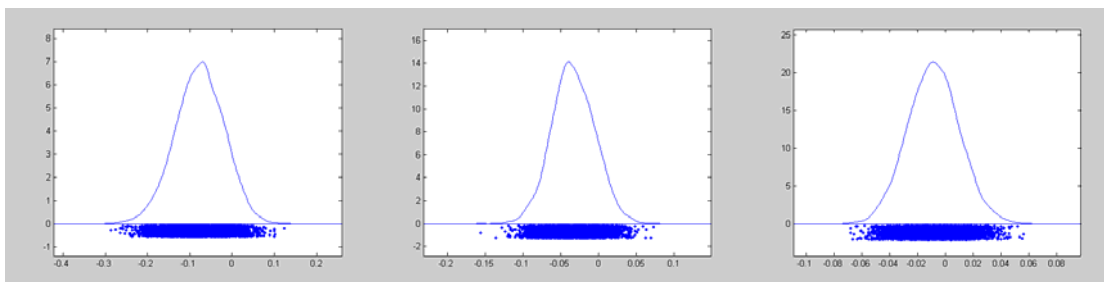
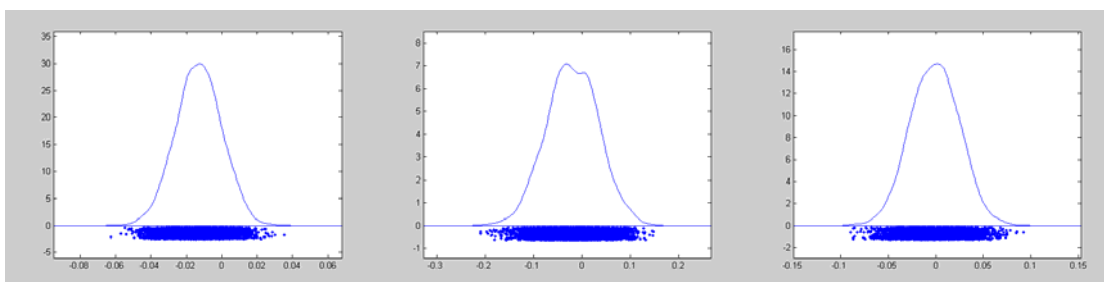
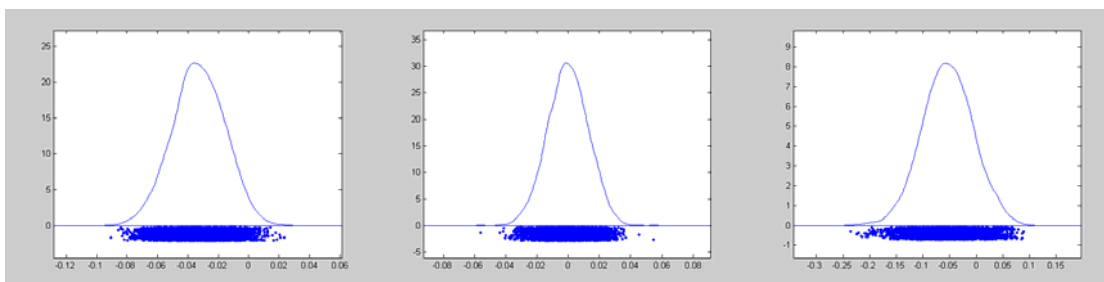
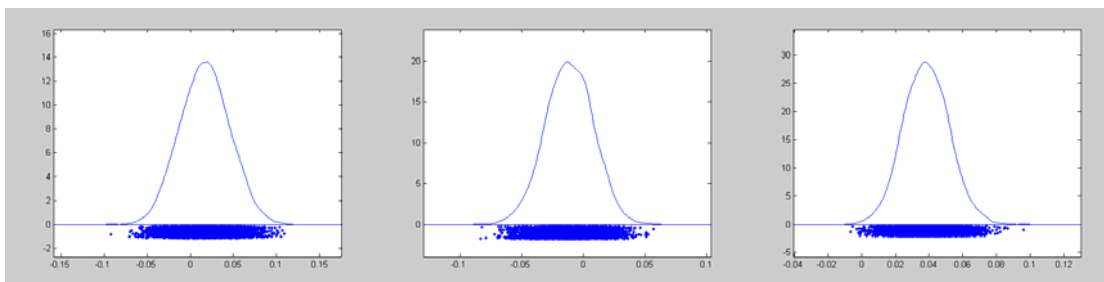
1. Resultados para a equação que tem variação da produção industrial no estado da Bahia como variável dependente

Dependent Variable ba			
R-squared	0.4813		
Rbar-squared	-0.9951		
sige	8.2452		
Nobs, Nvars	63, 8		
ndraws,nomit	5000, 100		
r-value	4		

Variável	Moda do Coef.	Variável	Moda do Coef.
ba lag1	0.059733	sp lag4	-0.000734
ba lag2	-0.055255	pr lag1	-0.054597
ba lag3	0.0012	pr lag2	0.016178
ba lag4	-0.008579	pr lag3	-0.011864
mg lag1	0.350248	pr lag4	0.03786
mg lag2	-0.207629	sc lag1	-0.105524
mg lag3	-0.142848	sc lag2	-0.075916
mg lag4	0.13745	sc lag3	-0.022817
es lag1	-0.218172	sc lag4	0.010946

es lag2	0.042262	rs lag1	-0.046826
es lag3	-0.131227	rs lag2	-0.032686
es lag4	-0.226286	rs lag3	-0.010954
rj lag1	-0.076937	rs lag4	0.011991
rj lag2	-0.033691	ec term ba	-0.011358
rj lag3	-0.00851	ec term mg	-1.207667
rj lag4	-0.013139	ec term es	-0.084879
sp lag1	-0.021423	ec term rj	0.203959
sp lag2	-0.002022	ec term sp	0.543363
sp lag3	-0.032866	constant	-24.66875

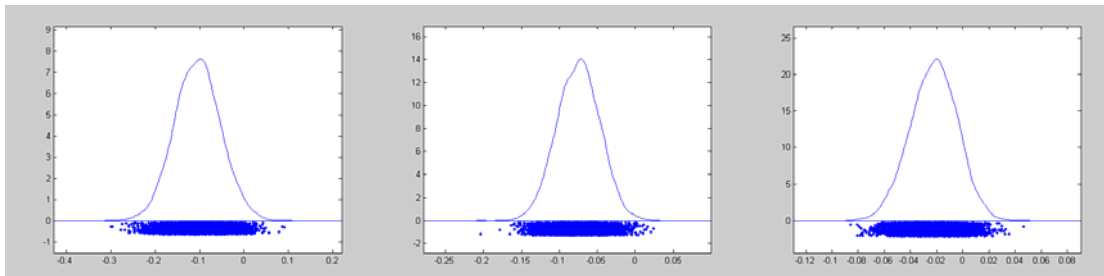
(1) $\beta_{E1,t-1}$ (2) $\beta_{E1,t-2}$ (3) $\beta_{E1,t-3}$ (4) $\beta_{E2,t-4}$ (5) $\beta_{MG,t-1}$ (6) $\beta_{MG,t-2}$ (7) $\beta_{MG,t-3}$ (8) $\beta_{MG,t-4}$ (9) $\beta_{E3,t-1}$

(10) $\beta_{ES,t-2}$ (11) $\beta_{ES,t-3}$ (12) $\beta_{ES,t-4}$ (13) $\beta_{RJ,t-1}$ (14) $\beta_{RJ,t-2}$ (15) $\beta_{RJ,t-3}$ (16) $\beta_{RJ,t-4}$ (17) $\beta_{SP,t-1}$ (18) $\beta_{SP,t-2}$ (19) $\beta_{SP,t-3}$ (20) $\beta_{SP,t-4}$ (21) $\beta_{PP,t-1}$ 

(22) $\beta_{PR,t-2}$

(23) $\beta_{PR,t-3}$

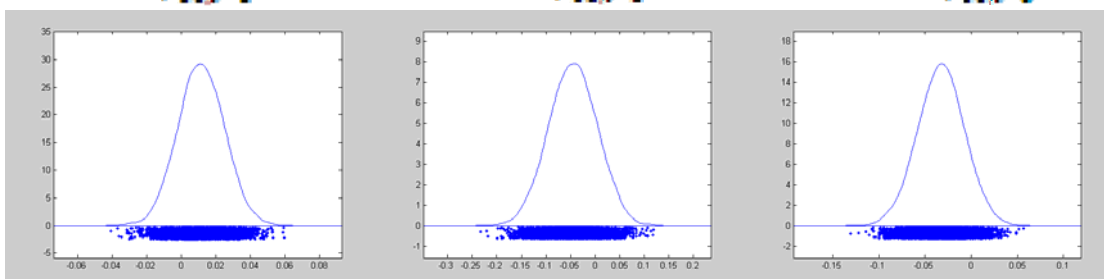
(24) $\beta_{PR,t-4}$



(25) $\beta_{SC,t-1}$

(26) $\beta_{SC,t-2}$

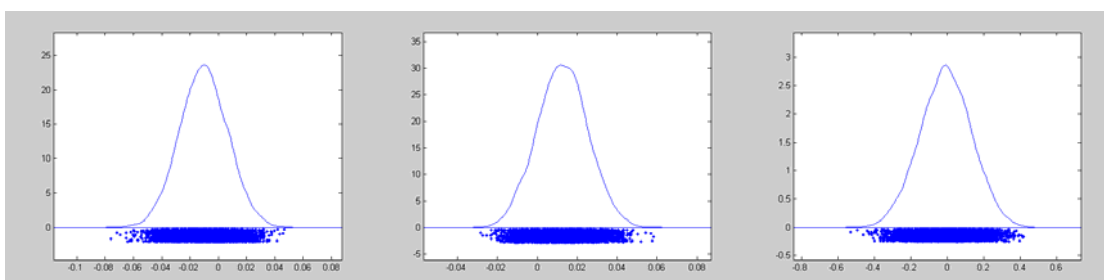
(27) $\beta_{SC,t-3}$



(28) $\beta_{SC,t-4}$

(29) $\beta_{RS,t-1}$

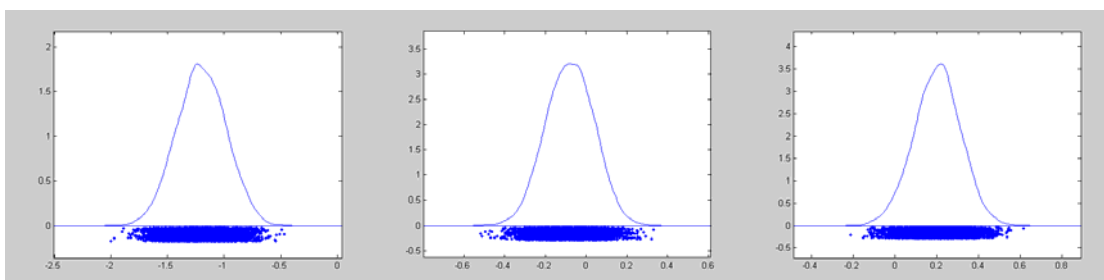
(30) $\beta_{RS,t-2}$



(31) $\beta_{RS,t-3}$

(32) $\beta_{RS,t-4}$

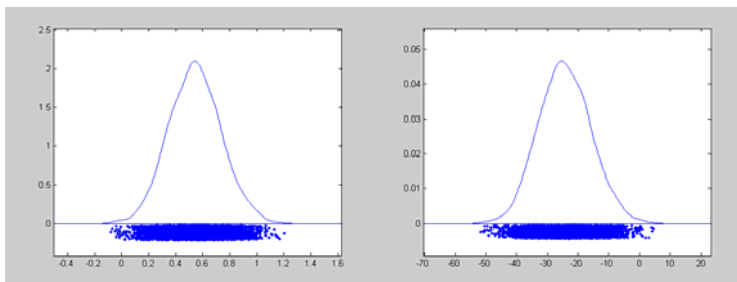
(33) ec term ba



(34) Ec term mg

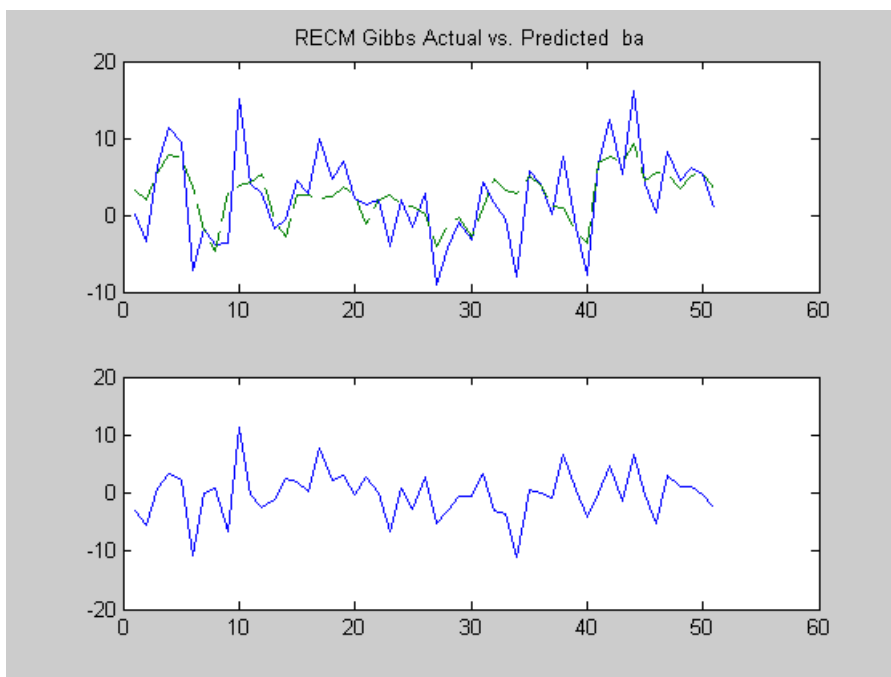
(35) ec term es

(36) ec term rj



(37) ec term sp

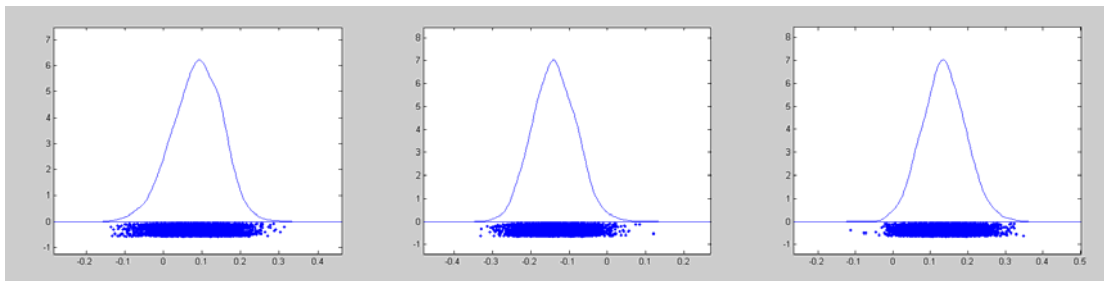
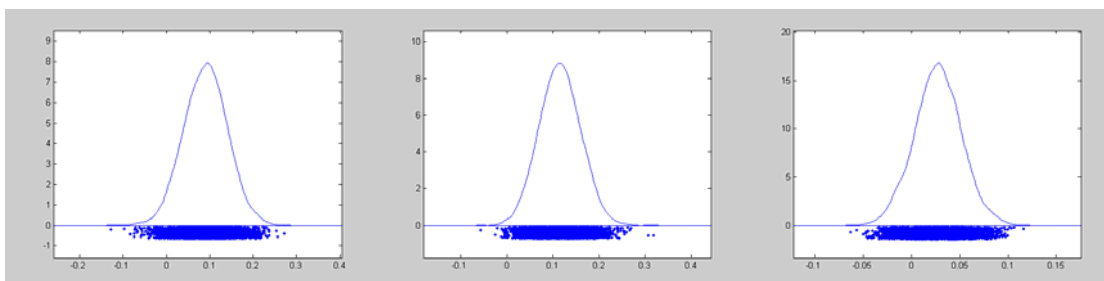
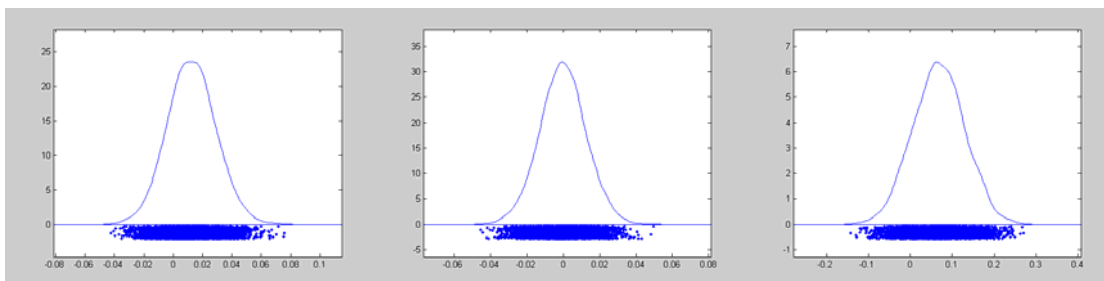
(38) C_{BA}

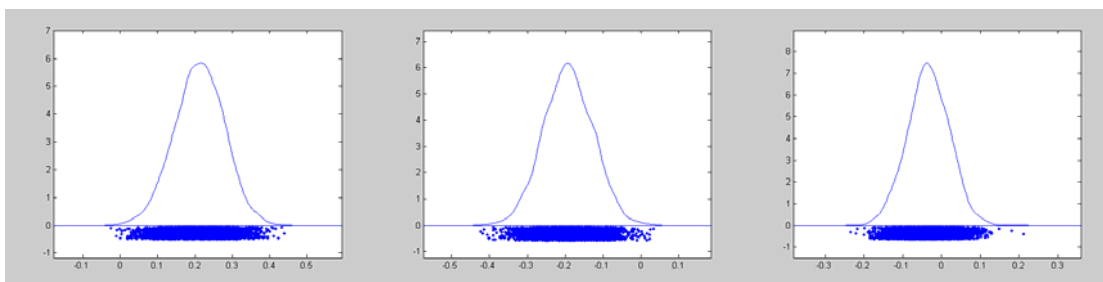
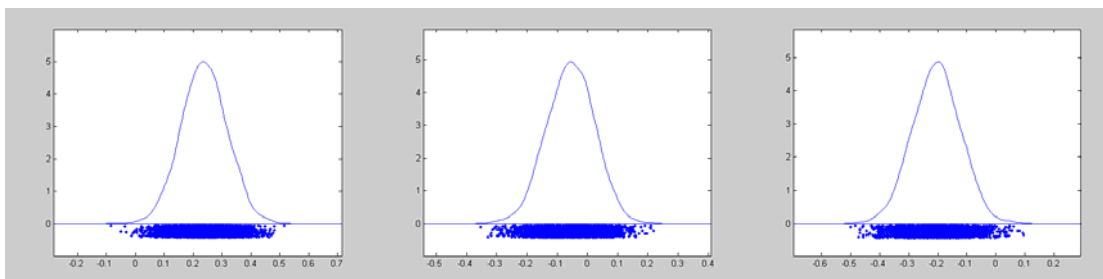
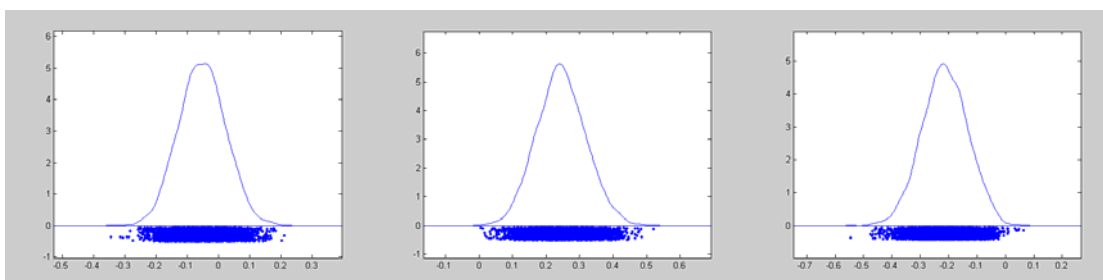
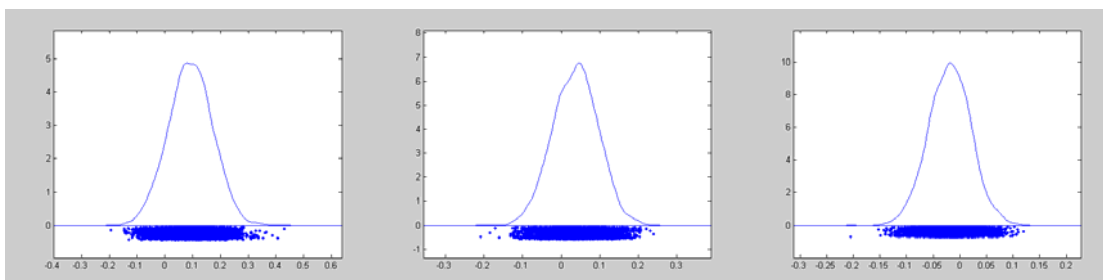
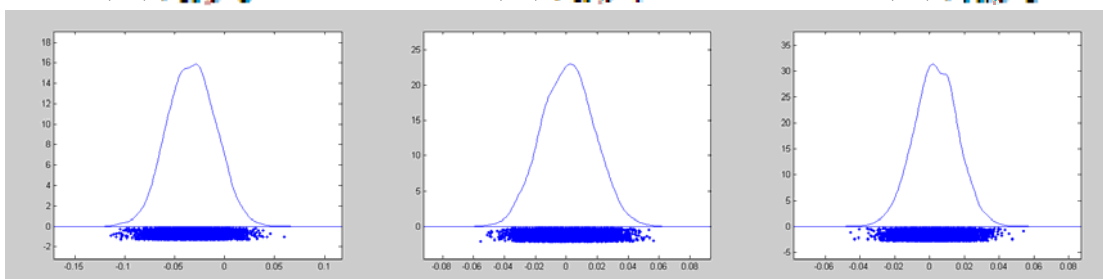


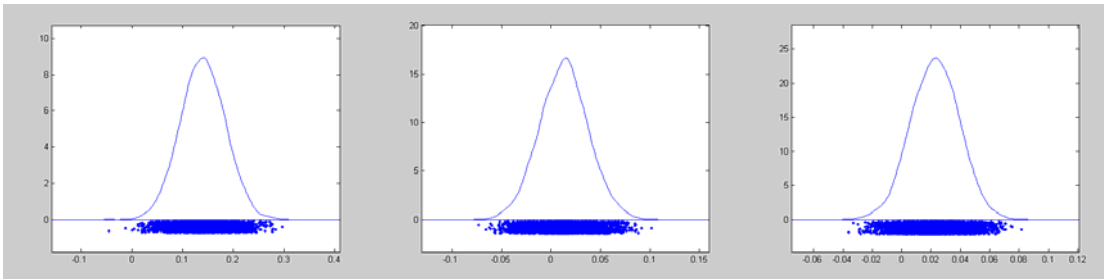
2. Resultados para a equação que tem variação da produção industrial no estado de Minas Gerais como variável dependente

Dependent Variable mg			
R-squared	0.6551		
Rbar-squared	-0.3264		
sige	3.0632		
Nobs, Nvars	63, 8		
ndraws,nomit	5000, 100		
r-value	4		
Variável	Moda do Coef.	Variável	Moda do Coef.
ba lag1	0.084338	sp lag4	0.033797
ba lag2	-0.138878	pr lag1	-0.019667
ba lag3	0.134565	pr lag2	-0.034014
ba lag4	0.087908	pr lag3	0.001047
mg lag1	0.11336	pr lag4	0.003097
mg lag2	0.027469	sc lag1	0.138452

mg lag3	0.012167	sc lag2	0.012058
mg lag4	-0.000074	sc lag3	0.021921
es lag1	0.067015	sc lag4	0.005999
es lag2	0.208154	rs lag1	-0.006597
es lag3	-0.193758	rs lag2	-0.014682
es lag4	-0.035671	rs lag3	-0.012682
rj lag1	0.235933	rs lag4	-0.01193
rj lag2	-0.057234	ec term ba	-0.375793
rj lag3	-0.207611	ec term mg	-0.085017
rj lag4	-0.05355	ec term es	-0.414301
sp lag1	0.241867	ec term rj	0.080884
sp lag2	-0.213851	ec term sp	0.429261
sp lag3	0.090588	constant	12.568592

(1) β_{BAT-1} (2) β_{BAT-2} (3) β_{BAT-3} (4) β_{BAT-4} (5) β_{MGT-1} (6) β_{MGT-2} (7) β_{MGT-3} (8) β_{MGT-4} (9) β_{ES-1}

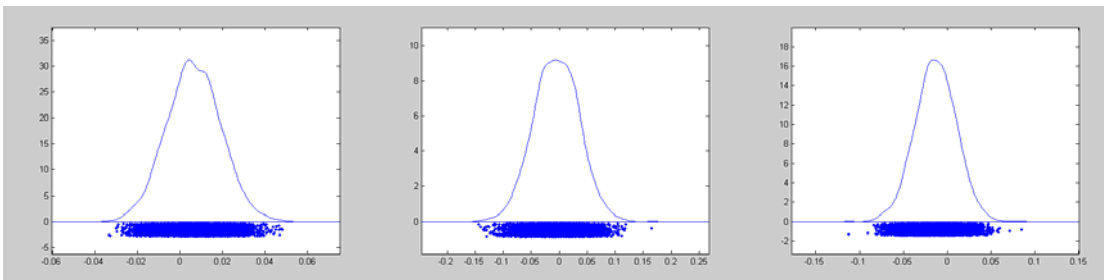
(10) $\beta_{ES,t-2}$ (11) $\beta_{ES,t-3}$ (12) $\beta_{ES,t-4}$ (13) $\beta_{RJ,t-1}$ (14) $\beta_{RJ,t-2}$ (15) $\beta_{RJ,t-3}$ (16) $\beta_{SP,t-4}$ (17) $\beta_{SP,t-1}$ (18) $\beta_{SP,t-2}$ (19) $\beta_{PR,t-3}$ (20) $\beta_{PR,t-4}$ (21) $\beta_{PR,t-1}$ (22) $\beta_{HML,t-2}$ (23) $\beta_{HML,t-3}$ (24) $\beta_{HML,t-4}$



(25) β_{SCI_T-1}

(26) β_{SCI_T-2}

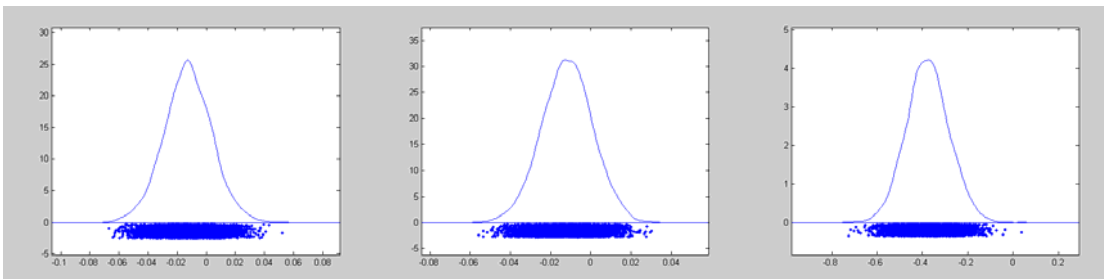
(27) β_{SCI_T-3}



(28) β_{SCI_T-4}

(29) β_{RS_T-1}

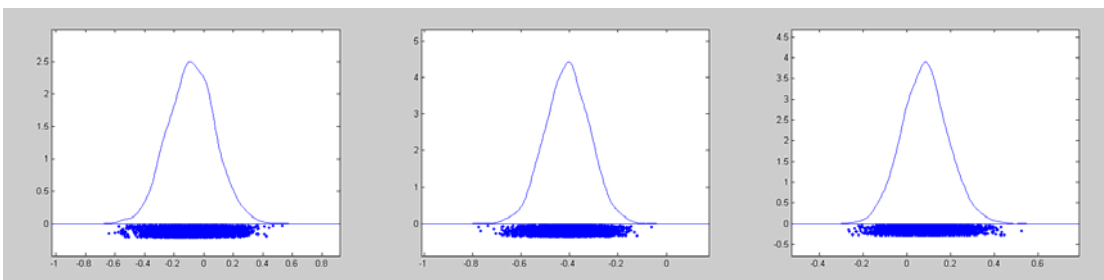
(30) β_{RS_T-2}



(31) β_{RS_T-3}

(32) β_{RS_T-4}

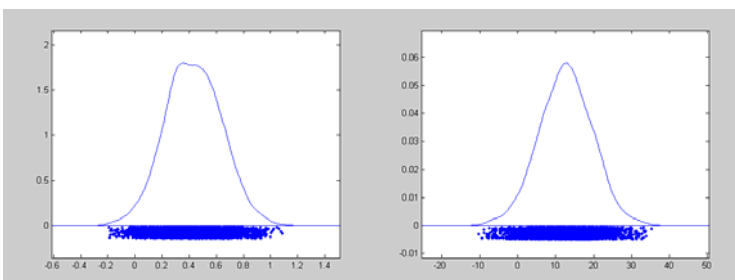
(33) ec term ba



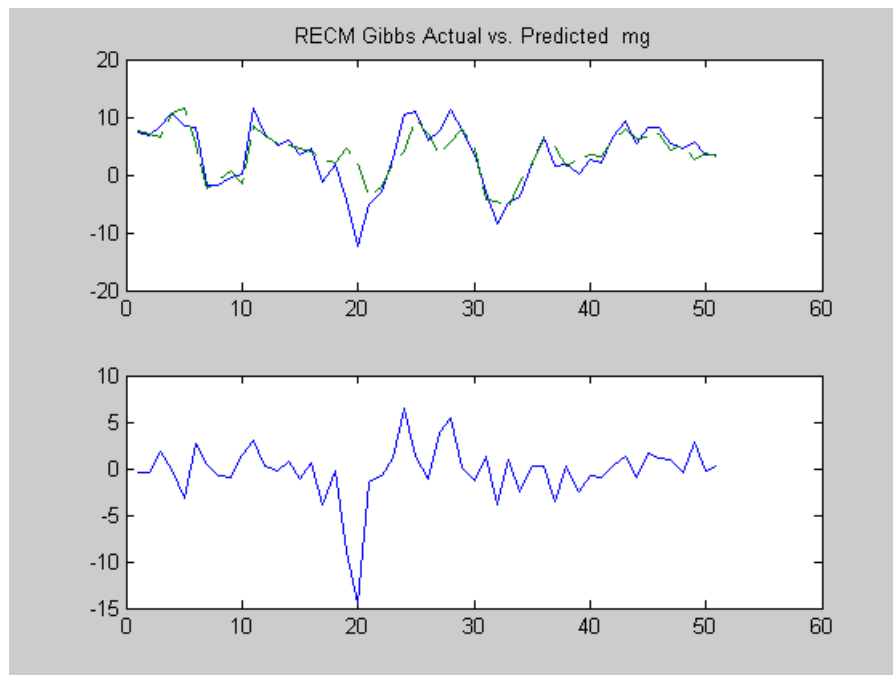
(34) Ec term mg

(35) ec term es

(36) ec term rj



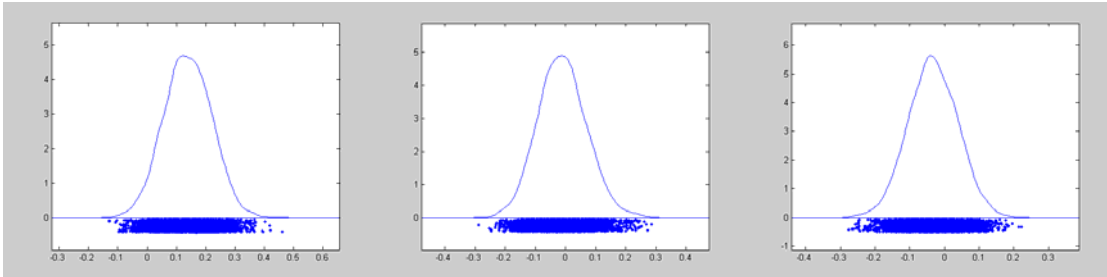
(37) ec term sp

(38) C_{MG} 

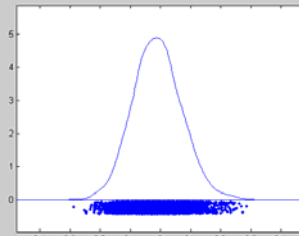
3. Resultados para a equação que tem variação da produção industrial no estado do Espírito Santo como variável dependente

Dependent Variable es			
R-squared	0.5964		
Rbar-squared	-0.5522		
sige	6.7494		
Nobs, Nvars	63, 8		
ndraws,nomit	5000, 100		
r-value	4		
Variável	Moda do Coef.	Variável	Moda do Coef.
ba lag1	0.134465	sp lag4	0.000514
ba lag2	-0.014229	pr lag1	-0.100924
ba lag3	-0.033098	pr lag2	-0.013046
ba lag4	-0.145451	pr lag3	-0.008086
mg lag1	0.327989	pr lag4	0.026305
mg lag2	-0.307179	sc lag1	0.004675
mg lag3	0.081627	sc lag2	0.002172
mg lag4	-0.405604	sc lag3	0.004832
es lag1	0.008087	sc lag4	-0.001245
es lag2	-0.013047	rs lag1	-0.033143
es lag3	-0.014045	rs lag2	0.005964
es lag4	-0.022757	rs lag3	-0.00255
rj lag1	0.309938	rs lag4	-0.000715

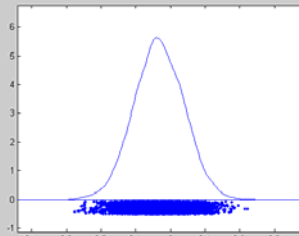
rj lag2	0.047882	ec term ba	-0.111397
rj lag3	0.194115	ec term mg	-0.449543
rj lag4	0.044422	ec term es	-0.258386
sp lag1	-0.074469	ec term rj	0.142374
sp lag2	-0.014622	ec term sp	-0.079971
sp lag3	-0.002564	constant	4.198378



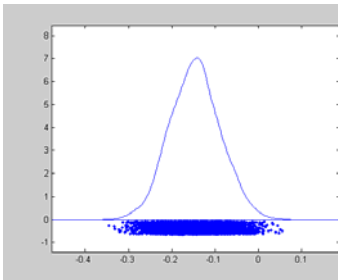
(1) $\beta_{BA,t-1}$



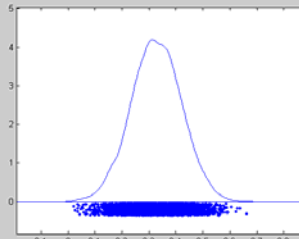
(2) $\beta_{BA,t-2}$



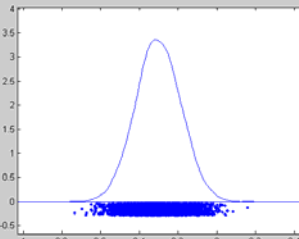
(3) $\beta_{BA,t-3}$



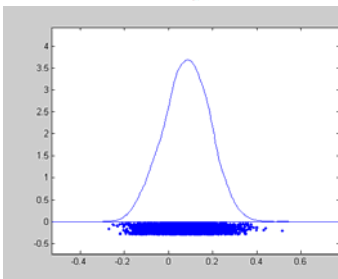
(4) $\beta_{BA,t-4}$



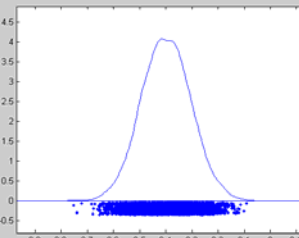
(5) $\beta_{MG,t-1}$



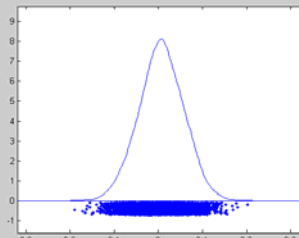
(6) $\beta_{MG,t-2}$



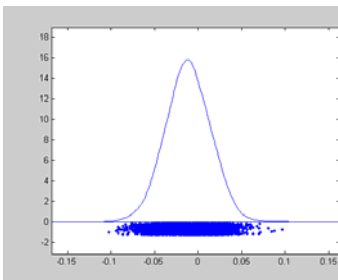
(7) $\beta_{MG,t-3}$



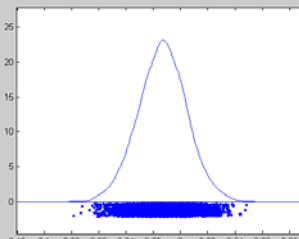
(8) $\beta_{MG,t-4}$



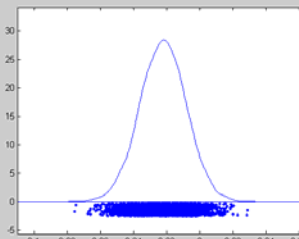
(9) $\beta_{ES,t-1}$



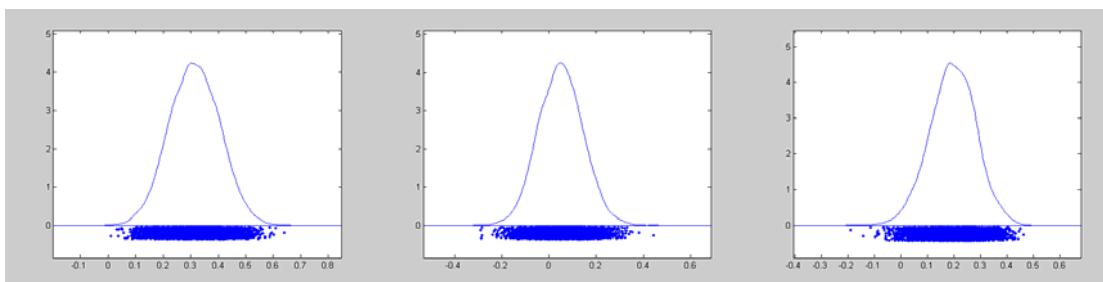
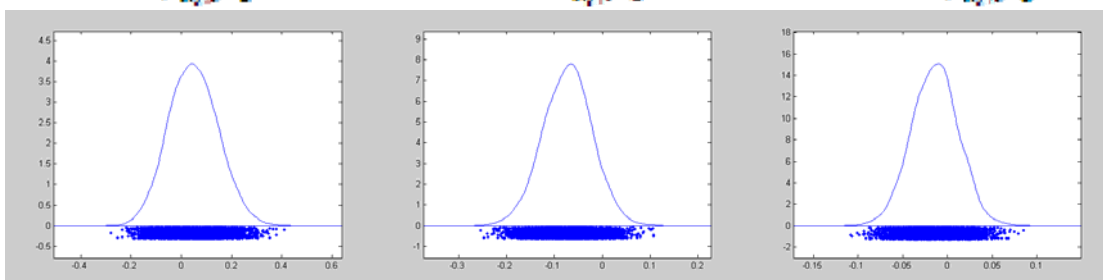
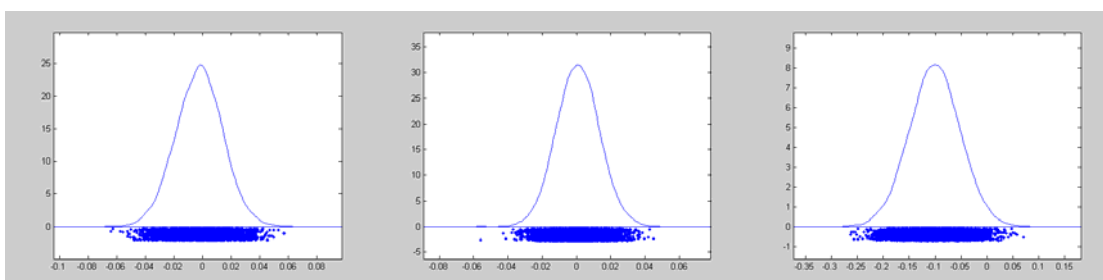
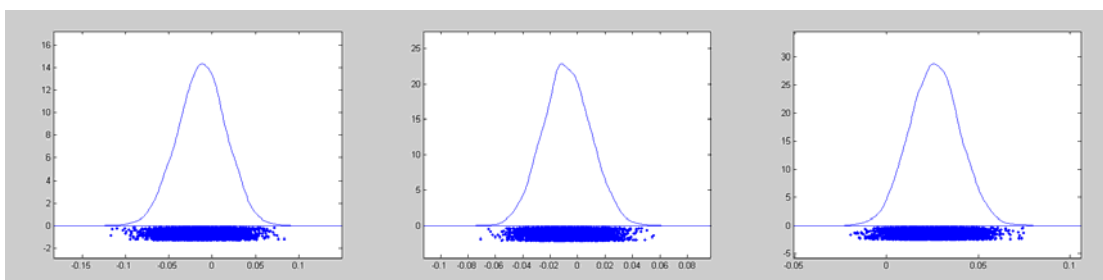
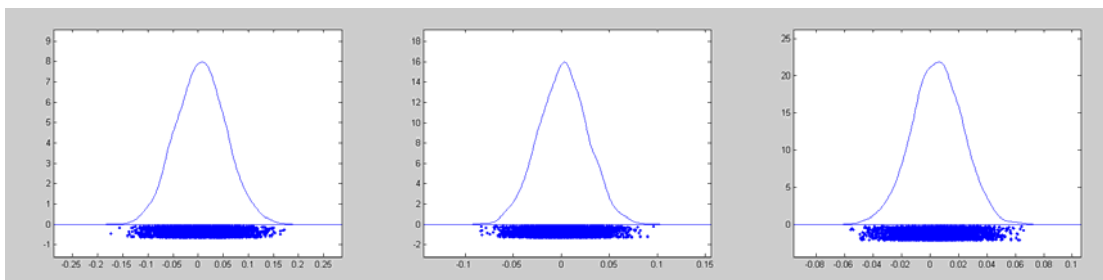
(10) $\beta_{ES,t-2}$

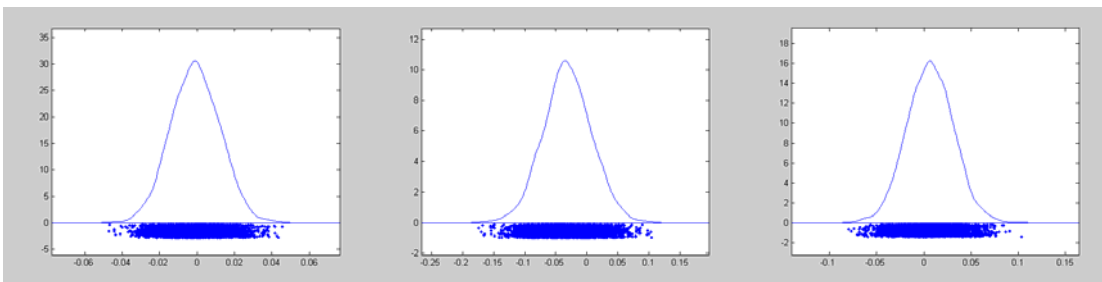


(11) $\beta_{ES,t-3}$



(12) $\beta_{ES,t-4}$

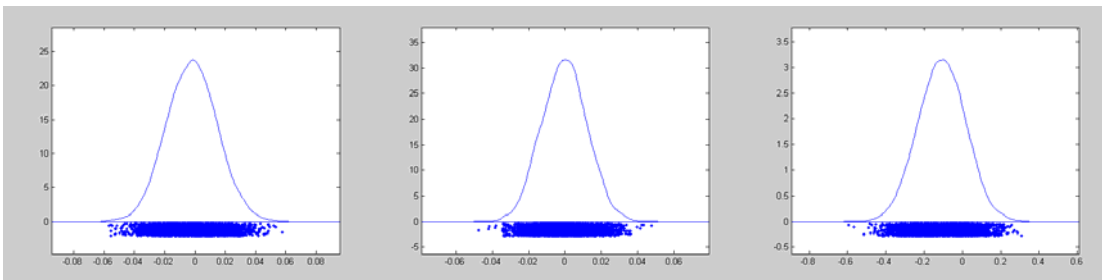
(13) $\beta_{R|S}-1$ (14) $\beta_{R|S}-2$ (15) $\beta_{R|S}-3$ (16) $\beta_{S|S}-1$ (17) $\beta_{S|S}-1$ (18) $\beta_{S|S}-2$ (19) $\beta_{S|L}-3$ (20) $\beta_{S|L}-4$ (21) $\beta_{S|L}-1$ (22) $\beta_{H|L}-2$ (23) $\beta_{H|L}-3$ (24) $\beta_{H|L}-4$ (25) $\beta_{S|C}-1$ (26) $\beta_{S|C}-2$ (27) $\beta_{S|C}-3$



(28) $\beta_{SC,t-1}$

(29) $\beta_{RF,t-1}$

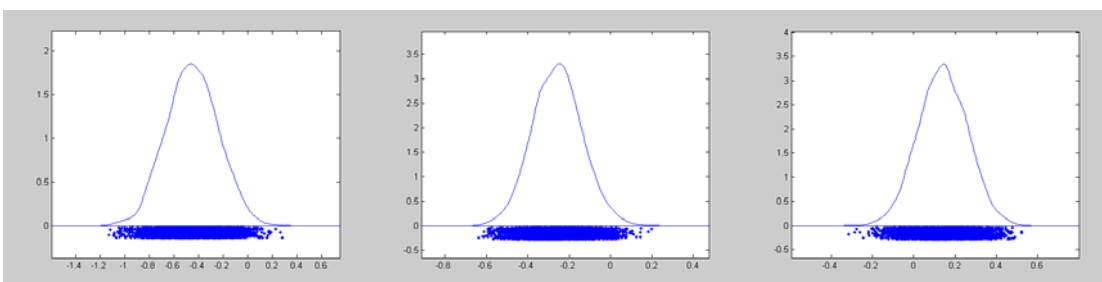
(30) $\beta_{RF,t-2}$



(31) $\beta_{RF,t-2}$

(32) $\beta_{RF,t-1}$

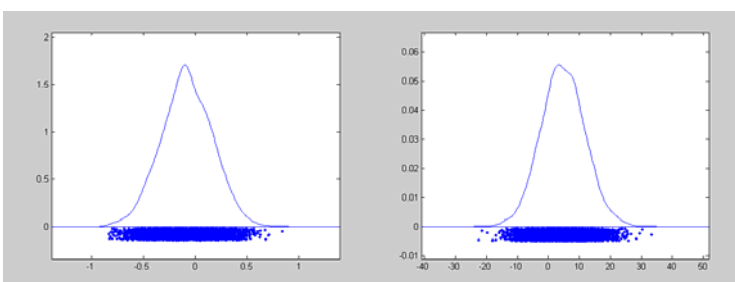
(33) ec term ba



(34) Ec term mg

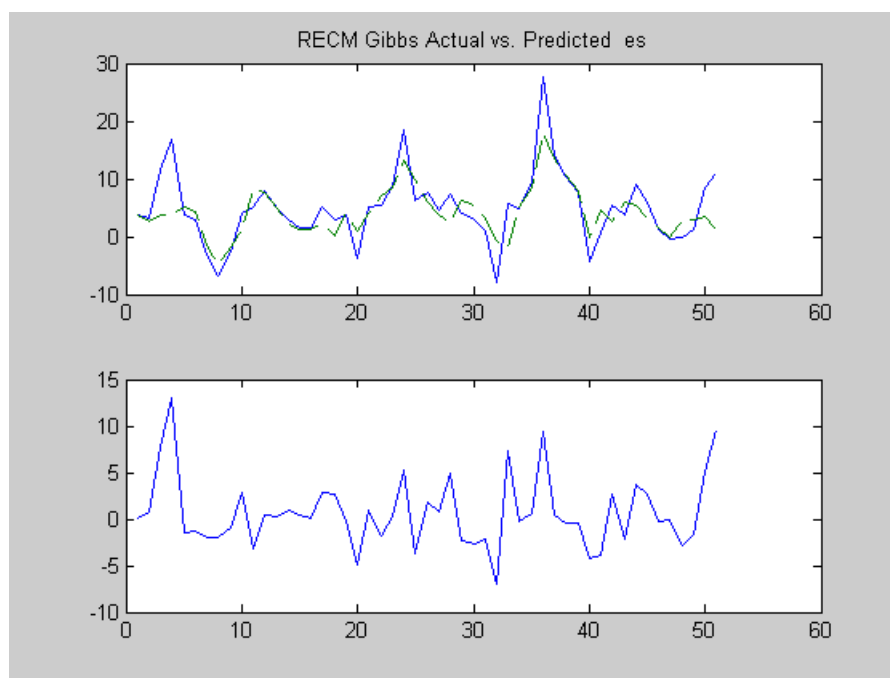
(35) ec term es

(36) ec term rj



(37) ec term sp

(38) C_{ES}

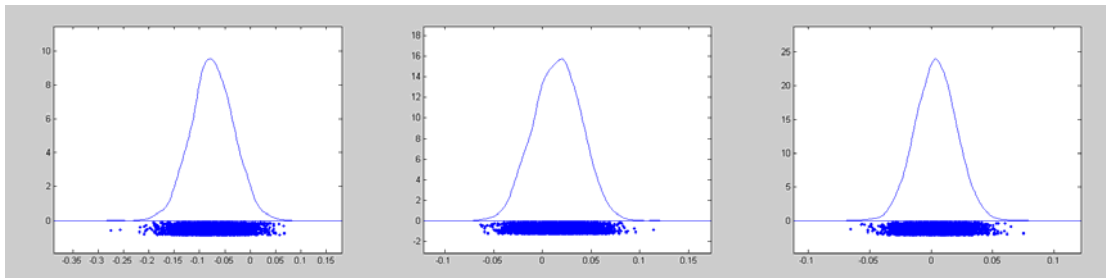


4. Resultados para a equação que tem variação da produção industrial no estado do Rio de Janeiro como variável dependente

Dependent Variable rj
 R-squared 0.6812
 Rbar-squared -0.2262
 sige 3.8561
 Nobs, Nvars 63, 8
 ndraws,nomit 5000, 100
 r-value 4

Variável	Moda do Coef.	Variável	Moda do Coef.
ba lag1	-0.075204	sp lag4	-0.067557
ba lag2	0.015449	pr lag1	-0.172954
ba lag3	0.003355	pr lag2	-0.037097
ba lag4	0.001873	pr lag3	-0.009837
mg lag1	0.145744	pr lag4	0.000887
mg lag2	-0.416011	sc lag1	-0.042696
mg lag3	0.151021	sc lag2	0.001473
mg lag4	0.088411	sc lag3	0.011715
es lag1	0.37216	sc lag4	0.01594
es lag2	-0.139775	rs lag1	-0.065963
es lag3	-0.184175	rs lag2	-0.008085
es lag4	0.010881	rs lag3	-0.001758
rj lag1	0.039682	rs lag4	-0.004641
rj lag2	-0.035313	ec term ba	0.600191

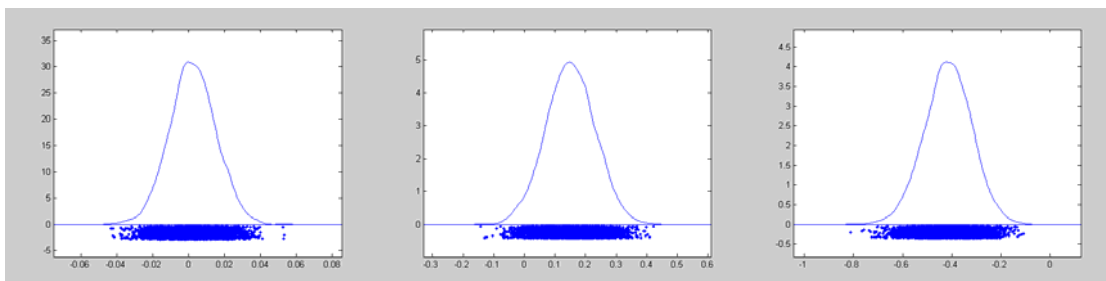
rj lag3	-0.022933	ec term mg	-0.354902
rj lag4	-0.01531	ec term es	-0.610012
sp lag1	0.075395	ec term rj	0.647984
sp lag2	0.430585	ec term sp	0.53047
sp lag3	-0.288606	constant	-13.564244



(1) $\beta_{BA,t-1}$

(2) $\beta_{BA,t-2}$

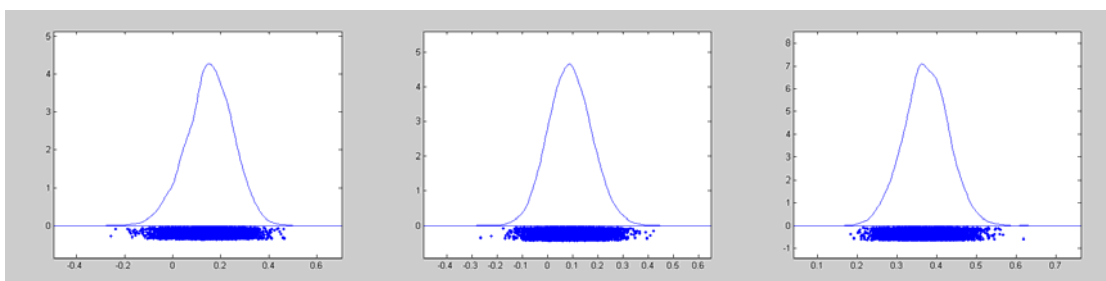
(3) $\beta_{BA,t-3}$



(4) $\beta_{BA,t-4}$

(5) $\beta_{MG,t-1}$

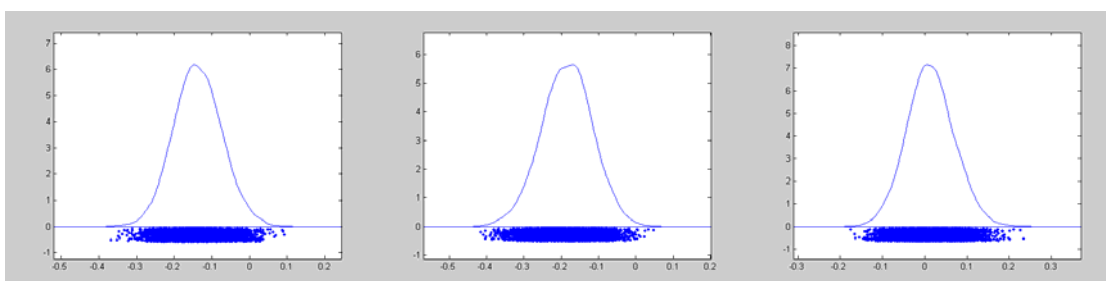
(6) $\beta_{MG,t-2}$



(7) $\beta_{MG,t-3}$

(8) $\beta_{MG,t-4}$

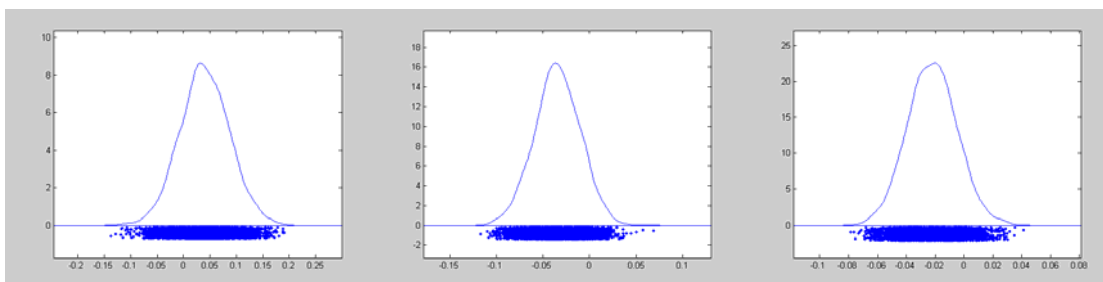
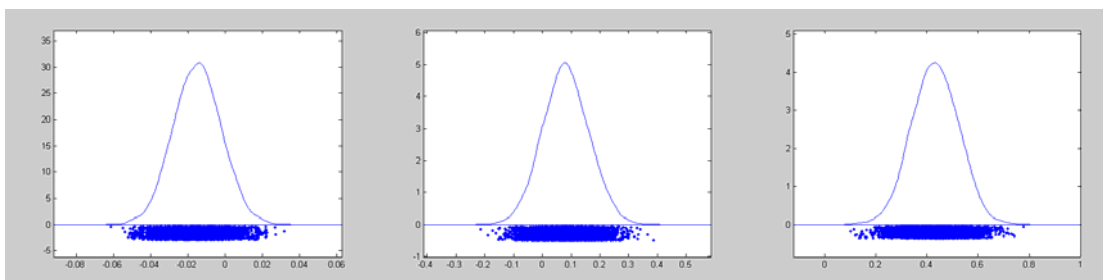
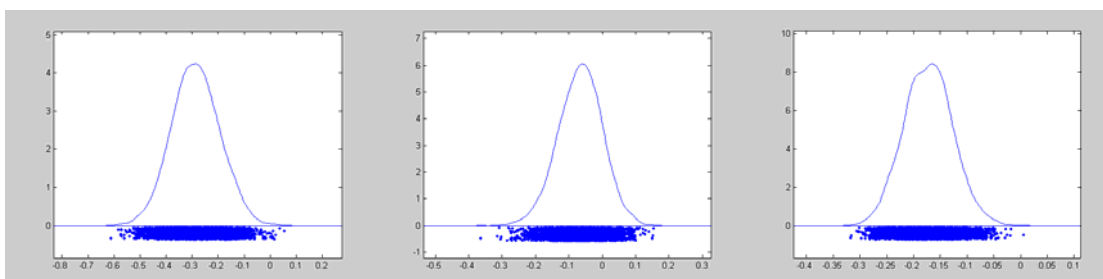
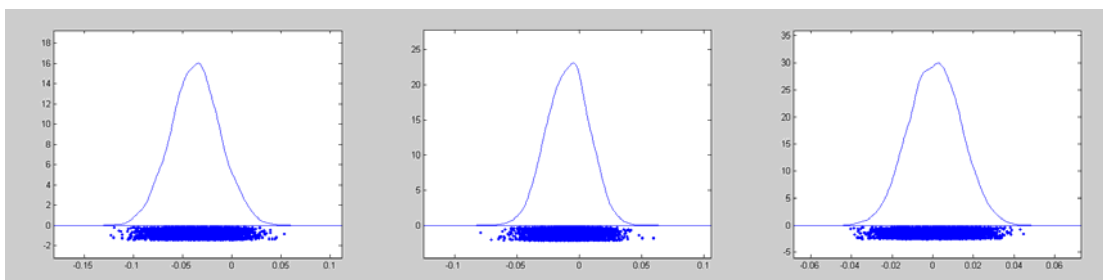
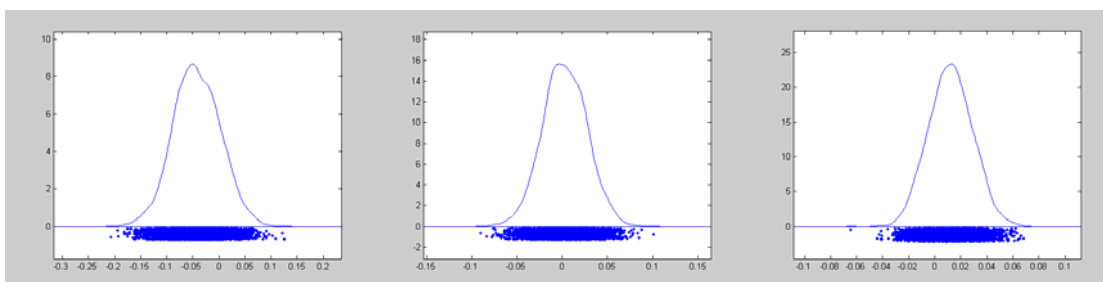
(9) $\beta_{ES,t-1}$

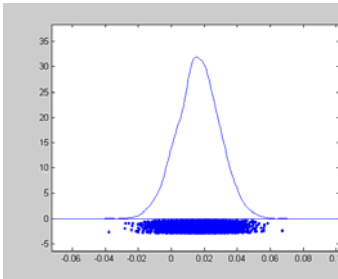
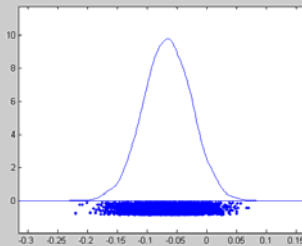
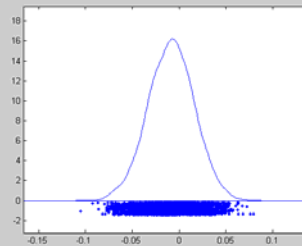
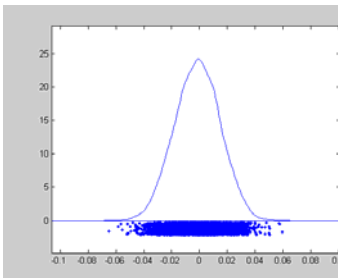
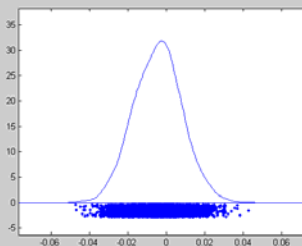
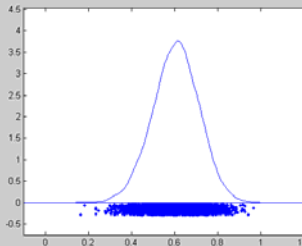
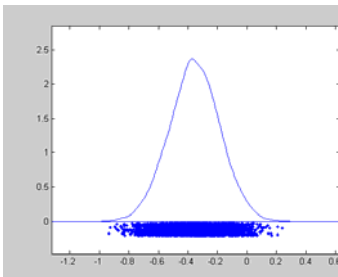
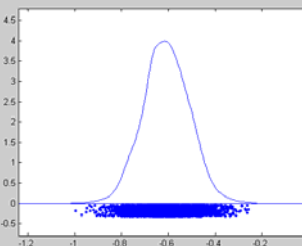


(10) $\beta_{ES,t-2}$

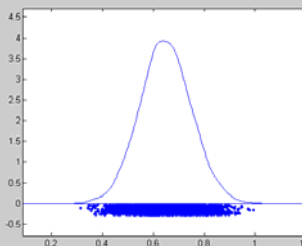
(11) $\beta_{ES,t-3}$

(12) $\beta_{ES,t-4}$

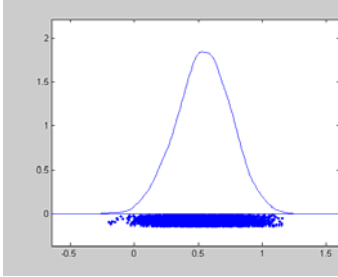
(13) $\beta_{R|t-1}$ (14) $\beta_{R|t-2}$ (15) $\beta_{R|t-3}$ (16) $\beta_{R|t-4}$ (17) $\beta_{SP|t-1}$ (18) $\beta_{SP|t-2}$ (19) $\beta_{SP|t-3}$ (20) $\beta_{SP|t-4}$ (21) $\beta_{PP|t-1}$ (22) $\beta_{PP|t-2}$ (23) $\beta_{PP|t-3}$ (24) $\beta_{PP|t-4}$ 

(25) $\beta_{SC,t-1}$ (26) $\beta_{SC,t-2}$ (27) $\beta_{SC,t-3}$ (28) $\beta_{SC,t-4}$ (29) $\beta_{RS,t-1}$ (30) $\beta_{RS,t-2}$ (31) $\beta_{RS,t-3}$ (32) $\beta_{RS,t-4}$ 

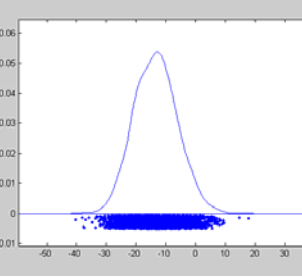
(33) ec term ba



(34) Ec term mg



(35) ec term es

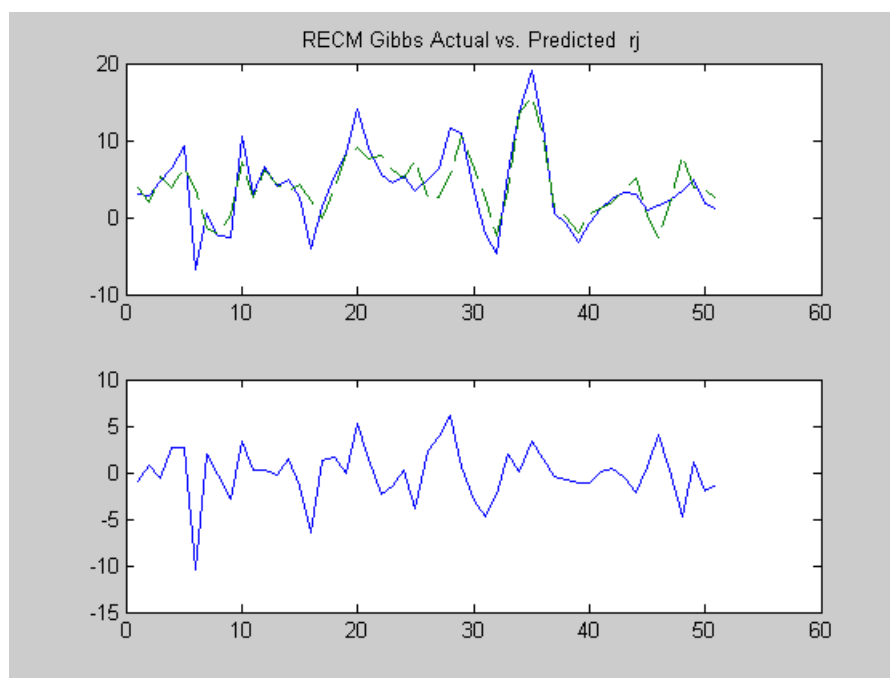


(36) ec term rj



(37) ec term sp

(38) c_{Rj} 

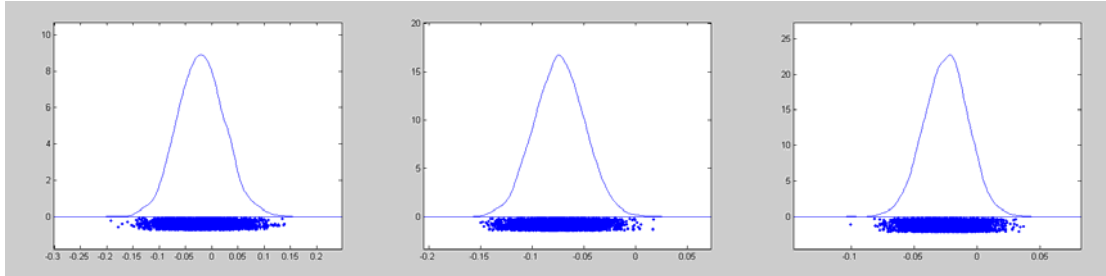
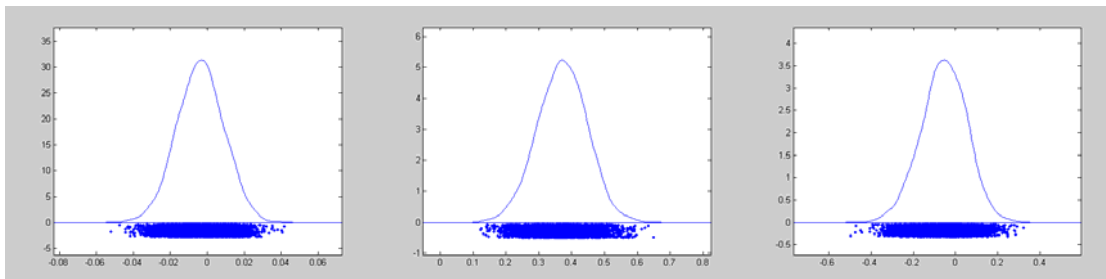
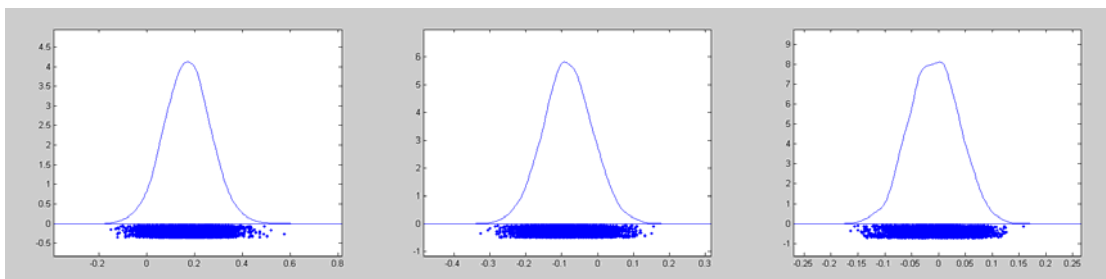
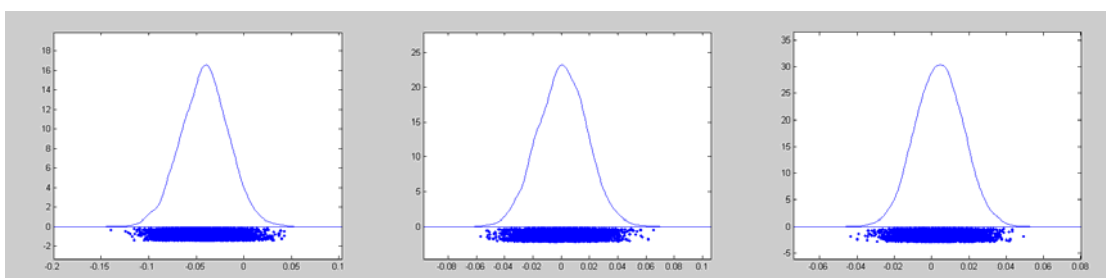


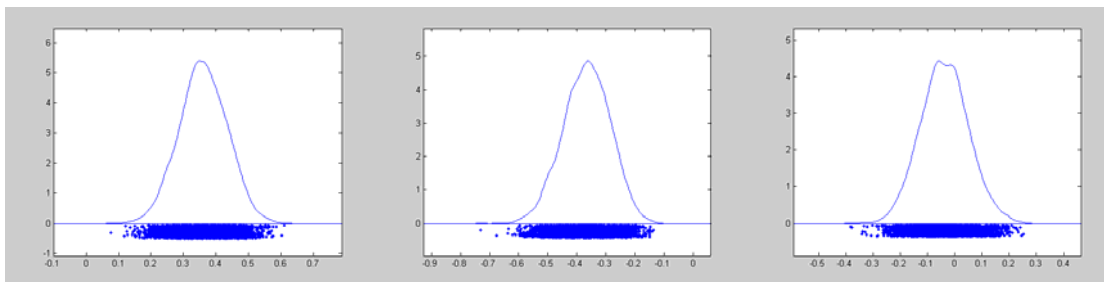
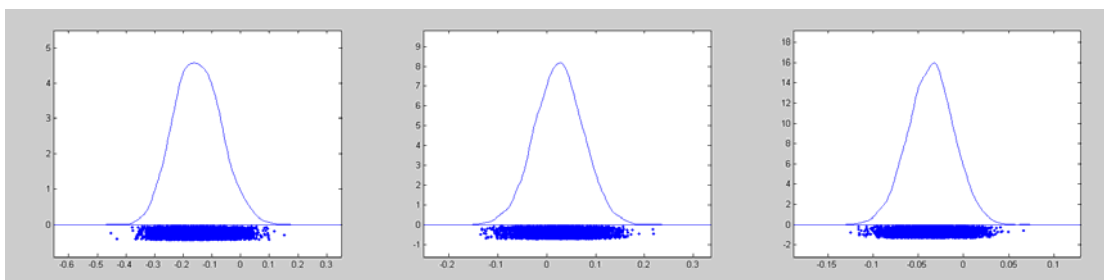
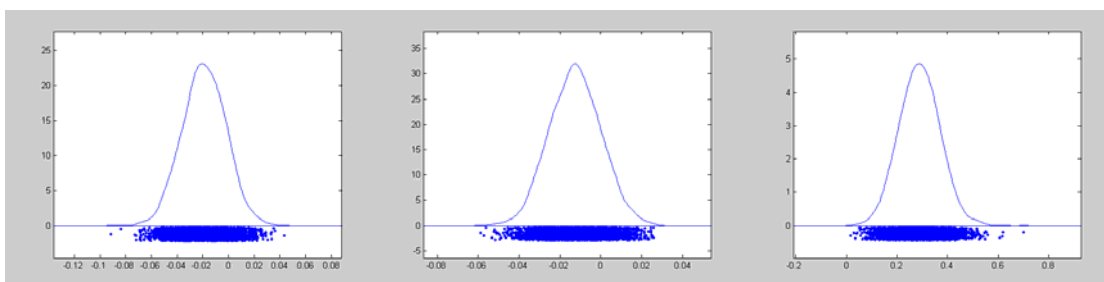
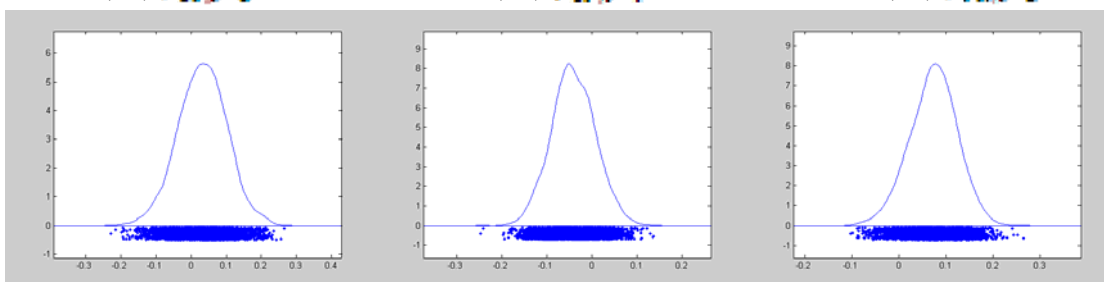
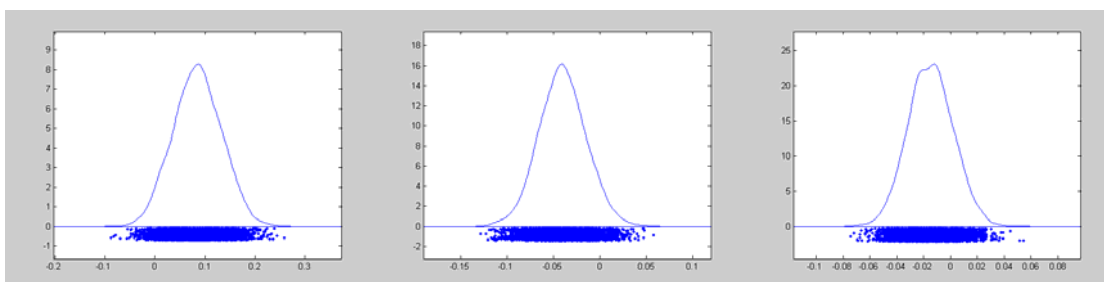
5. Resultados para a equação que tem variação da produção industrial no estado de São Paulo como variável dependente

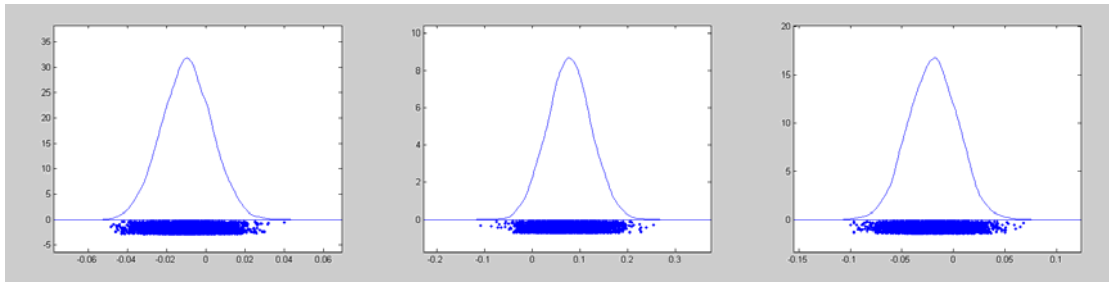
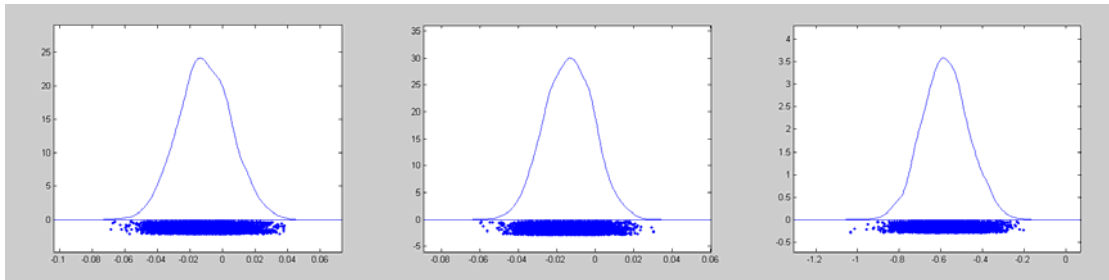
Dependent Variable sp
 R-squared 0.8538
 Rbar-squared 0.4376
 sige 4.0526
 Nobs, Nvars 63, 8
 ndraws,nomit 5000, 100
 r-value 4

Variável	Moda do Coef.	Variável	Moda do Coef.
ba lag1	-0.021143	sp lag4	-0.012814
ba lag2	-0.074695	pr lag1	0.284804
ba lag3	-0.026037	pr lag2	0.033978
ba lag4	-0.00438	pr lag3	-0.042421
mg lag1	0.368663	pr lag4	0.072459
mg lag2	-0.056445	sc lag1	0.082836
mg lag3	0.166773	sc lag2	-0.041279
mg lag4	-0.085576	sc lag3	-0.015473
es lag1	-0.008323	sc lag4	-0.009772
es lag2	-0.041464	rs lag1	0.073773
es lag3	0.000699	rs lag2	-0.01932
es lag4	0.004155	rs lag3	-0.010832
rj lag1	0.363205	rs lag4	-0.014312
rj lag2	-0.369222	ec term ba	-0.584048
rj lag3	-0.037968	ec term mg	-0.318823

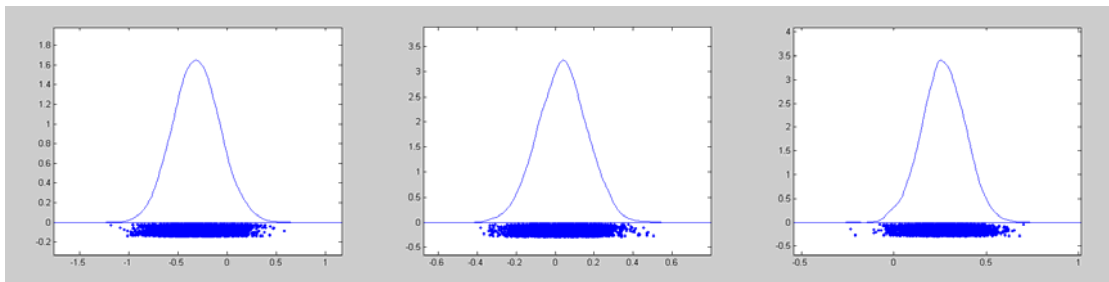
ry lag4	-0.156283	ec term es	0.03243
sp lag1	0.02563	ec term ry	0.260765
sp lag2	-0.037012	ec term sp	0.493096
sp lag3	-0.020034	constant	-19.189934

(1) $\beta_{BA,t-1}$ (2) $\beta_{BA,t-2}$ (3) $\beta_{BA,t-3}$ (4) $\beta_{BA,t-4}$ (5) $\beta_{MG,t-1}$ (6) $\beta_{MG,t-2}$ (7) $\beta_{MG,t-3}$ (8) $\beta_{MG,t-4}$ (9) $\beta_{ES,t-1}$ (10) $\beta_{ES,t-2}$ (11) $\beta_{ES,t-3}$ (12) $\beta_{ES,t-4}$

(13) $\beta_{R|t-1}$ (14) $\beta_{R|t-2}$ (15) $\beta_{R|t-3}$ (16) $\beta_{R|t-4}$ (17) $\beta_{SP|t-1}$ (18) $\beta_{SP|t-2}$ (19) $\beta_{SP|t-3}$ (20) $\beta_{SP|t-4}$ (21) $\beta_{PP|t-1}$ (22) $\beta_{PP|t-2}$ (23) $\beta_{PP|t-3}$ (24) $\beta_{PP|t-4}$ (25) $\beta_{S|t-1}$ (26) $\beta_{S|t-2}$ (27) $\beta_{S|t-3}$

(28) $\beta_{SC,t-1}$ (29) $\beta_{RF,t-1}$ (30) $\beta_{RF,t-2}$ (31) $\beta_{RF,t-3}$ (32) $\beta_{RF,t-4}$

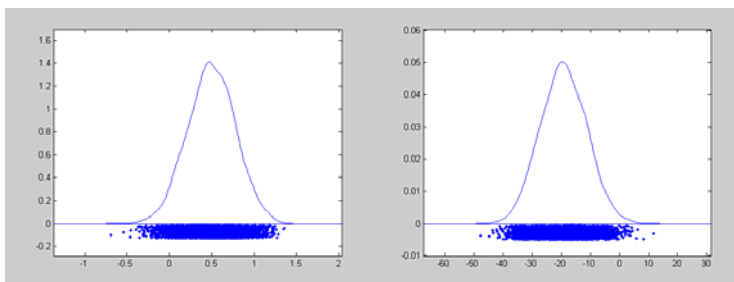
(33) ec term ba



(34) Ec term mg

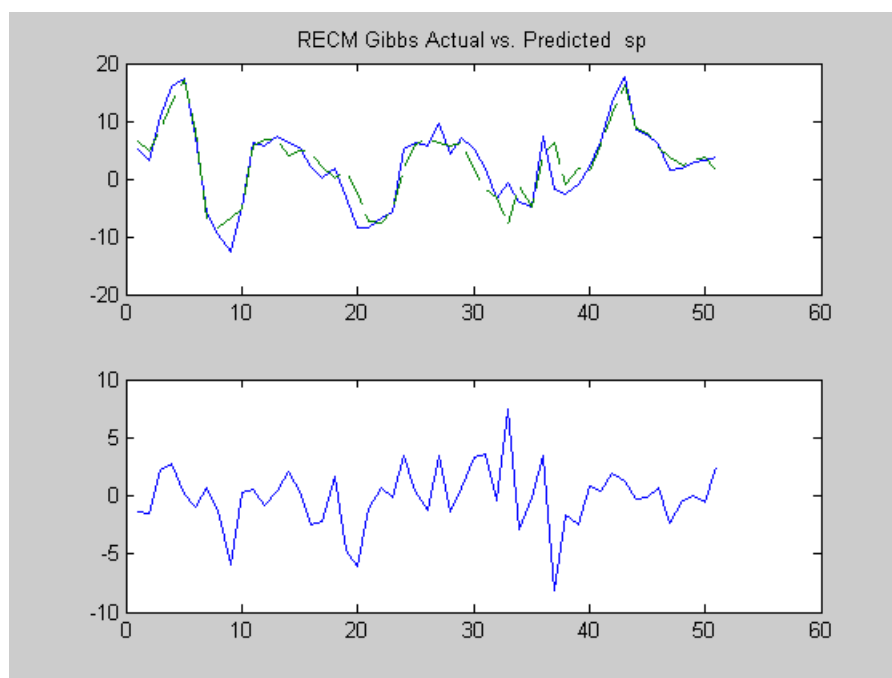
(35) ec term es

(36) ec term rj



(37) ec term sp

(38) C_{sp}

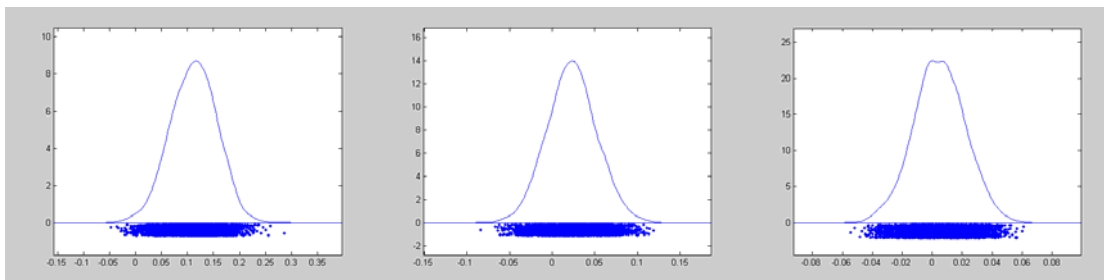


6. Resultados para a equação que tem variação da produção industrial no estado do Paraná como variável dependente

Dependent Variable pr	
R-squared	0.6786
Rbar-squared	-0.236
sige	8.0849
Nobs, Nvars	63, 8
ndraws,nomit	5000, 100
r-value	4

Variável	Moda do Coef.	Variável	Moda do Coef.
ba lag1	0.11193	sp lag4	-0.038937
ba lag2	0.023188	pr lag1	0.121993
ba lag3	0.003585	pr lag2	0.086034
ba lag4	-0.024071	pr lag3	0.002142
mg lag1	-0.082482	pr lag4	-0.018305
mg lag2	-0.046421	sc lag1	0.374677
mg lag3	-0.001732	sc lag2	-0.187214
mg lag4	-0.001464	sc lag3	-0.354132
es lag1	-0.018677	sc lag4	0.018931
es lag2	-0.058642	rs lag1	0.077283
es lag3	-0.016904	rs lag2	-0.011063
es lag4	-0.021833	rs lag3	-0.005006
rj lag1	0.088706	rs lag4	-0.00155
rj lag2	0.001547	ec term ba	-0.172861
rj lag3	-0.011136	ec term mg	0.425751
rj lag4	-0.026611	ec term es	-0.612045

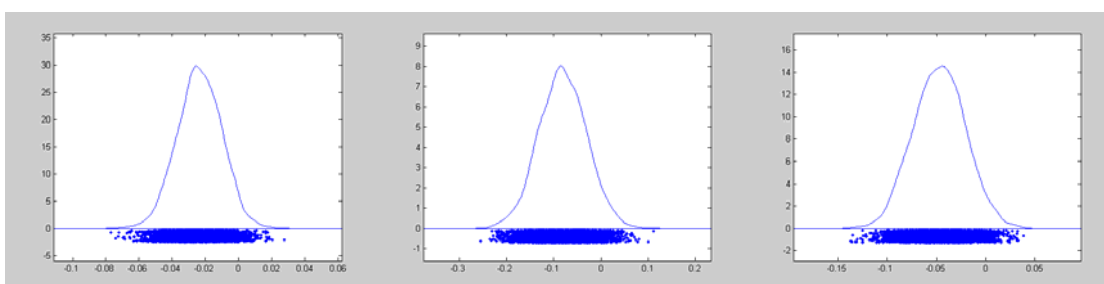
sp lag1	0.255148	ec term rj	-0.442406
sp lag2	0.147308	ec term sp	-0.495417
sp lag3	0.150339	constant	62.808212



(1) $\beta_{BA,t-1}$

(2) $\beta_{BA,t-2}$

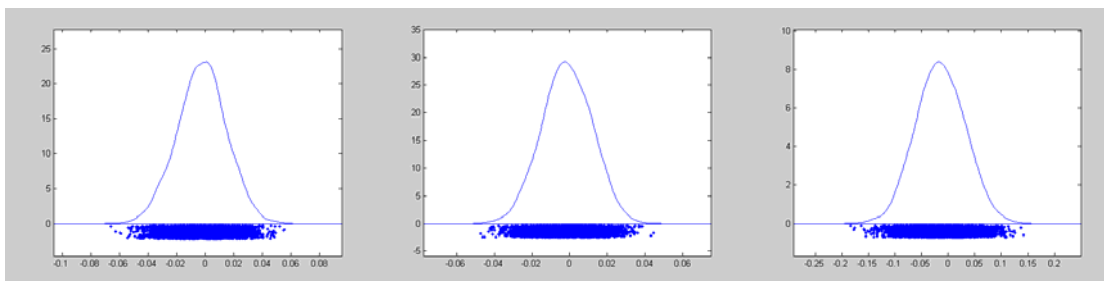
(3) $\beta_{BA,t-3}$



(4) $\beta_{BA,t-4}$

(5) $\beta_{MG,t-1}$

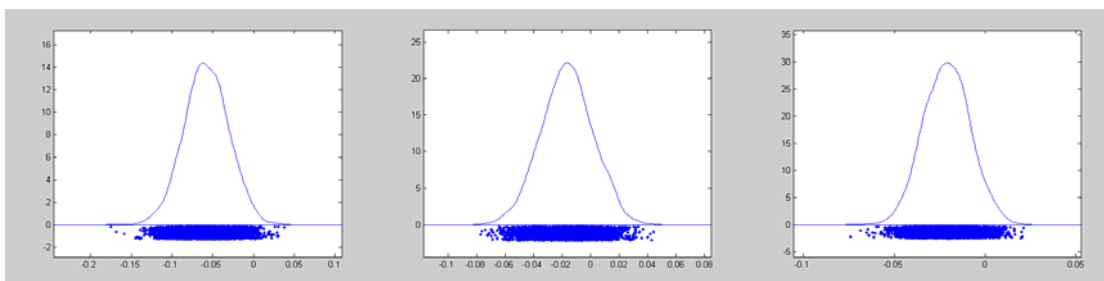
(6) $\beta_{MG,t-2}$



(7) $\beta_{MG,t-3}$

(8) $\beta_{MG,t-4}$

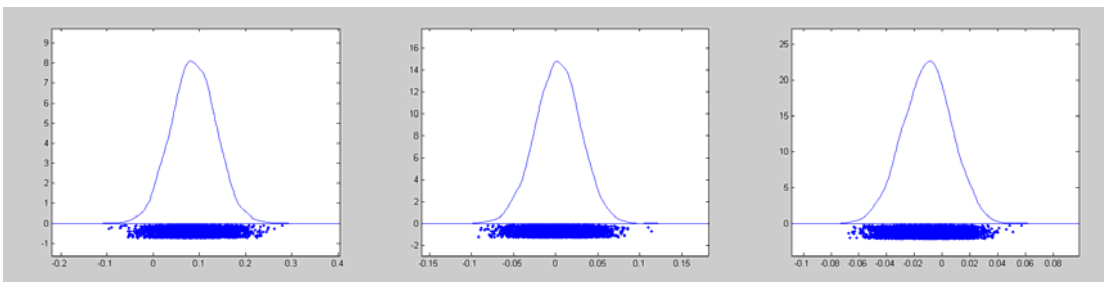
(9) $\beta_{ES,t-1}$



(10) $\beta_{ES,t-2}$

(11) $\beta_{ES,t-3}$

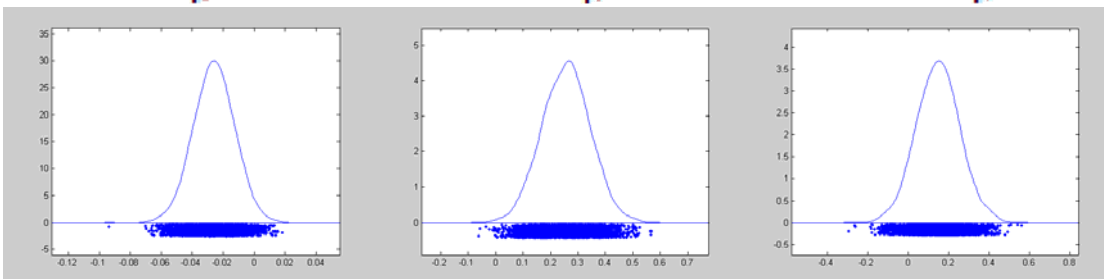
(12) $\beta_{ES,t-4}$



(13) $\beta_{R/J,t-1}$

(14) $\beta_{R/J,t-2}$

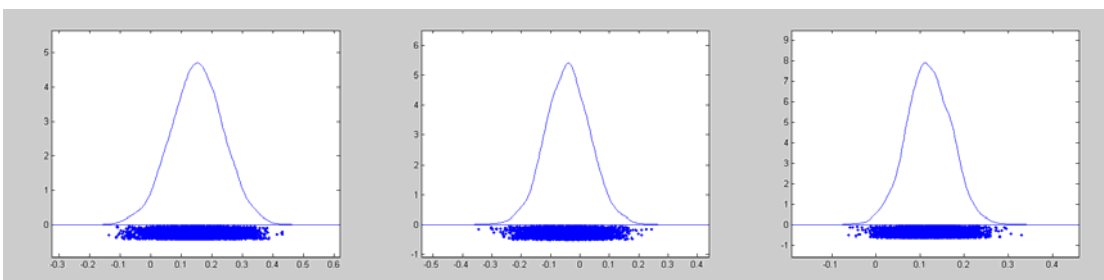
(15) $\beta_{R/J,t-3}$



(16) $\beta_{R/J,t-4}$

(17) $\beta_{SP,t-1}$

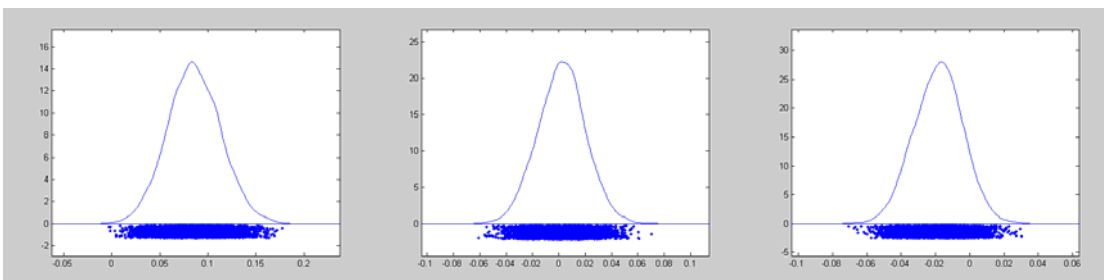
(18) $\beta_{SP,t-2}$



(19) $\beta_{SP,t-3}$

(20) $\beta_{SP,t-4}$

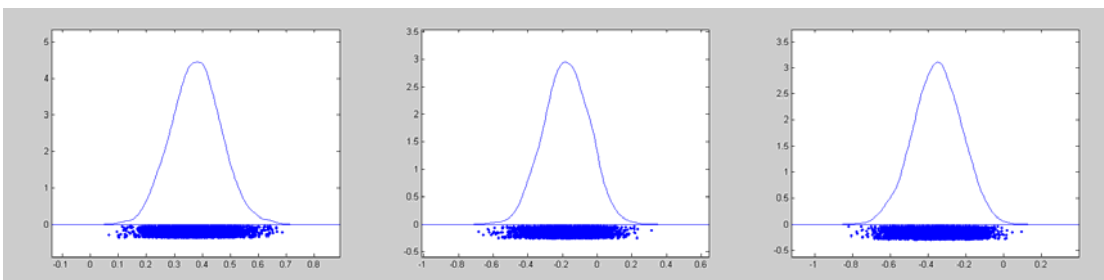
(21) $\beta_{PP,t-1}$



(22) $\beta_{PP,t-2}$

(23) $\beta_{PP,t-3}$

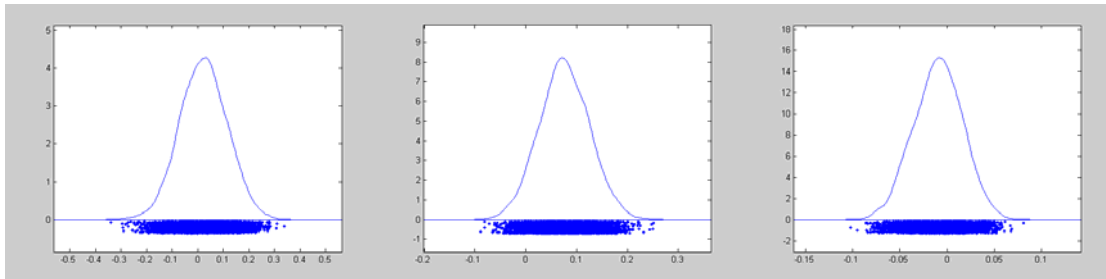
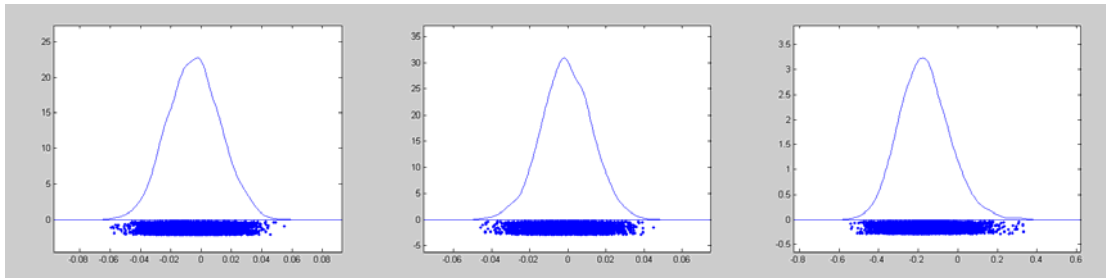
(24) $\beta_{PP,t-4}$



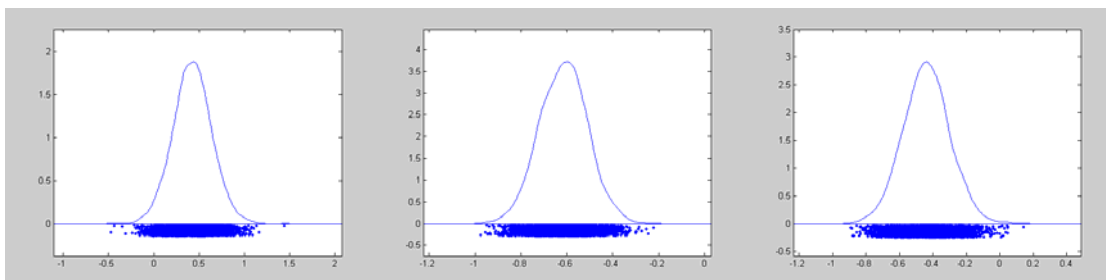
(25) $\beta_{SCC,t-1}$

(26) $\beta_{SCC,t-2}$

(27) $\beta_{SCC,t-3}$

(28) β_{SCt-1} (29) β_{RFt-1} (30) β_{RFt-2} (31) β_{RFt-2} (32) β_{RFt-1}

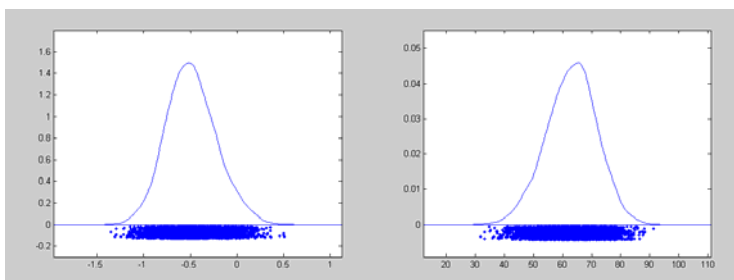
(33) ec term ba



(34) Ec term mg

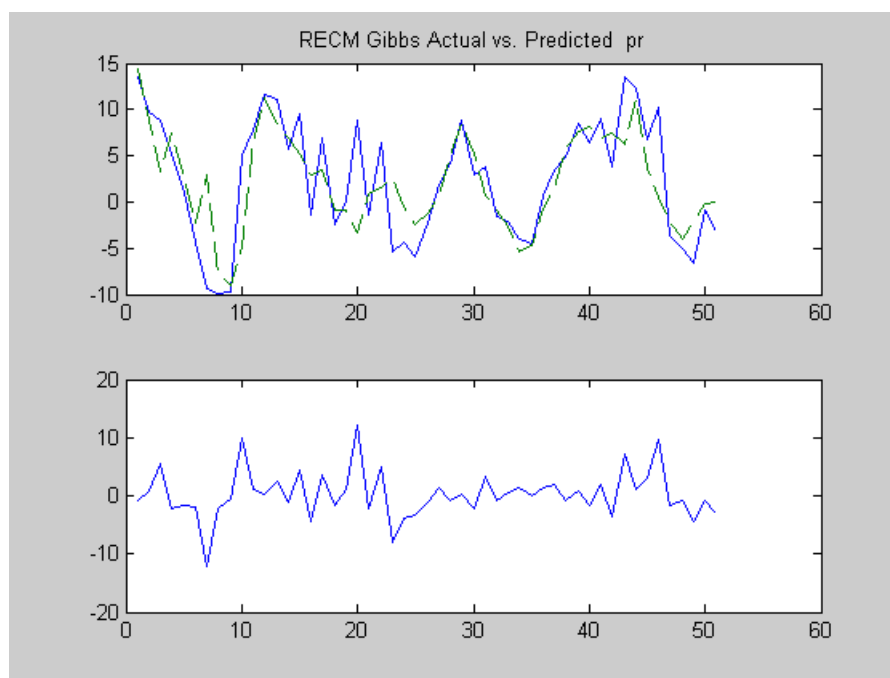
(35) ec term es

(36) ec term rj



(37) ec term sp

(38) C_{RR}



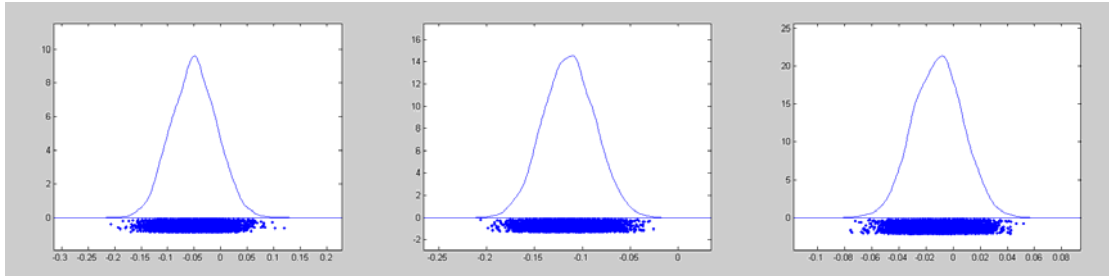
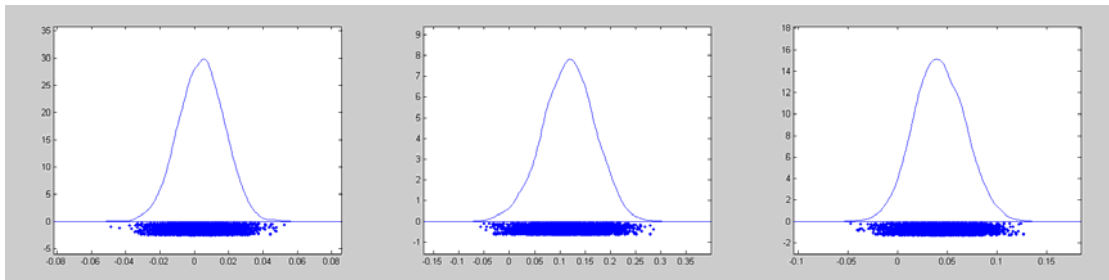
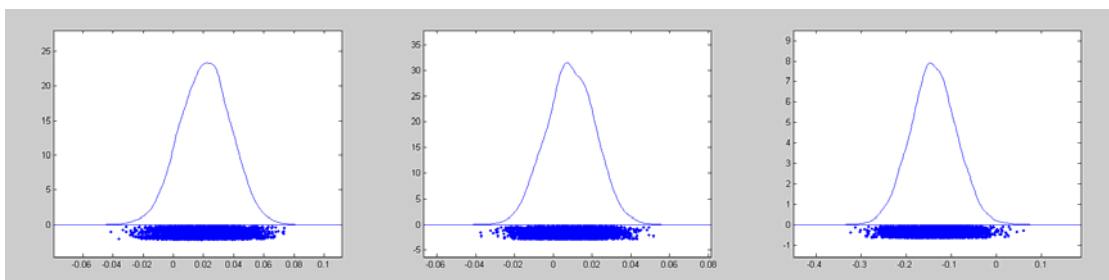
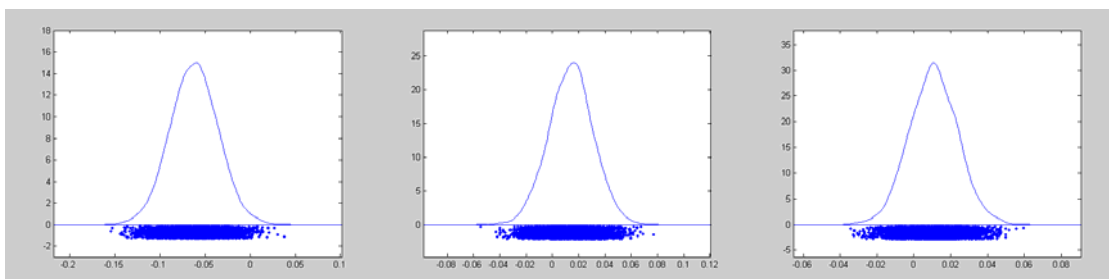
7. Resultados para a equação que tem variação da produção industrial no estado de Santa Catarina como variável dependente

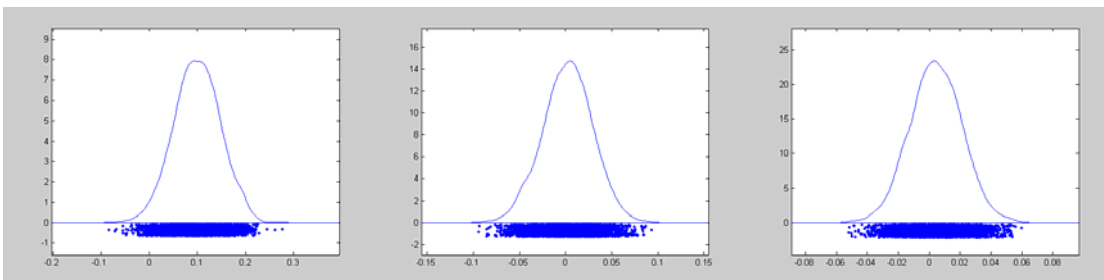
Dependent Variable sc

R-squared	0.7612
Rbar-squared	0.0817
sige	5.4287
Nobs, Nvars	63, 8
ndraws,nomit	5000, 100
r-value	4

Variável	Moda do Coef.	Variável	Moda do Coef.
ba lag1	-0.050915	sp lag4	-0.002132
ba lag2	-0.113544	pr lag1	0.349645
ba lag3	-0.011053	pr lag2	-0.071705
ba lag4	0.003478	pr lag3	-0.034125
mg lag1	0.119727	pr lag4	0.13623
mg lag2	0.042329	sc lag1	0.20881
mg lag3	0.021041	sc lag2	0.012751
mg lag4	0.008402	sc lag3	0.007769
es lag1	-0.141634	sc lag4	-0.004494
es lag2	-0.061355	rs lag1	0.248587
es lag3	0.013956	rs lag2	-0.203077
es lag4	0.01	rs lag3	-0.035594
rj lag1	0.09978	rs lag4	-0.111628
rj lag2	0.001043	ec term ba	-0.166282
rj lag3	0.004065	ec term mg	-0.163909

ry lag4	-0.016135	ec term es	0.212369
sp lag1	0.077957	ec term ry	-0.058967
sp lag2	0.021222	ec term sp	0.853938
sp lag3	0.007424	constant	-19.028144

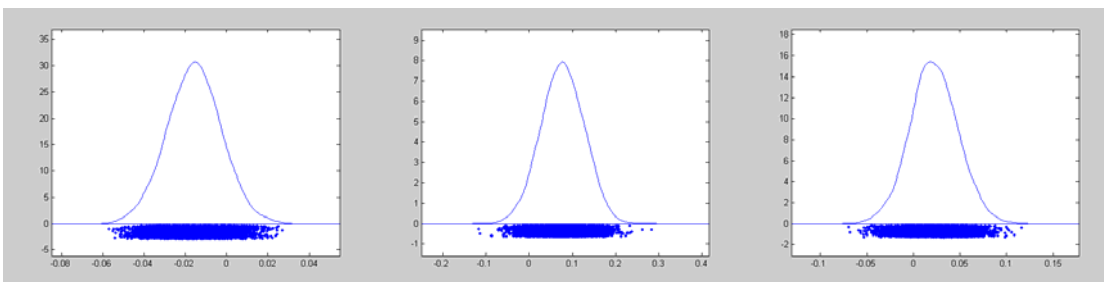
(1) $\beta_{BA,t-1}$ (2) $\beta_{BA,t-2}$ (3) $\beta_{BA,t-3}$ (4) $\beta_{BA,t-4}$ (5) $\beta_{MG,t-1}$ (6) $\beta_{MG,t-2}$ (7) $\beta_{MG,t-3}$ (8) $\beta_{MG,t-4}$ (9) $\beta_{ES,t-1}$ (10) $\beta_{ES,t-2}$ (11) $\beta_{ES,t-3}$ (12) $\beta_{ES,t-4}$



(13) $\beta_{RJ,t-1}$

(14) $\beta_{RJ,t-2}$

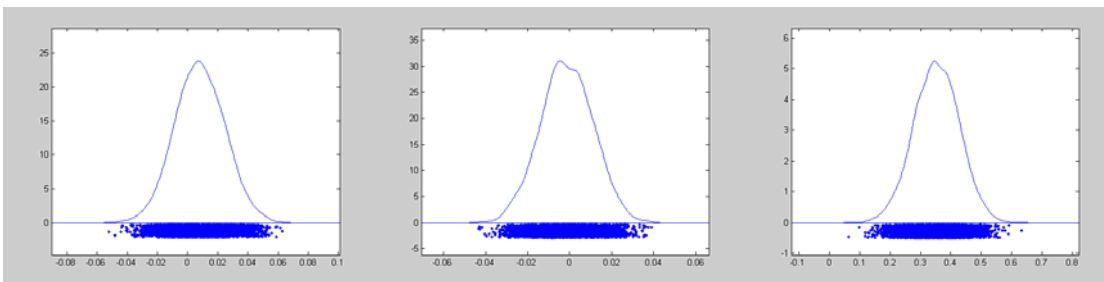
(15) $\beta_{RJ,t-3}$



(16) $\beta_{RJ,t-4}$

(17) $\beta_{SP,t-1}$

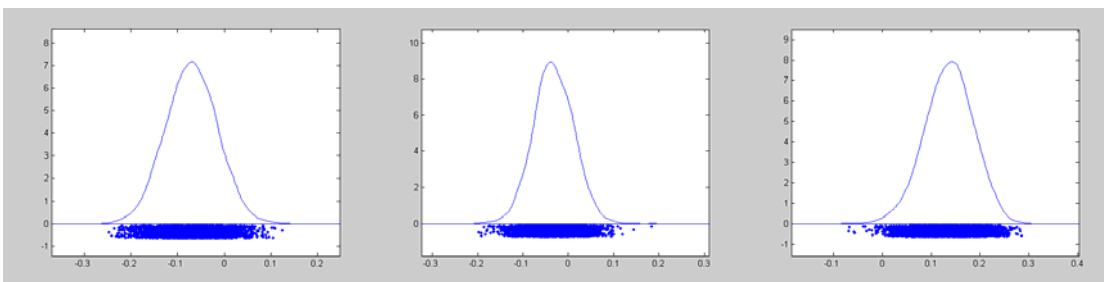
(18) $\beta_{SP,t-2}$



(19) $\beta_{SP,t-3}$

(20) $\beta_{SP,t-4}$

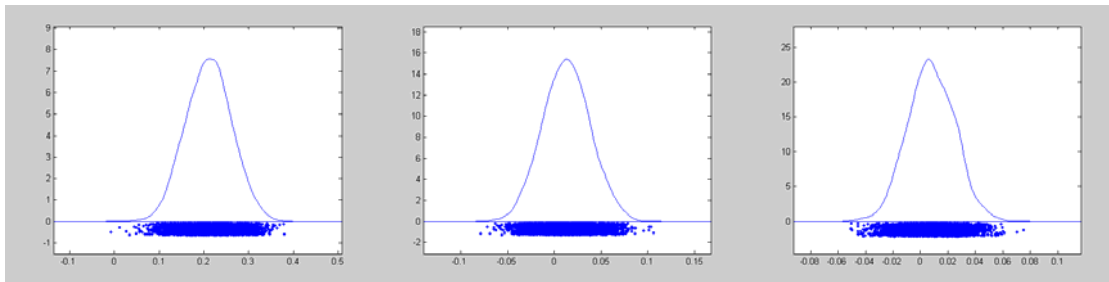
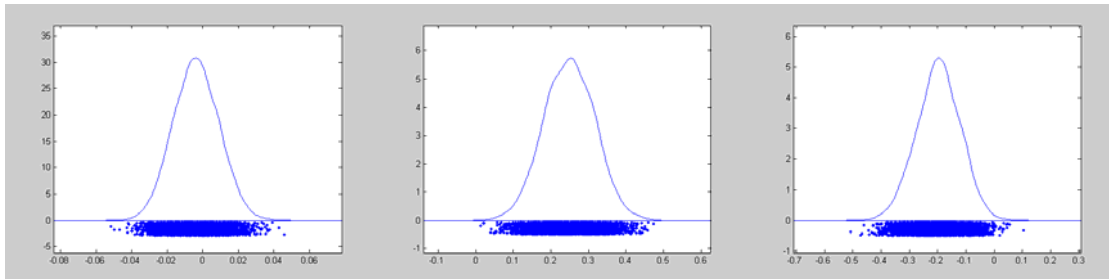
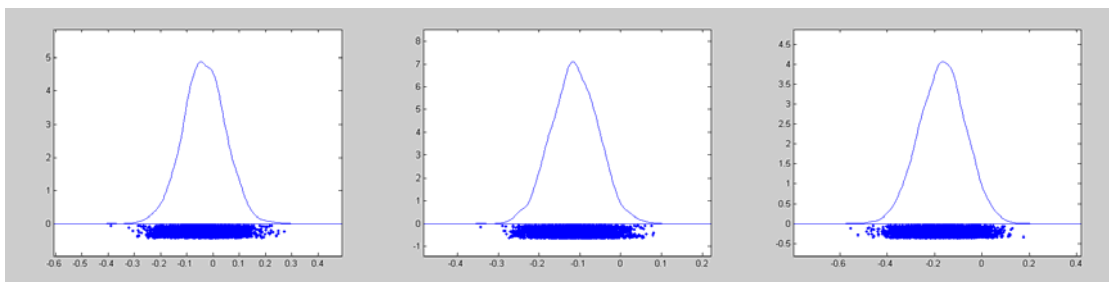
(21) $\beta_{PP,t-1}$



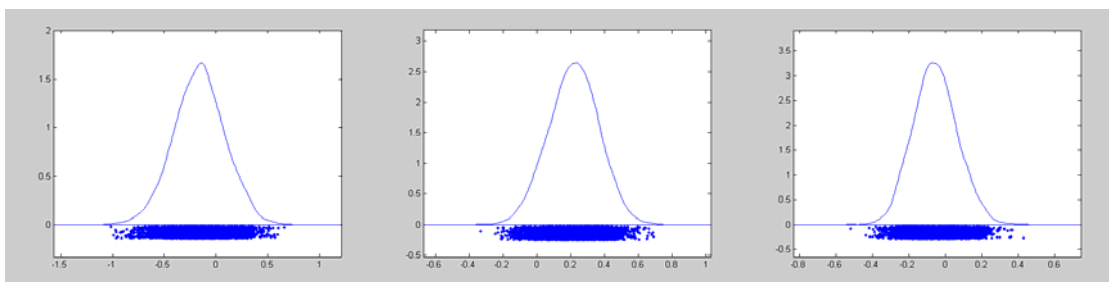
(22) $\beta_{PP,t-2}$

(23) $\beta_{PP,t-3}$

(24) $\beta_{PP,t-4}$

(25) $\beta_{SC,t-1}$ (26) $\beta_{SC,t-2}$ (27) $\beta_{SC,t-3}$ (28) $\beta_{SC,t-4}$ (29) $\beta_{RS,t-1}$ (30) $\beta_{RS,t-2}$ (31) $\beta_{RS,t-3}$ (32) $\beta_{RS,t-4}$

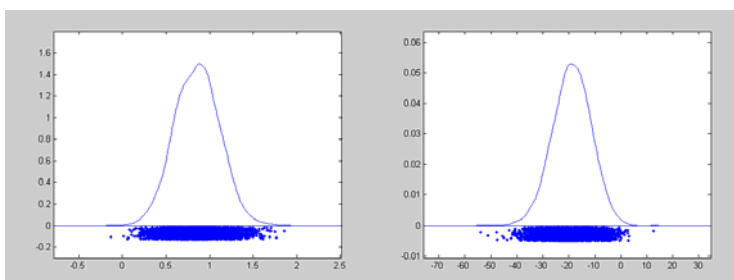
(33) ec term ba



(34) Ec term mg

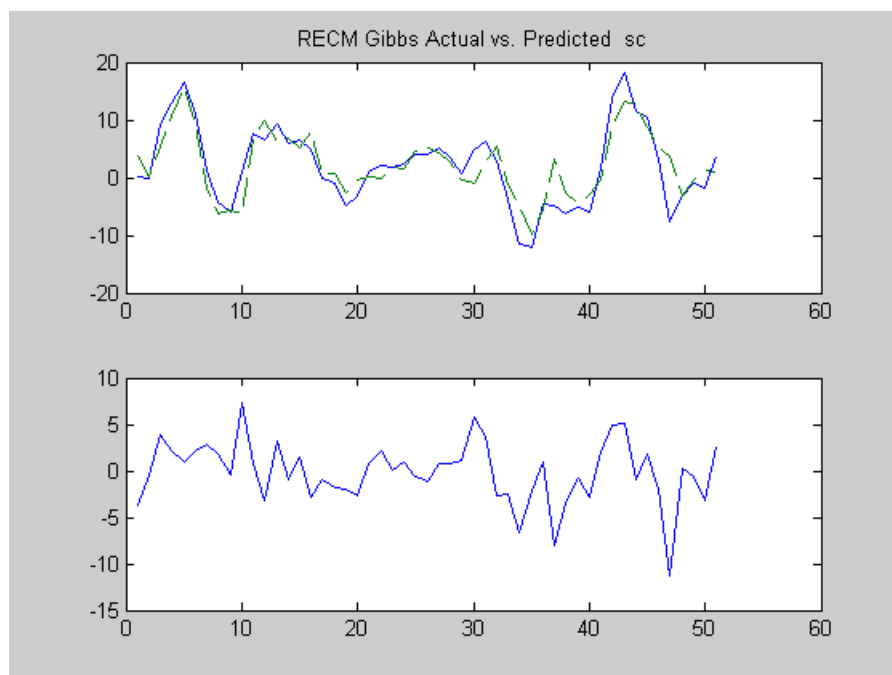
(35) ec term es

(36) ec term rj



(37) ec term sp

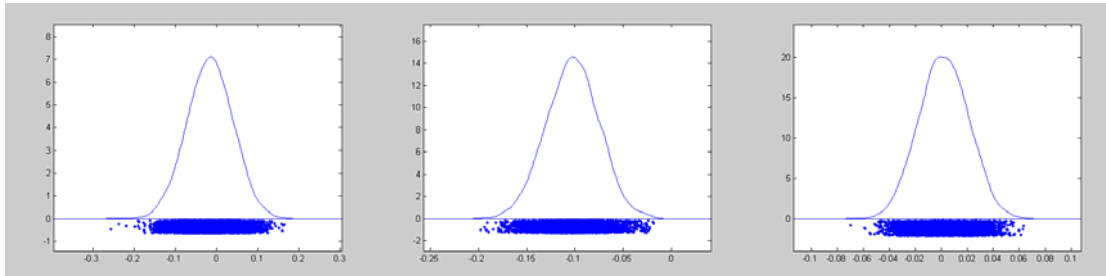
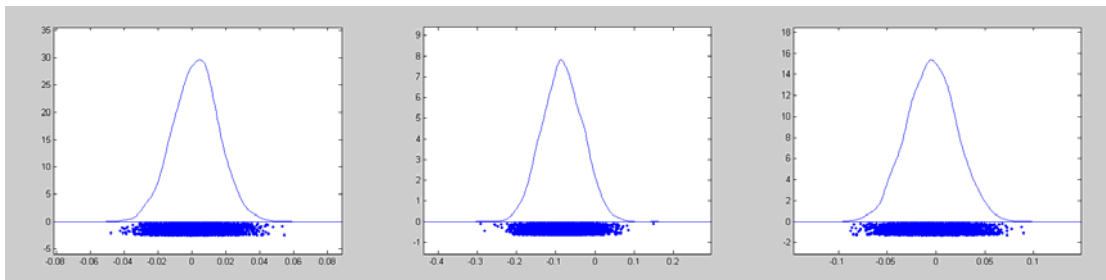
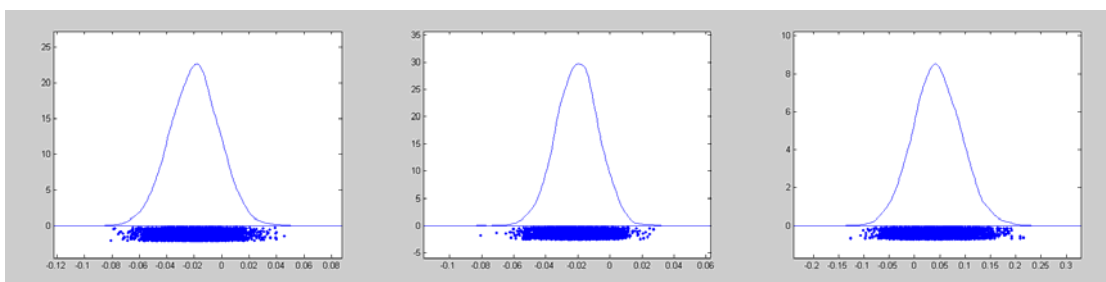
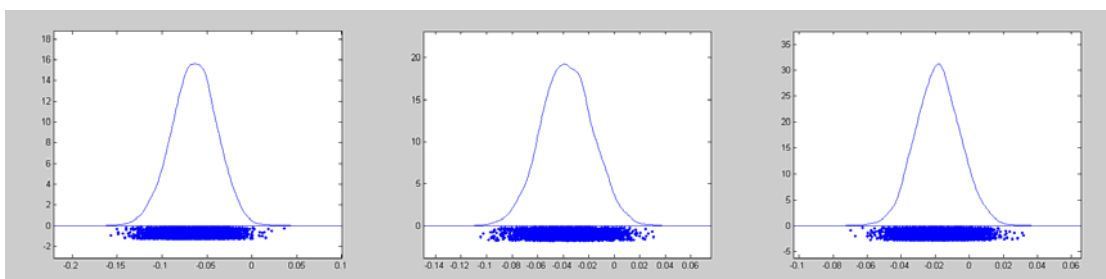
(38) C_{SC}

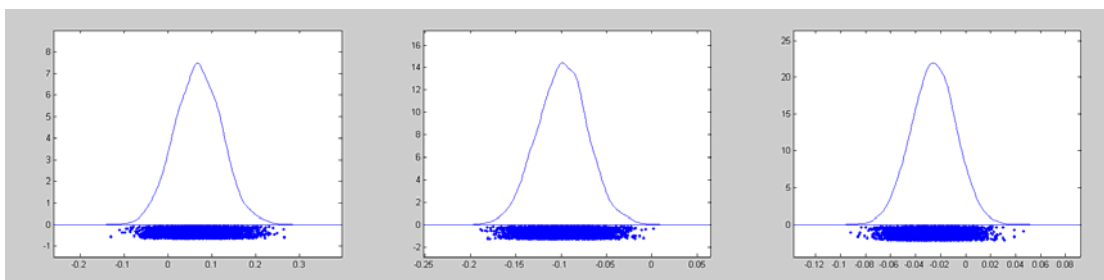
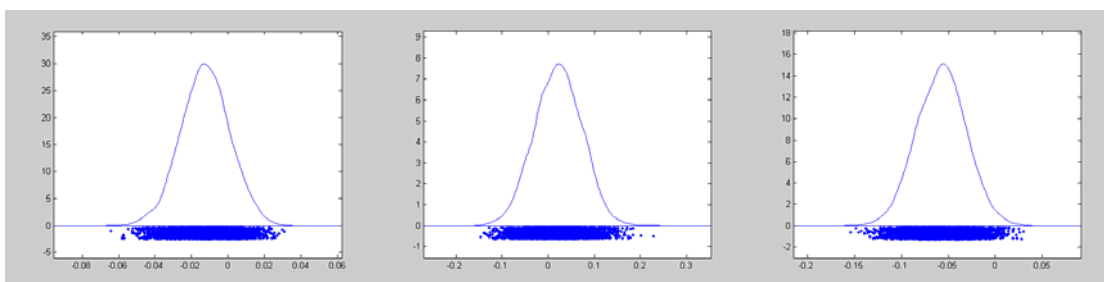
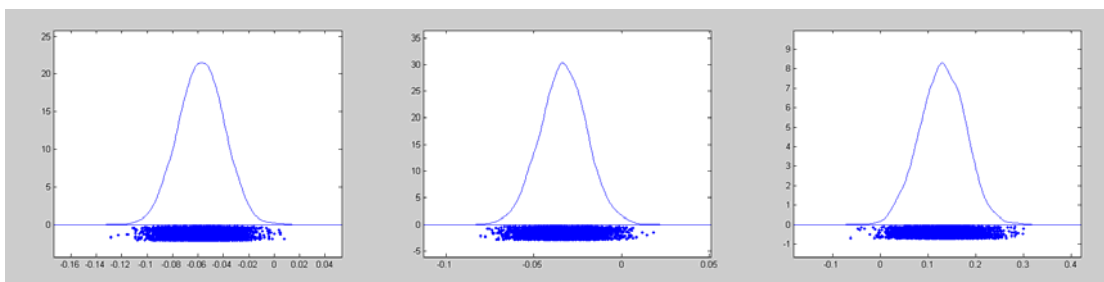
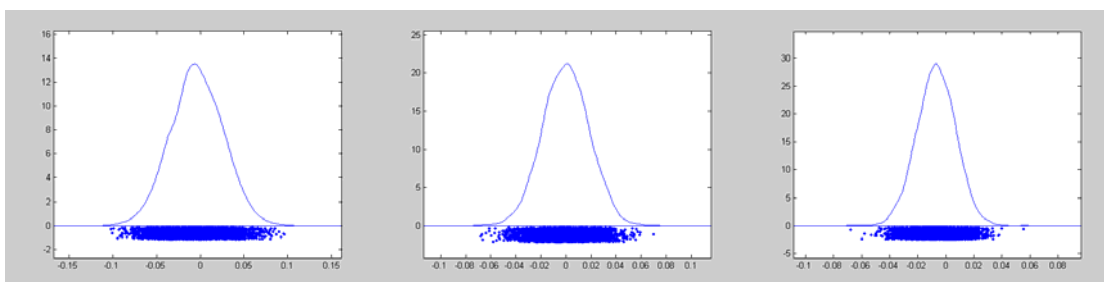


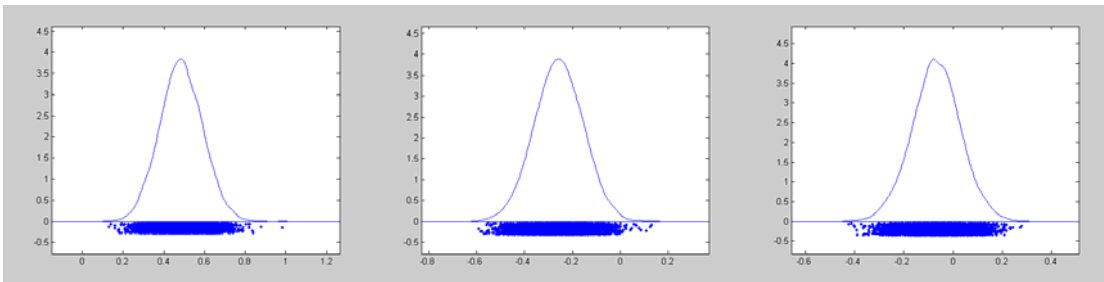
8. Resultados para a equação que tem variação da produção industrial no estado do Rio Grande do Sul como variável dependente

Dependent Variable rs			
R-squared	0.8215		
Rbar-squared	0.3135		
sige	8.0915		
Nobs, Nvars	63, 8		
ndraws,nomit	5000, 100		
r-value	4		
Variável	Moda do Coef.	Variável	Moda do Coef.
ba lag1	-0.019835	sp lag4	-0.033834
ba lag2	-0.102106	pr lag1	0.127956
ba lag3	0.002251	pr lag2	-0.004535
ba lag4	0.001559	pr lag3	-0.000306
mg lag1	-0.085208	pr lag4	-0.007216
mg lag2	-0.005252	sc lag1	0.481168
mg lag3	-0.02017	sc lag2	-0.260047
mg lag4	-0.020833	sc lag3	-0.071536
es lag1	0.043327	sc lag4	0.013158
es lag2	-0.064987	rs lag1	0.44732
es lag3	-0.038076	rs lag2	0.008315
es lag4	-0.01865	rs lag3	-0.020481
rj lag1	0.065276	rs lag4	-0.023558
rj lag2	-0.099138	ec term ba	-0.230313

rj lag3	-0.025701	ec term mg	0.103398
rj lag4	-0.012852	ec term es	-0.439008
sp lag1	0.019542	ec term rj	0.370209
sp lag2	-0.057309	ec term sp	0.564771
sp lag3	-0.058087	constant	0.62958

(1) $\beta_{MG,t-1}$ (2) $\beta_{MG,t-2}$ (3) $\beta_{MG,t-3}$ (4) $\beta_{ES,t-4}$ (5) $\beta_{MG,t-1}$ (6) $\beta_{MG,t-2}$ (7) $\beta_{MG,t-3}$ (8) $\beta_{MG,t-4}$ (9) $\beta_{ES,t-1}$ (10) $\beta_{ES,t-2}$ (11) $\beta_{ES,t-3}$ (12) $\beta_{ES,t-4}$

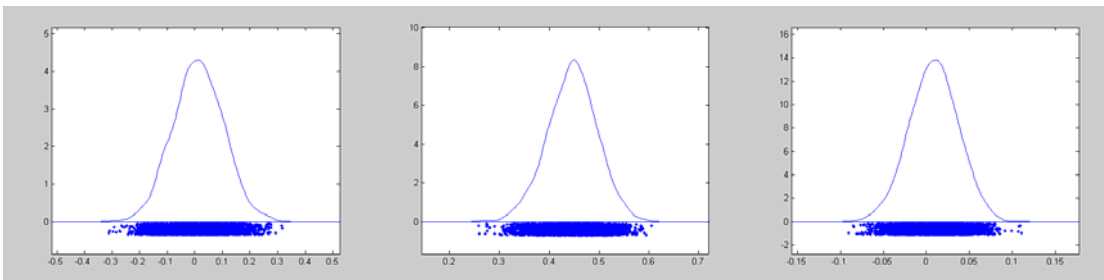
(13) $\beta_{RJ,t-1}$ (14) $\beta_{RJ,t-2}$ (15) $\beta_{RJ,t-3}$ (16) $\beta_{RJ,t-4}$ (17) $\beta_{SP,t-1}$ (18) $\beta_{SP,t-2}$ (19) $\beta_{SP,t-3}$ (20) $\beta_{SP,t-4}$ (21) $\beta_{PP,t-1}$ (22) $\beta_{PP,t-2}$ (23) $\beta_{PP,t-3}$ (24) $\beta_{PP,t-4}$



(25) $\beta_{SC,t-1}$

(26) $\beta_{SC,t-2}$

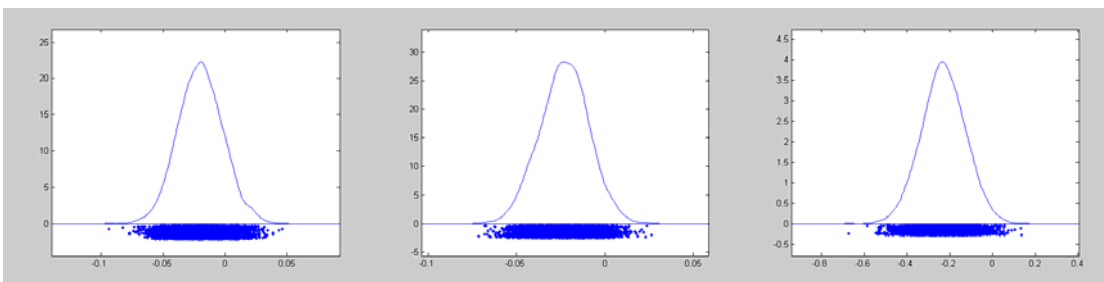
(27) $\beta_{SC,t-3}$



(28) $\beta_{SC,t-4}$

(29) $\beta_{RS,t-1}$

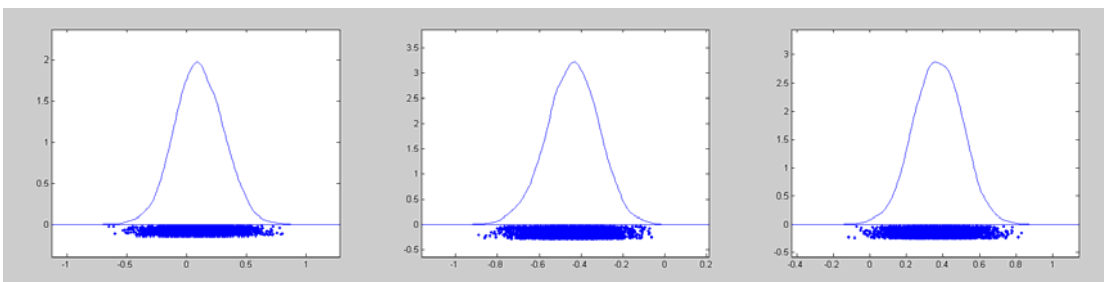
(30) $\beta_{RS,t-2}$



(31) $\beta_{RS,t-3}$

(32) $\beta_{RS,t-4}$

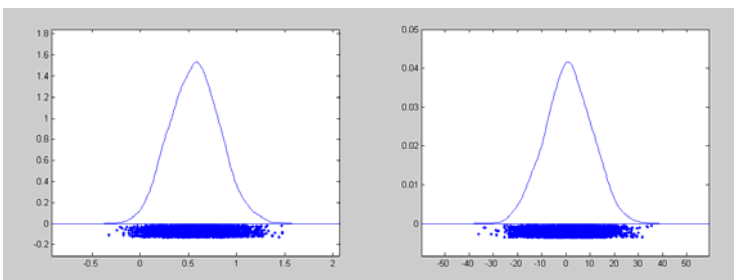
(33) ec term ba



(34) Ec term mg

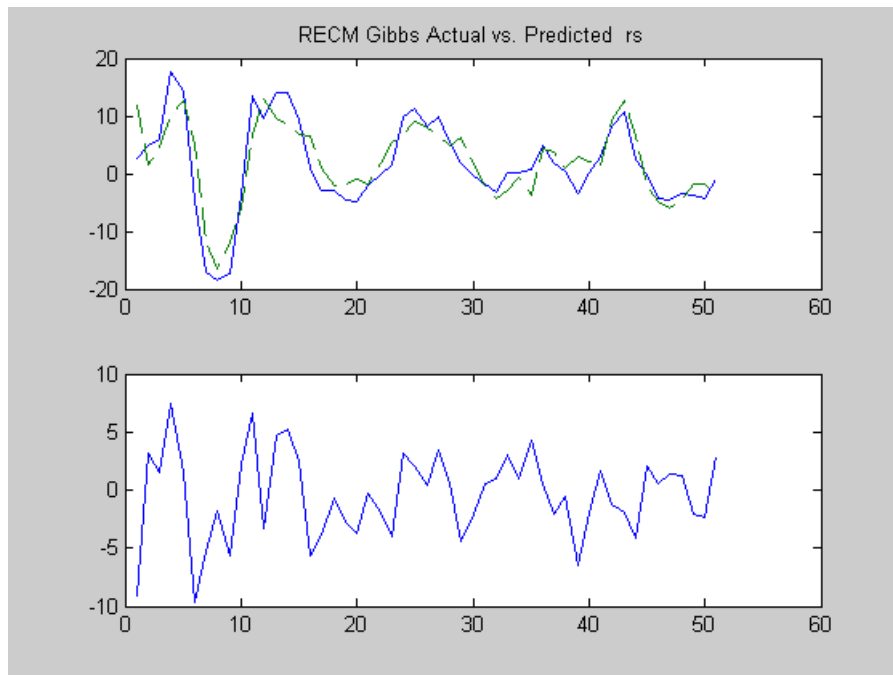
(35) ec term es

(36) ec term rj



(37) ec term sp

(38) C_{RS}



Apêndice C Apresentação dos resultados empíricos do modelo VEC Bayesiano utilizando informação *a priori* de Minnesota modificada estimado por Gibbs Sampling

PRIOR hyperparameters

tightness = 0.10

decay = 0.10

Weights matrix

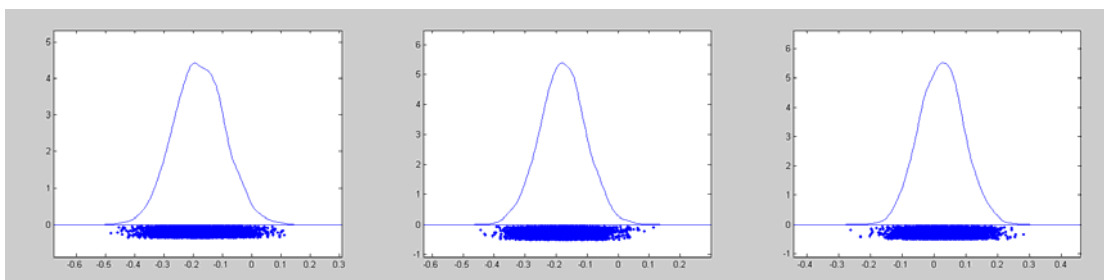
Variable	ba	mg	es	rj	sp	pr	sc	rs
ba	1	1	1	0.1	0.1	0.1	0.1	0.1
mg	1	1	1	1	1	0.1	0.1	0.1
es	1	1	1	1	0.1	0.1	0.1	0.1
rj	0.1	1	1	1	1	0.1	0.1	0.1
sp	0.1	1	0.1	1	1	1	0.1	0.1
pr	0.1	0.1	0.1	0.1	1	1	1	0.1
sc	0.1	0.1	0.1	0.1	0.1	1	1	1
rs	0.1	0.1	0.1	0.1	0.1	0.1	1	1

1. Resultados para a equação que tem variação da produção industrial no estado da Bahia como variável dependente

Dependent Variable ba	
R-squared	0.5216
Rbar-squared	-0.3633
sige	4.1394
Nobs, Nvars	63, 8
ndraws,nomit	5000, 100
r-value	4

Variável	Moda do Coef.	Variável	Moda do Coef.
ba lag1	-0.18085	sp lag4	-0.020015
ba lag2	-0.180046	pr lag1	-0.026793
ba lag3	0.021967	pr lag2	-0.002814
ba lag4	-0.065973	pr lag3	-0.028996
mg lag1	0.137671	pr lag4	0.017136
mg lag2	-0.151049	sc lag1	0.023281
mg lag3	0.08697	sc lag2	-0.024868
mg lag4	-0.129905	sc lag3	-0.029127
es lag1	-0.218416	sc lag4	-0.009274
es lag2	-0.029894	rs lag1	0.0039
es lag3	-0.235742	rs lag2	-0.004146

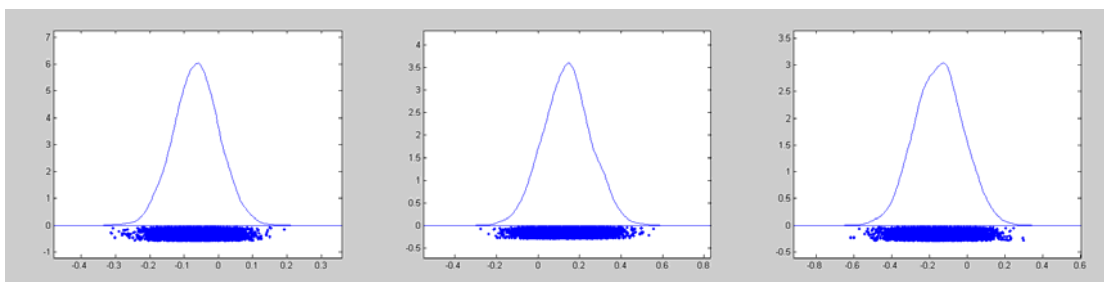
es lag4	-0.25113	rs lag3	-0.010267
rj lag1	-0.029227	rs lag4	-0.004967
rj lag2	0.026785	ec term ba	-0.160416
rj lag3	0.015252	ec term mg	-0.413503
rj lag4	0.002208	ec term es	-0.116217
sp lag1	-0.005675	ec term rj	-0.16637
sp lag2	0.015101	ec term sp	-0.556375
sp lag3	-0.012544	constant	17.298635



(1) $\beta_{BA,t-1}$

(2) $\beta_{BA,t-2}$

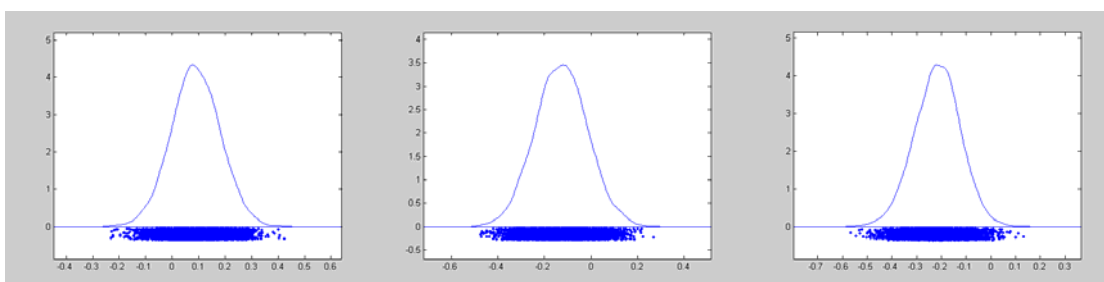
(3) $\beta_{BA,t-3}$



(4) $\beta_{BA,t-4}$

(5) $\beta_{MG,t-1}$

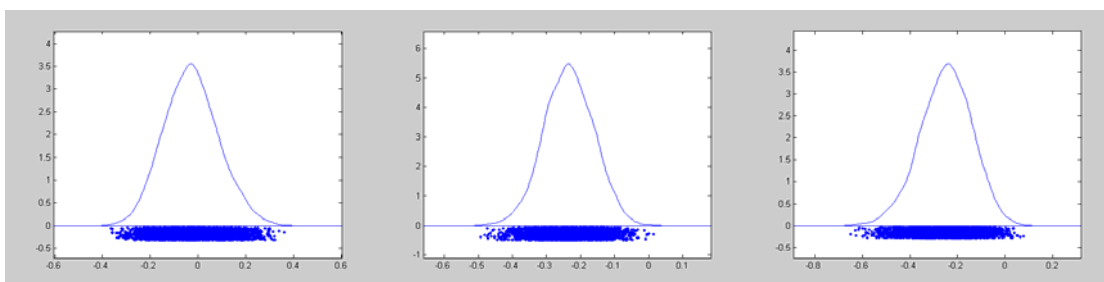
(6) $\beta_{MG,t-2}$



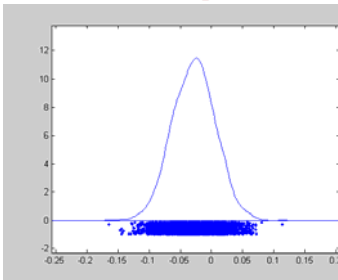
(7) $\beta_{MG,t-3}$

(8) $\beta_{MG,t-4}$

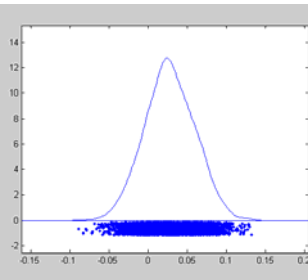
(9) $\beta_{ES,t-1}$



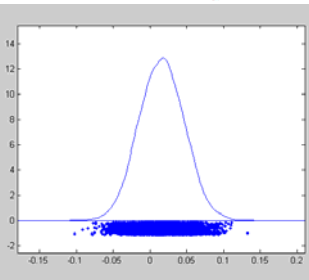
(10) $\beta_{ES,t-2}$



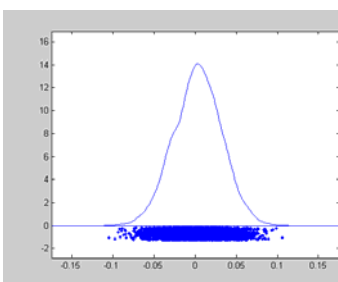
(11) $\beta_{ES,t-3}$



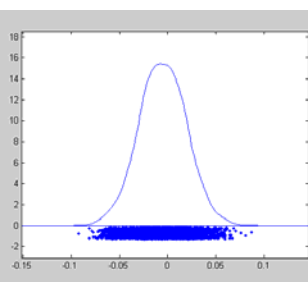
(12) $\beta_{ES,t-4}$



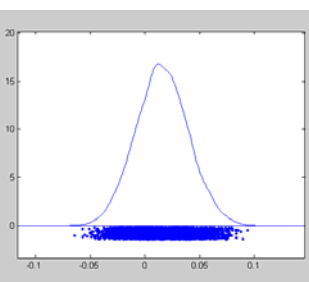
(13) $\beta_{RJ,t-1}$



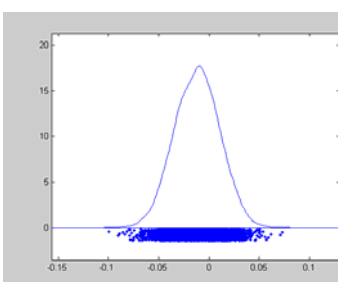
(14) $\beta_{RJ,t-2}$



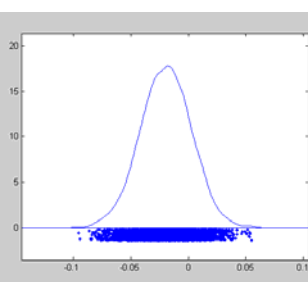
(15) $\beta_{RJ,t-3}$



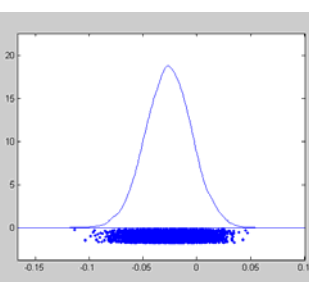
(16) $\beta_{KJ,t-4}$



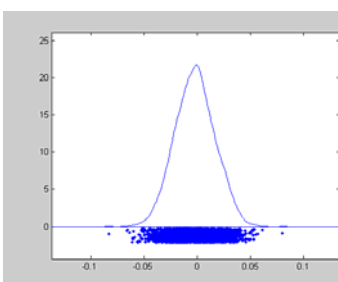
(17) $\beta_{SP,t-1}$



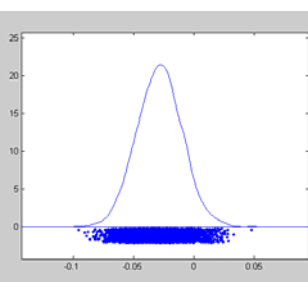
(18) $\beta_{SP,t-2}$



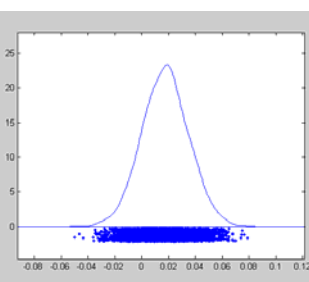
(19) $\beta_{SP,t-3}$



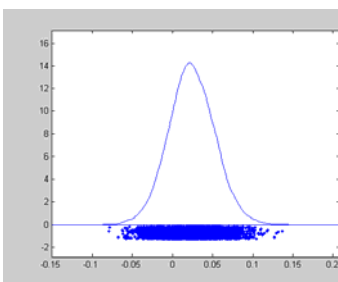
(20) $\beta_{SP,t-4}$



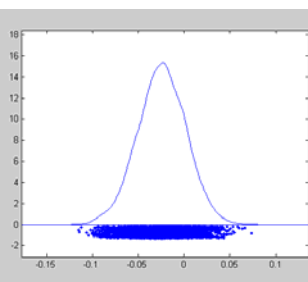
(21) $\beta_{PR,t-1}$



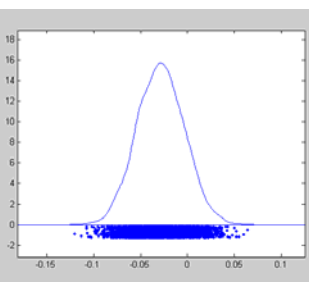
(22) $\beta_{PR,t-2}$



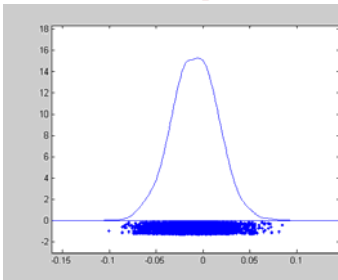
(23) $\beta_{PR,t-3}$



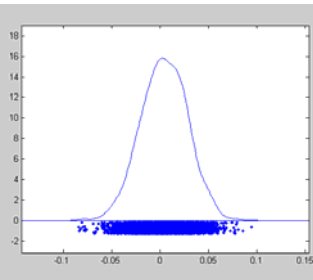
(24) $\beta_{PR,t-4}$



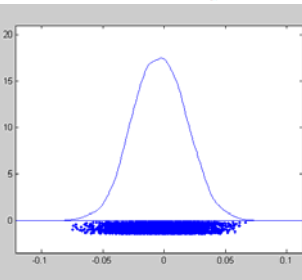
(25) $\beta_{SC,t-1}$



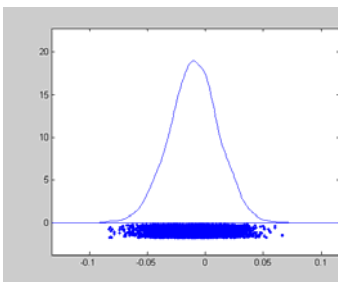
(26) $\beta_{SC,t-2}$



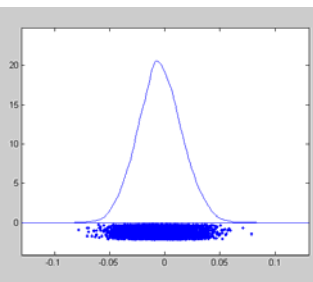
(27) $\beta_{SC,t-3}$



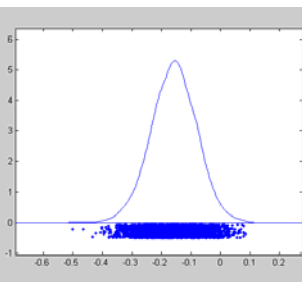
(28) $\beta_{SC,t-4}$



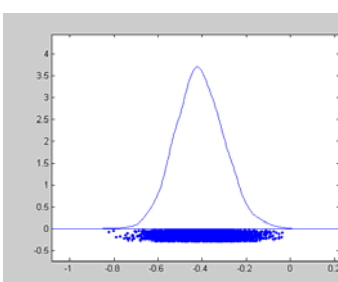
(29) $\beta_{RS,t-1}$



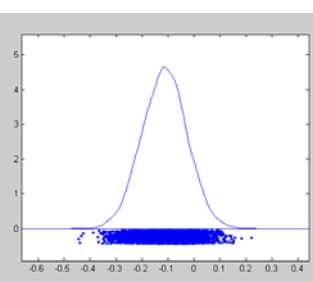
(30) $\beta_{RS,t-2}$



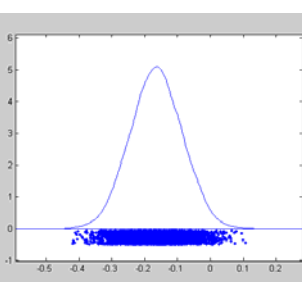
(31) $\beta_{RS,t-3}$



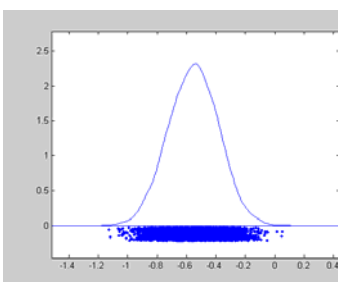
(32) $\beta_{RS,t-4}$



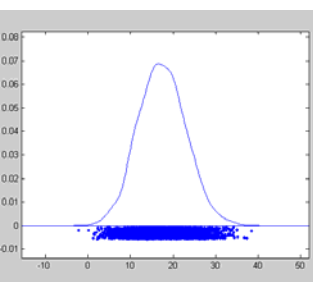
(33) ec term ba



(34) Ec term mg



(35) ec term es



(36) ec term rj

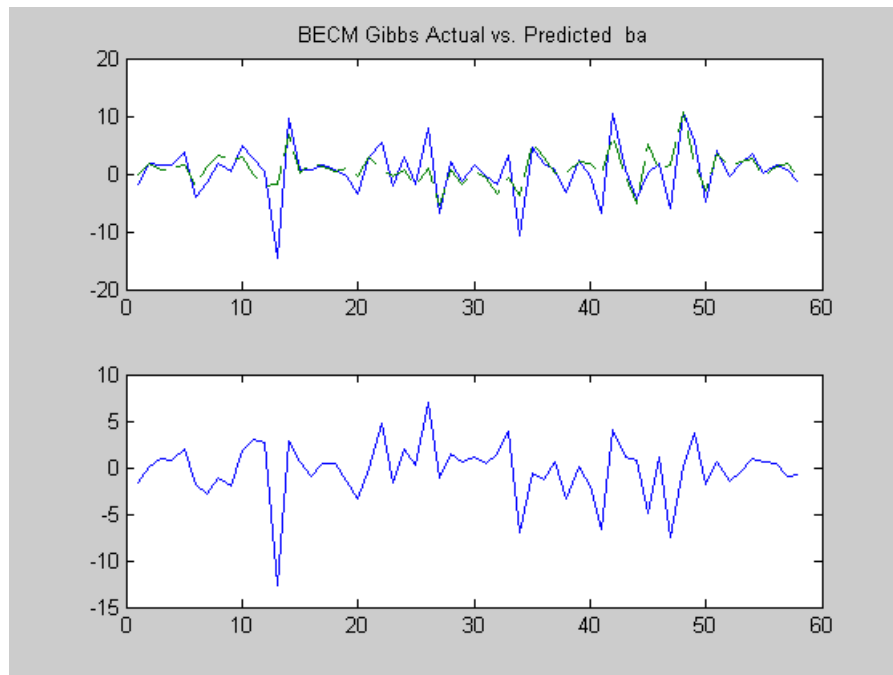


(37) ec term sp



(38) C_{BH}

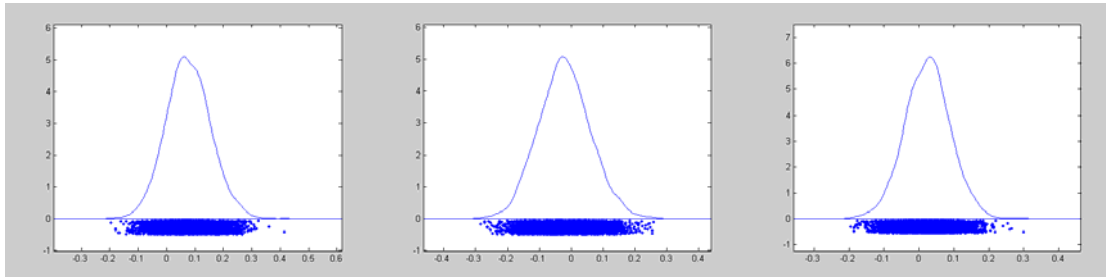
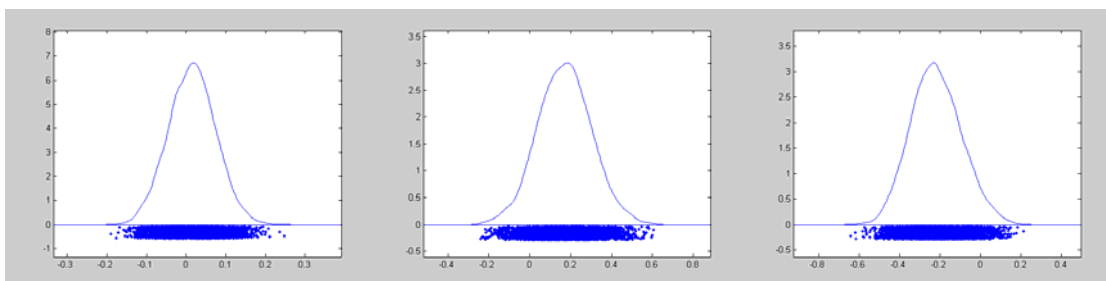
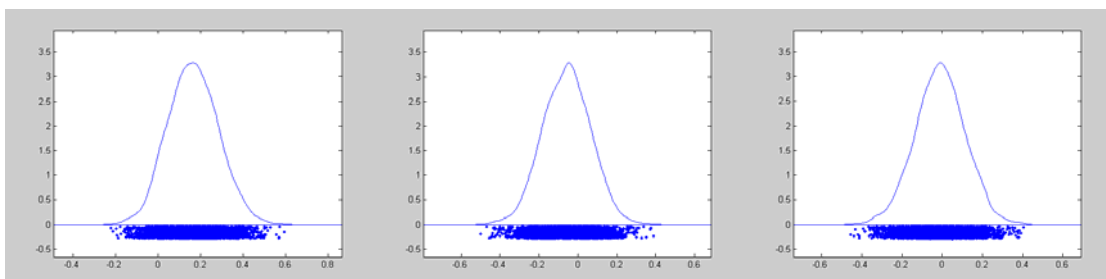
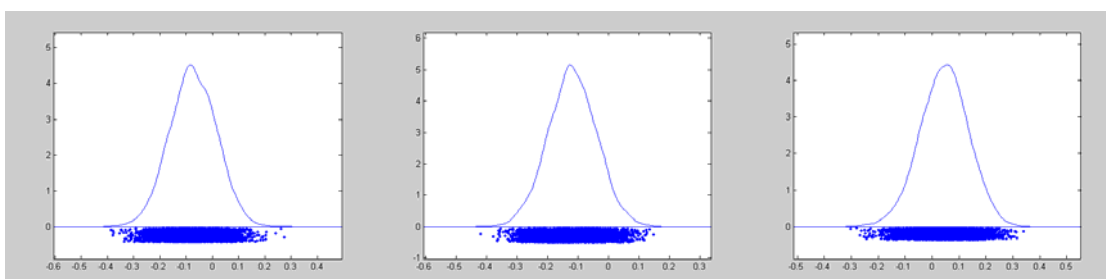


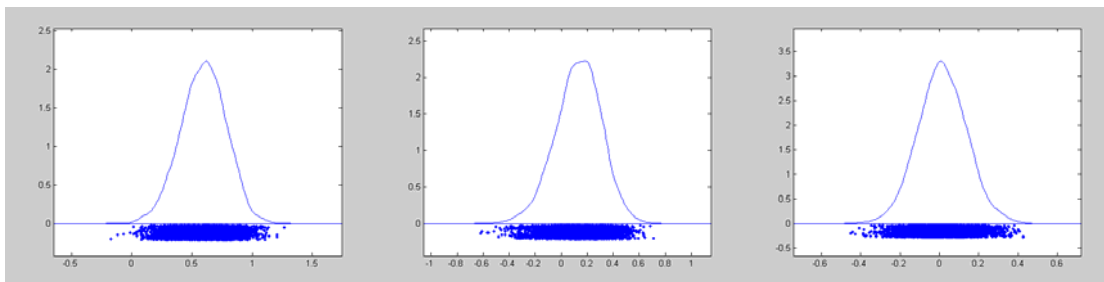
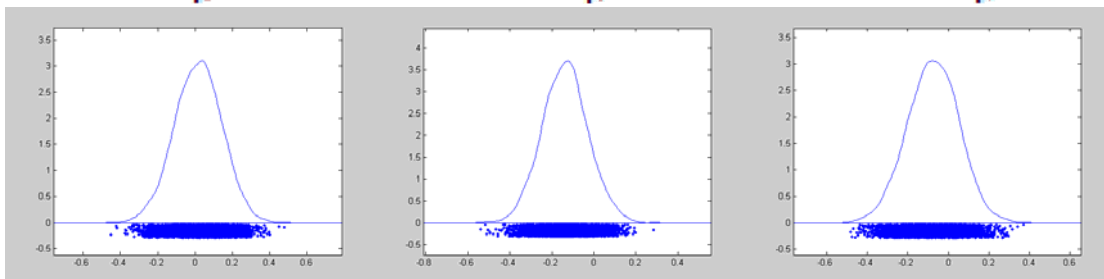
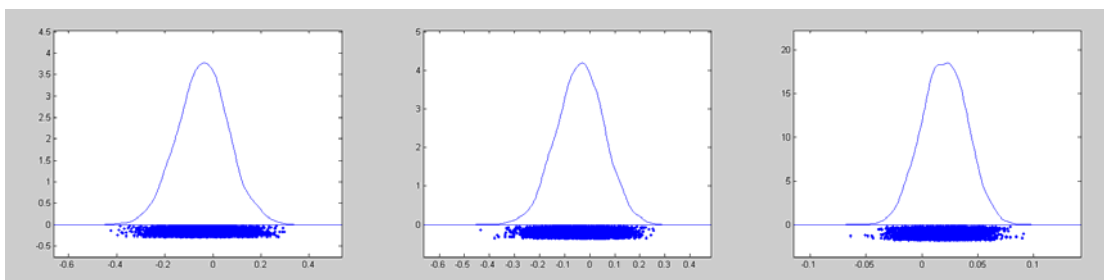
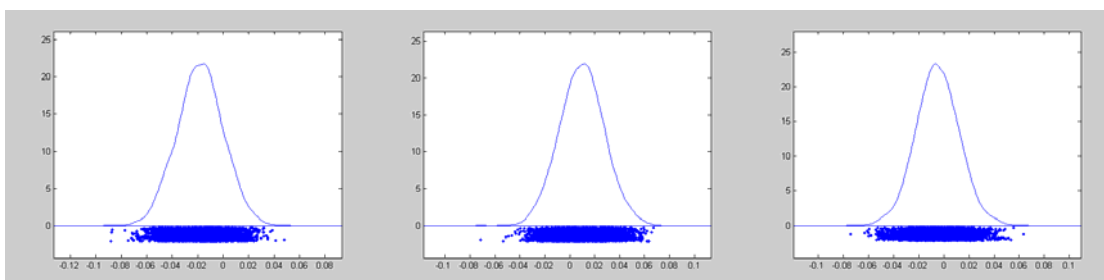
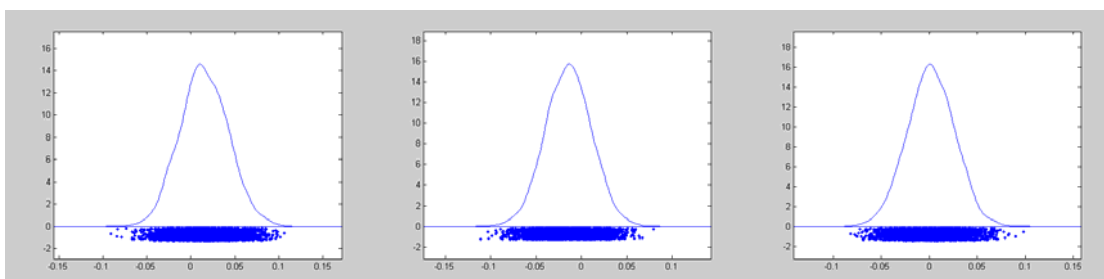


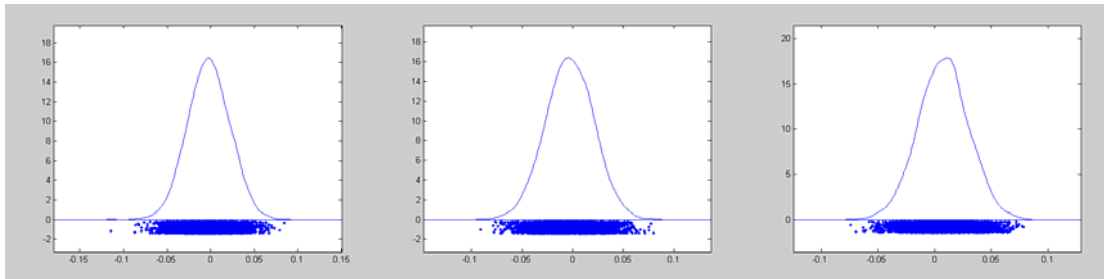
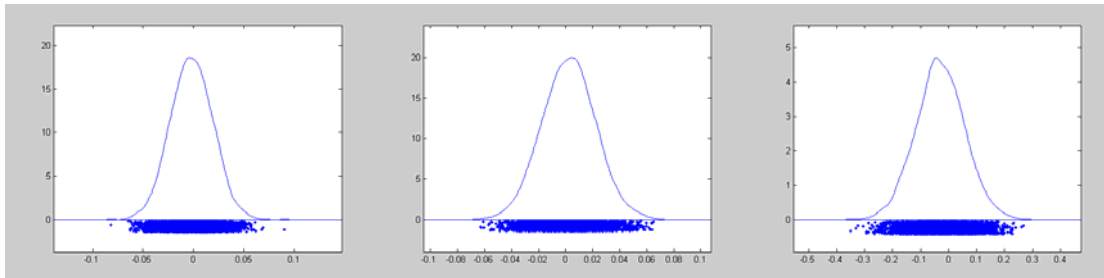
2. Resultados para a equação que tem variação da produção industrial no estado de Minas Gerais como variável dependente

Dependent Variable mg			
R-squared	0.479		
Rbar-squared	-0.4848		
sige	1.8386		
Nobs, Nvars	63, 8		
ndraws,nomit	5000, 100		
r-value	4		
Variável	Moda do Coef.	Variável	Moda do Coef.
ba lag1	0.051209	sp lag4	0.014819
ba lag2	-0.073361	pr lag1	0.117724
ba lag3	-0.004204	pr lag2	-0.058192
ba lag4	0.033776	pr lag3	0.037861
mg lag1	0.283535	pr lag4	-0.013657
mg lag2	-0.237648	sc lag1	0.100931
mg lag3	0.201865	sc lag2	-0.168671
mg lag4	-0.107405	sc lag3	0.035878
es lag1	-0.024039	sc lag4	-0.081079
es lag2	-0.026129	rs lag1	-0.072849
es lag3	-0.080126	rs lag2	0.055285
es lag4	0.027552	rs lag3	-0.022677
rj lag1	0.356387	rs lag4	-0.02712
rj lag2	-0.067909	ec term ba	0.018815
rj lag3	-0.048303	ec term mg	-0.142289

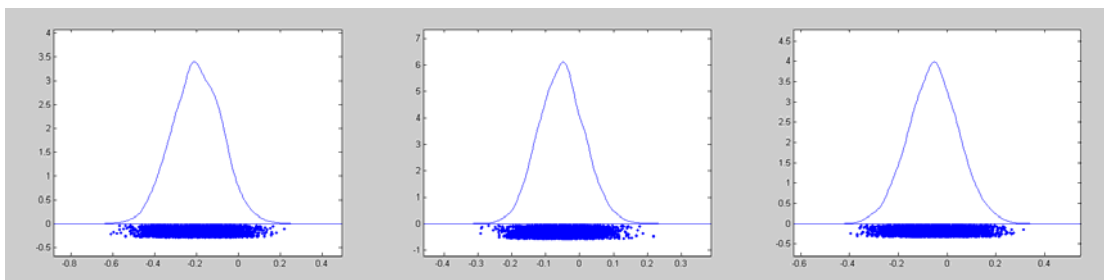
sj lag4	-0.038802	ec term es	-0.029894
sp lag1	-0.120206	ec term sj	0.006258
sp lag2	-0.025076	ec term sp	0.427333
sp lag3	-0.051765	constant	-6.04196

(1) $\beta_{BA,t-1}$ (2) $\beta_{BA,t-2}$ (3) $\beta_{BA,t-3}$ (4) $\beta_{BA,t-4}$ (5) $\beta_{MG,t-1}$ (6) $\beta_{MG,t-2}$ (7) $\beta_{MG,t-3}$ (8) $\beta_{MG,t-4}$ (9) $\beta_{ES,t-1}$ (10) $\beta_{ES,t-2}$ (11) $\beta_{ES,t-3}$ (12) $\beta_{ES,t-4}$

(13) $\beta_{R|t-1}$ (14) $\beta_{R|t-2}$ (15) $\beta_{R|t-3}$ (16) $\beta_{S|t-1}$ (17) $\beta_{SP,t-1}$ (18) $\beta_{SP,t-2}$ (19) $\beta_{SP,t-3}$ (20) $\beta_{SP,t-4}$ (21) $\beta_{PR,t-1}$ (22) $\beta_{PK,t-1}$ (23) $\beta_{PK,t-2}$ (24) $\beta_{PK,t-3}$ (25) $\beta_{SC,t-1}$ (26) $\beta_{SC,t-2}$ (27) $\beta_{SC,t-3}$

(28) $\beta_{SC,t-1}$ (29) $\beta_{RF,t-1}$ (30) $\beta_{RF,t-2}$ (31) $\beta_{RF,t-2}$ (32) $\beta_{RF,t-1}$

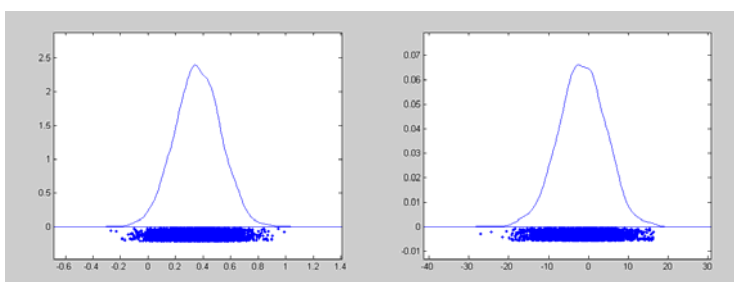
(33) ec term ba



(34) Ec term mg

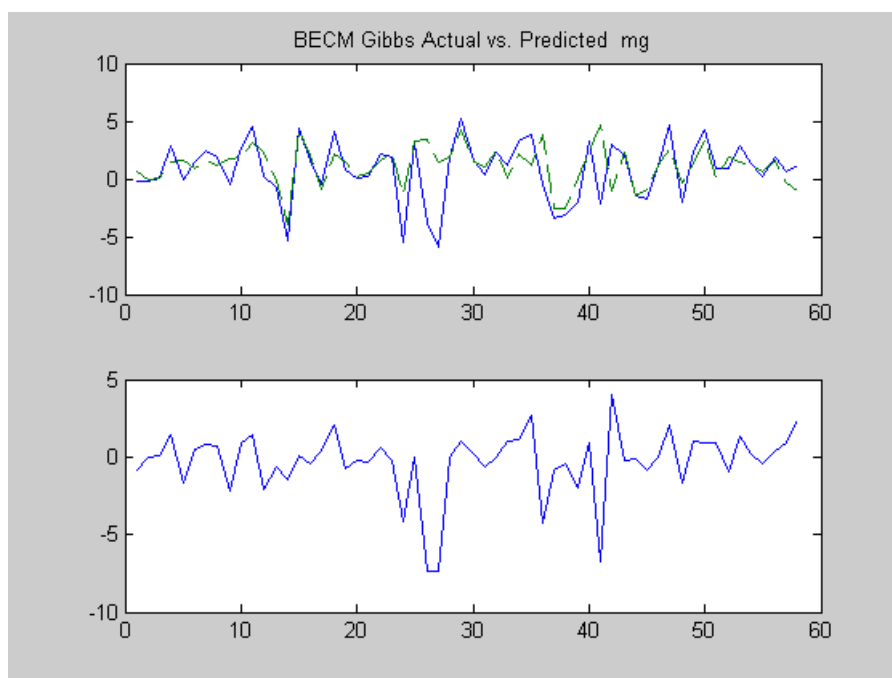
(35) ec term es

(36) ec term rj



(37) ec term sp

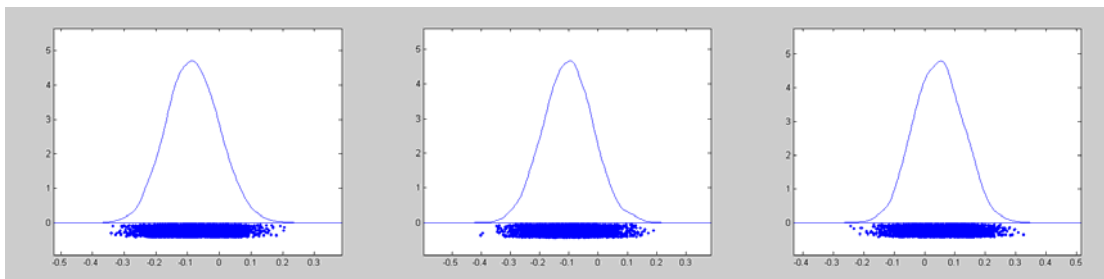
(38) C_{MG}



3. Resultados para a equação que tem variação da produção industrial no estado do Espírito Santo como variável dependente

Dependent Variable es			
R-squared	0.4581		
Rbar-squared	-0.5444		
sige	2.7694		
Nobs, Nvars	63, 8		
ndraws,nomit	5000, 100		
r-value	4		
Variável	Moda do Coef.	Variável	Moda do Coef.
ba lag1	-0.084851	sp lag4	0.00007
ba lag2	-0.104551	pr lag1	0.003427
ba lag3	0.04442	pr lag2	-0.007351
ba lag4	0.100237	pr lag3	-0.013549
mg lag1	0.159361	pr lag4	0.003476
mg lag2	-0.196022	sc lag1	0.016922
mg lag3	0.139126	sc lag2	-0.005715
mg lag4	-0.227781	sc lag3	0.003675
es lag1	0.196999	sc lag4	-0.01352
es lag2	0.057684	rs lag1	-0.00857
es lag3	-0.07506	rs lag2	0.001348
es lag4	-0.181935	rs lag3	-0.010243
rj lag1	-0.534865	rs lag4	0.004345
rj lag2	-0.374701	ec term ba	0.20223
rj lag3	0.070062	ec term mg	0.054952
rj lag4	-0.04636	ec term es	-0.317443

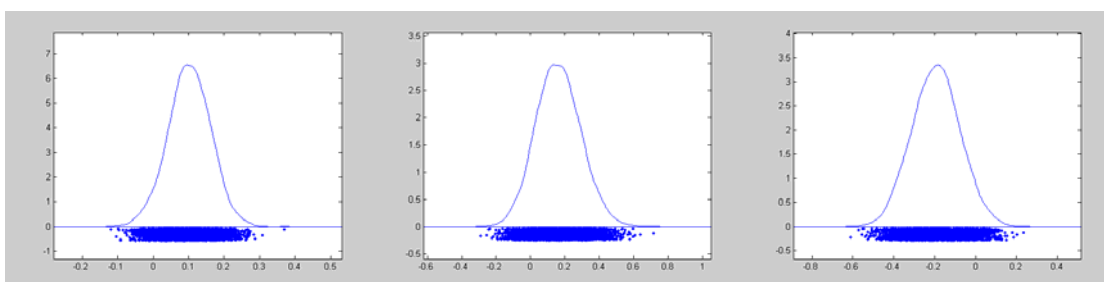
sp lag1	-0.008142	ec term rj	0.617632
sp lag2	-0.001942	ec term sp	0.016442
sp lag3	0.003301	constant	-12.562717



(1) $\beta_{BA,t-1}$

(2) $\beta_{BA,t-2}$

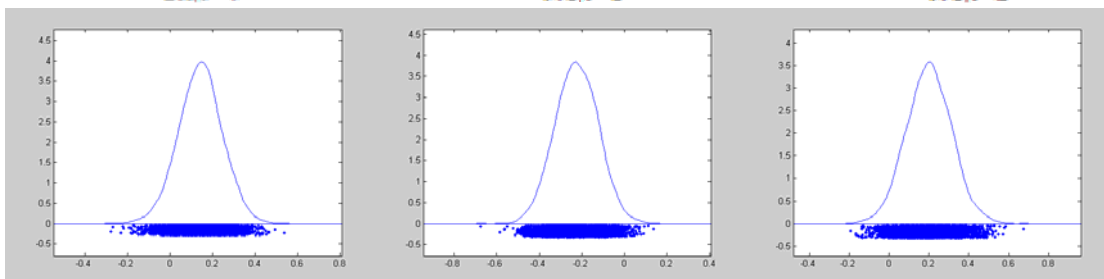
(3) $\beta_{BA,t-3}$



(4) $\beta_{BA,t-4}$

(5) $\beta_{MG,t-1}$

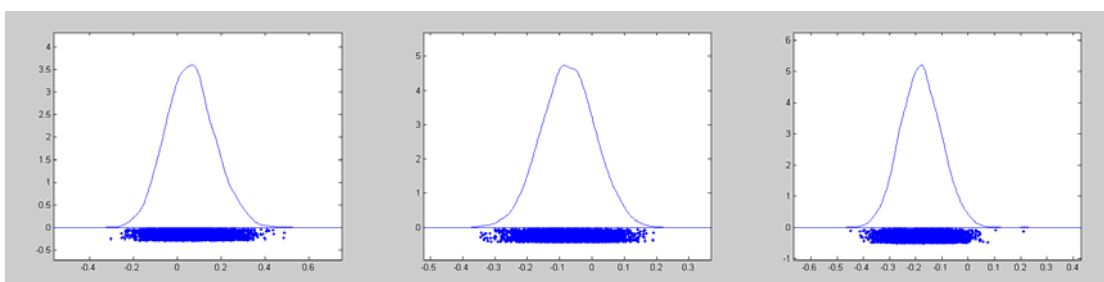
(6) $\beta_{MG,t-2}$



(7) $\beta_{MG,t-3}$

(8) $\beta_{MG,t-4}$

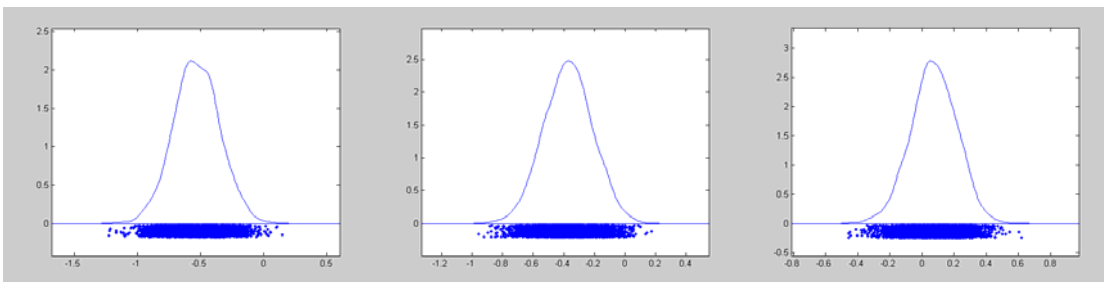
(9) $\beta_{ES,t-1}$



(10) $\beta_{ES,t-2}$

(11) $\beta_{ES,t-3}$

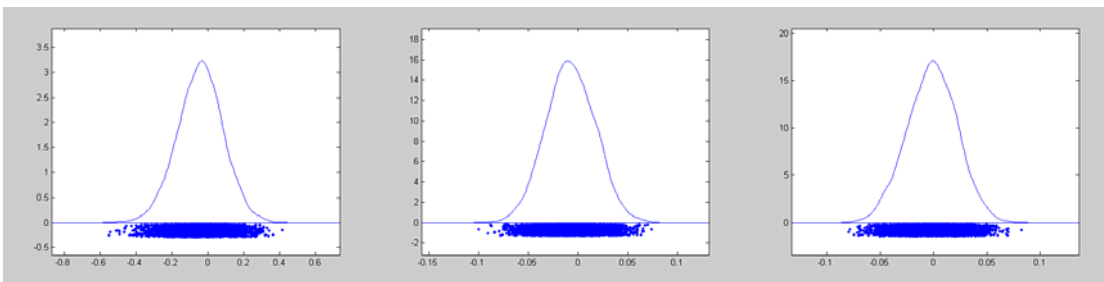
(12) $\beta_{ES,t-4}$



(13) $\beta_{RJ,t-1}$

(14) $\beta_{RJ,t-2}$

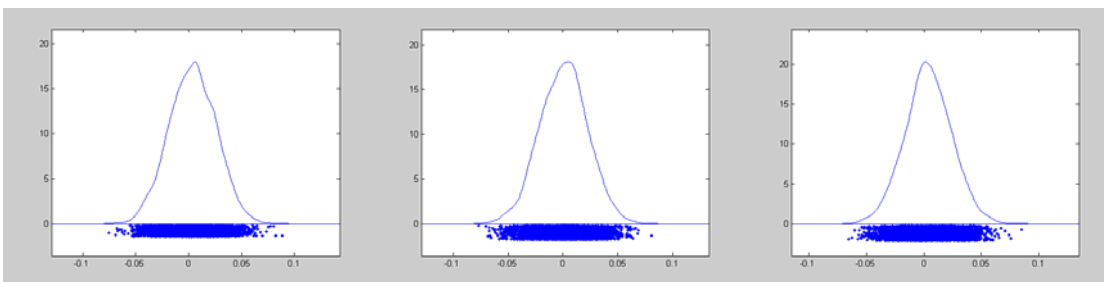
(15) $\beta_{RJ,t-3}$



(16) $\beta_{RJ,t-4}$

(17) $\beta_{SP,t-1}$

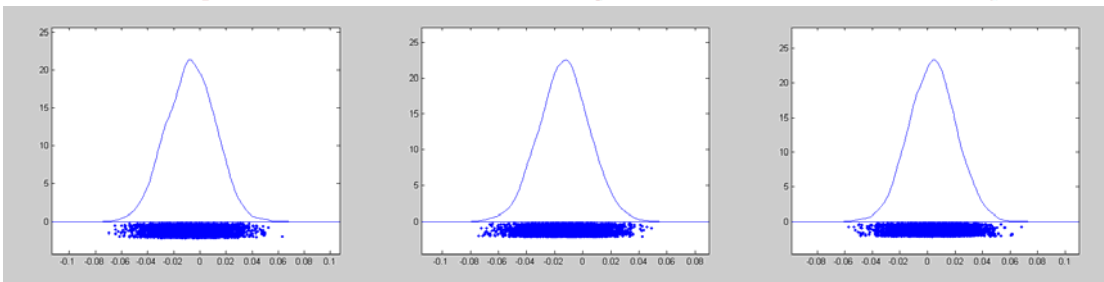
(18) $\beta_{SP,t-2}$



(19) $\beta_{SP,t-3}$

(20) $\beta_{SP,t-4}$

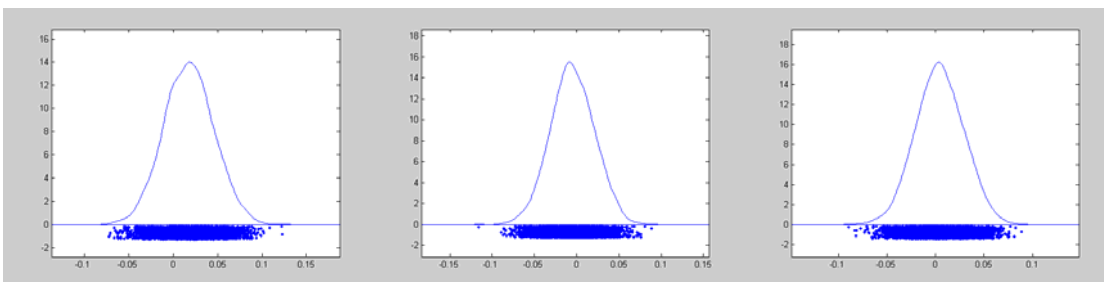
(21) $\beta_{PP,t-1}$



(22) $\beta_{PP,t-2}$

(23) $\beta_{PP,t-3}$

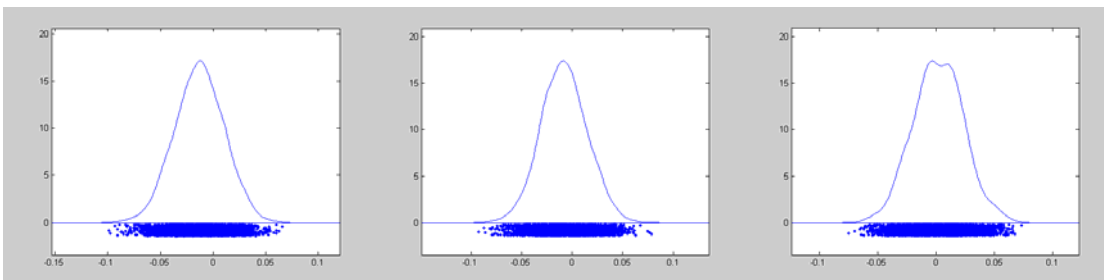
(24) $\beta_{PP,t-4}$



(25) $\beta_{SVC,t-1}$

(26) $\beta_{SVC,t-2}$

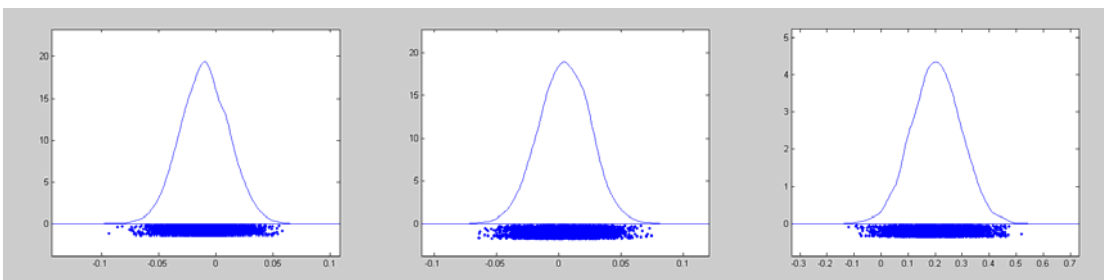
(27) $\beta_{SVC,t-3}$



(28) $\beta_{SC,t-1}$

(29) $\beta_{RF,t-1}$

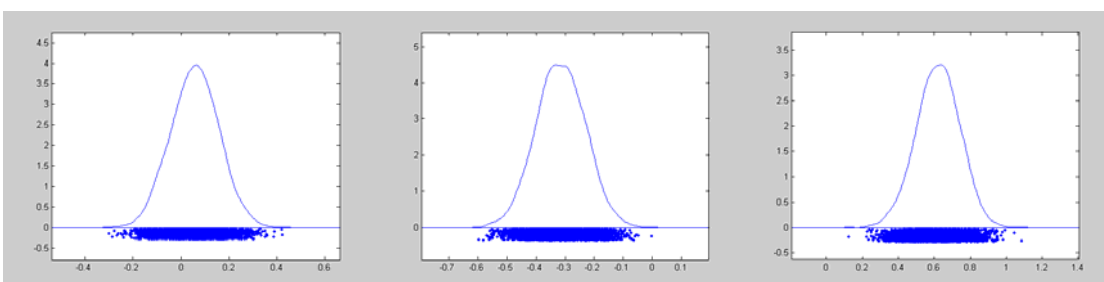
(30) $\beta_{RF,t-2}$



(31) $\beta_{RF,t-3}$

(32) $\beta_{RF,t-4}$

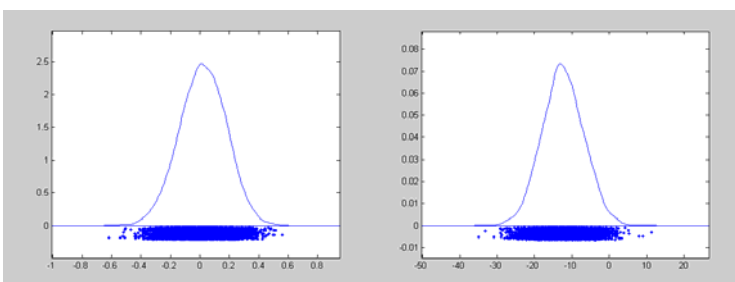
(33) ec term ba



(34) Ec term mg

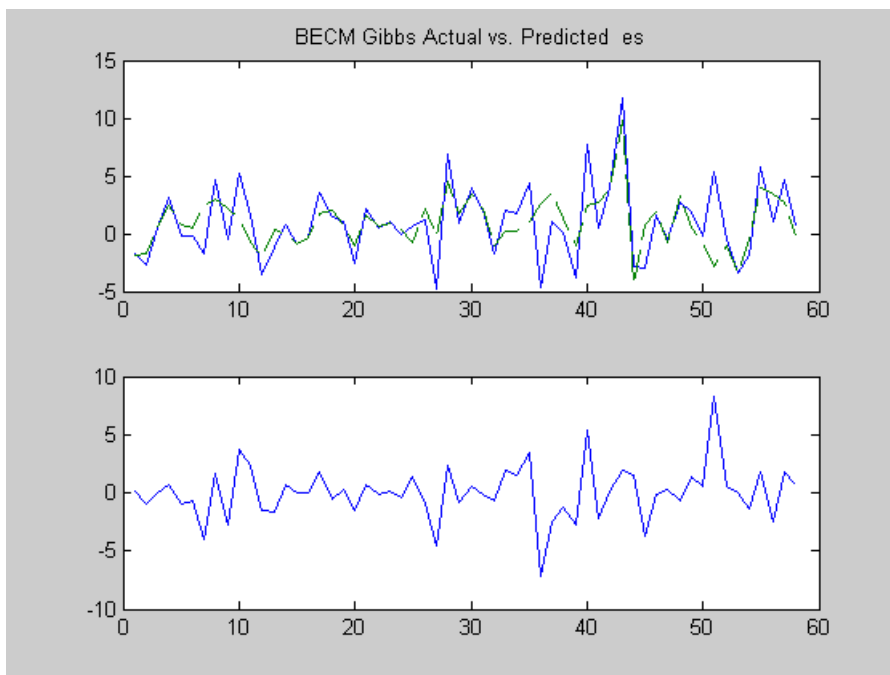
(35) ec term es

(36) ec term rj



(37) ec term sp

(38) C_{ES}



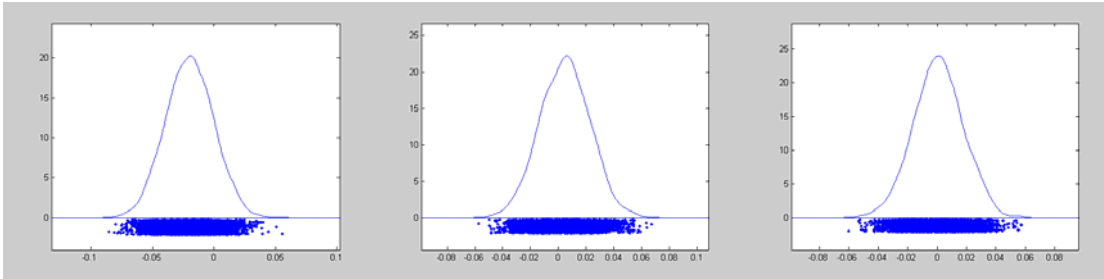
4. Resultados para a equação que tem variação da produção industrial no estado do Rio de Janeiro como variável dependente

Dependent Variable: rj

R-squared	0.419
Rbar-squared	-0.6559
sige	1.5188
Nobs, Nvars	58, 8
ndraws,nomit	5000, 100
r-value	4

Variável	Moda do Coef.	Variável	Moda do Coef.
ba lag1	-0.020187	sp lag4	-0.090042
ba lag2	0.004311	pr lag1	-0.011506
ba lag3	0.000456	pr lag2	-0.004543
ba lag4	0.001162	pr lag3	-0.002413
mg lag1	0.156157	pr lag4	-0.002862
mg lag2	-0.086898	sc lag1	-0.011986
mg lag3	-0.04416	sc lag2	-0.002326
mg lag4	-0.031959	sc lag3	-0.003543
es lag1	-0.014675	sc lag4	-0.001736
es lag2	-0.084455	rs lag1	0.001004
es lag3	-0.119406	rs lag2	-0.002366
es lag4	0.009389	rs lag3	0.003691
rj lag1	0.207055	rs lag4	-0.001127
rj lag2	0.064793	ec term ba	-0.027228
rj lag3	-0.250432	ec term mg	-0.215178
rj lag4	-0.033164	ec term es	0.063717

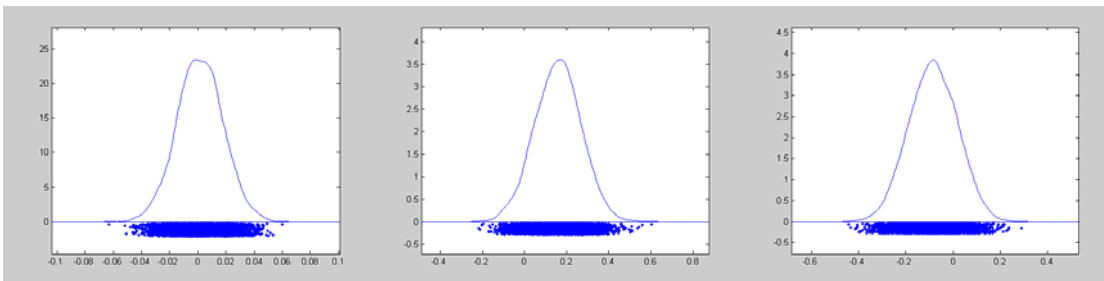
sp lag1	-0.307772	ec term rj	-0.120902
sp lag2	0.149132	ec term sp	0.157661
sp lag3	-0.048129	constant	-1.298941



(1) $\beta_{BA,t-1}$

(2) $\beta_{BA,t-2}$

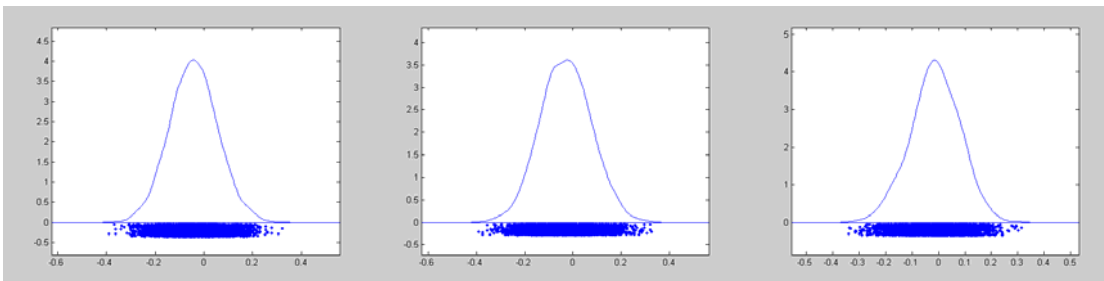
(3) $\beta_{BA,t-3}$



(4) $\beta_{BA,t-4}$

(5) $\beta_{MG,t-1}$

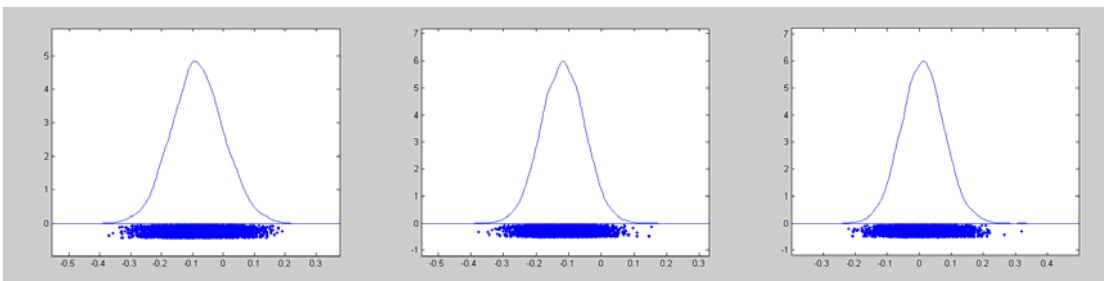
(6) $\beta_{MG,t-2}$



(7) $\beta_{MG,t-3}$

(8) $\beta_{MG,t-4}$

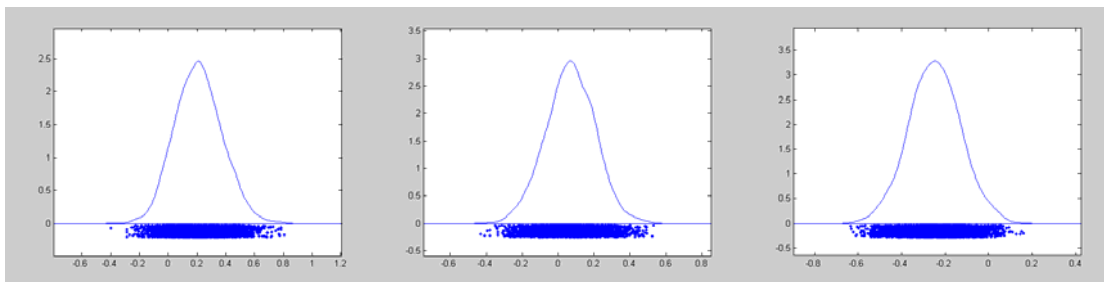
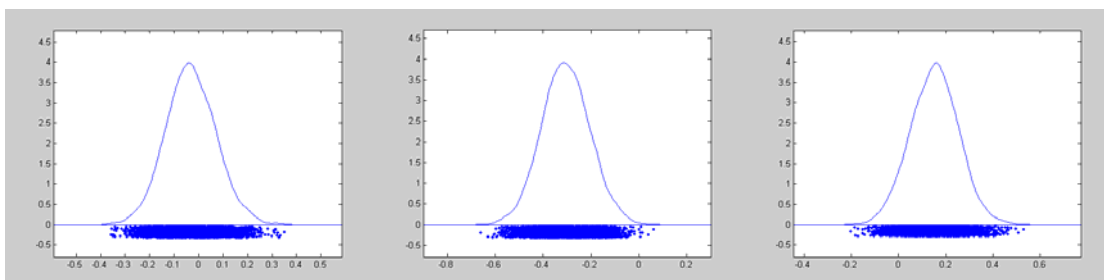
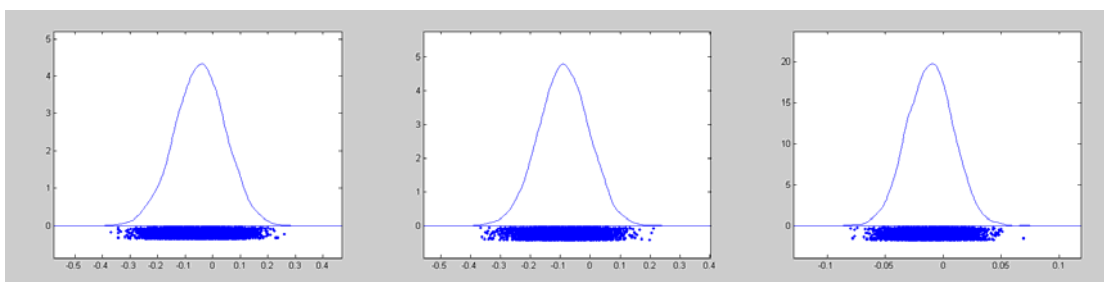
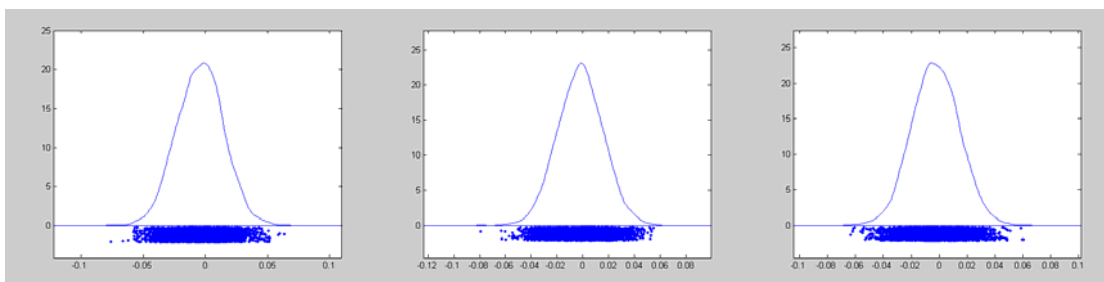
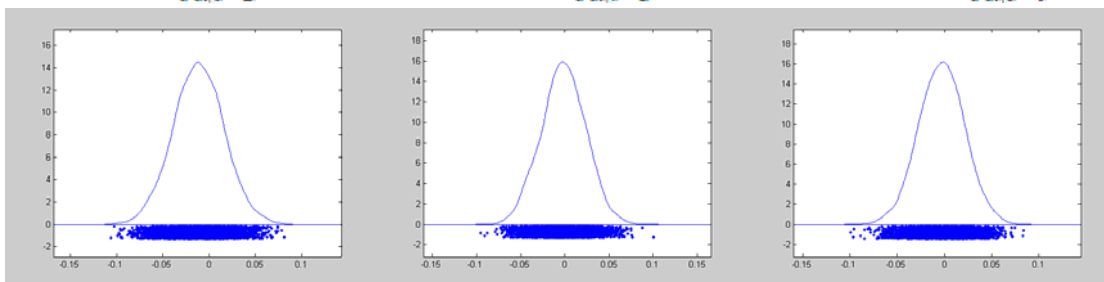
(9) $\beta_{ES,t-1}$

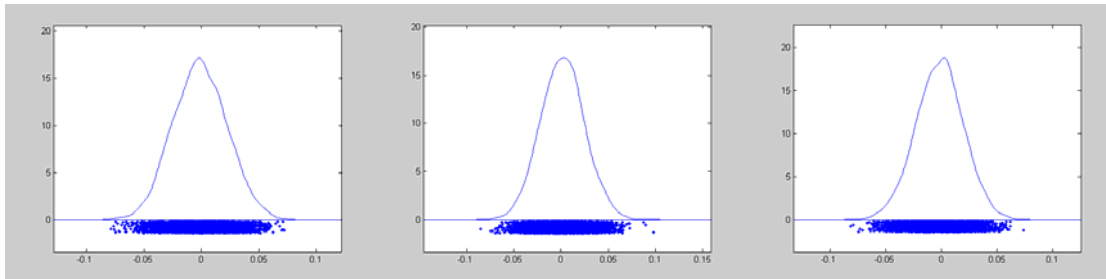
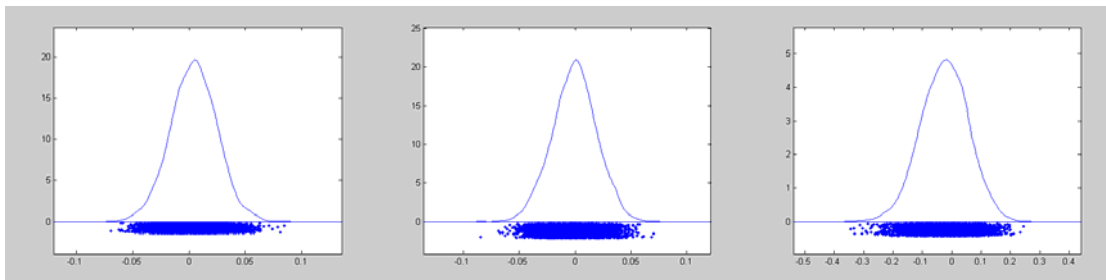


(10) $\beta_{ES,t-2}$

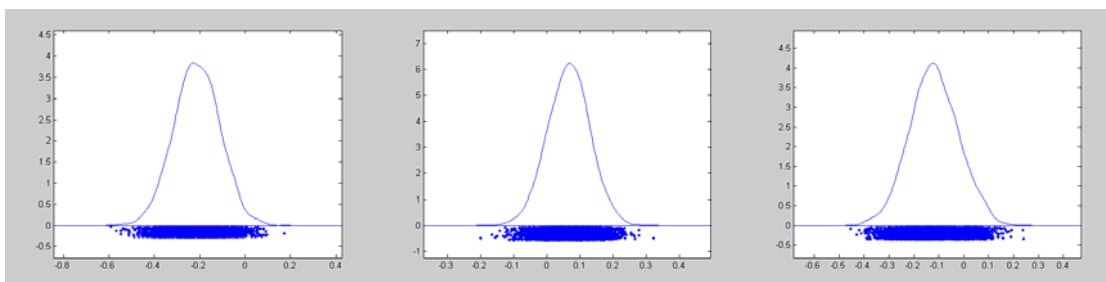
(11) $\beta_{ES,t-3}$

(12) $\beta_{ES,t-4}$

(13) $\beta_{R,t-1}$ (14) $\beta_{R,t-2}$ (15) $\beta_{R,t-3}$ (16) $\beta_{R,t-4}$ (17) $\beta_{SP,t-1}$ (18) $\beta_{SP,t-2}$ (19) $\beta_{SP,t-3}$ (20) $\beta_{SP,t-4}$ (21) $\beta_{PR,t-1}$ (22) $\beta_{PR,t-2}$ (23) $\beta_{PR,t-3}$ (24) $\beta_{PR,t-4}$ (25) $\beta_{SUC,t-1}$ (26) $\beta_{SUC,t-2}$ (27) $\beta_{SUC,t-3}$

(28) $\beta_{SCI,t-1}$ (29) $\beta_{RF,t-1}$ (30) $\beta_{RF,t-2}$ (31) $\beta_{RF,t-3}$ (32) $\beta_{RF,t-4}$

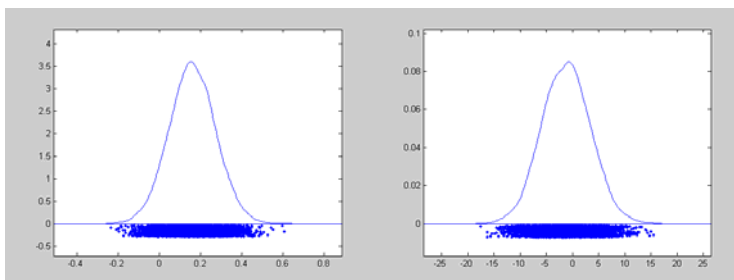
(33) ec term ba



(34) Ec term mg

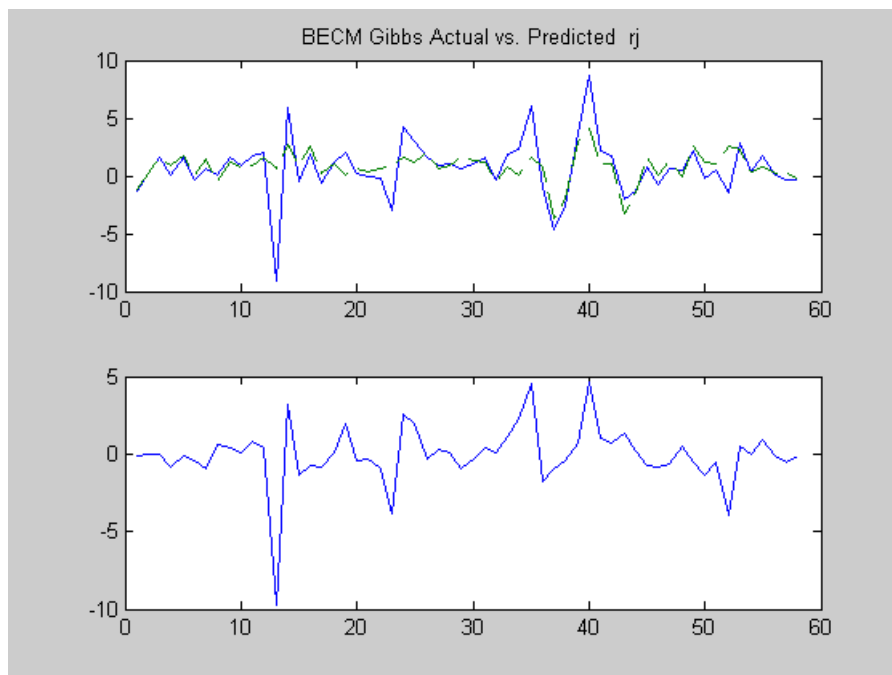
(35) ec term es

(36) ec term rj



(37) ec term sp

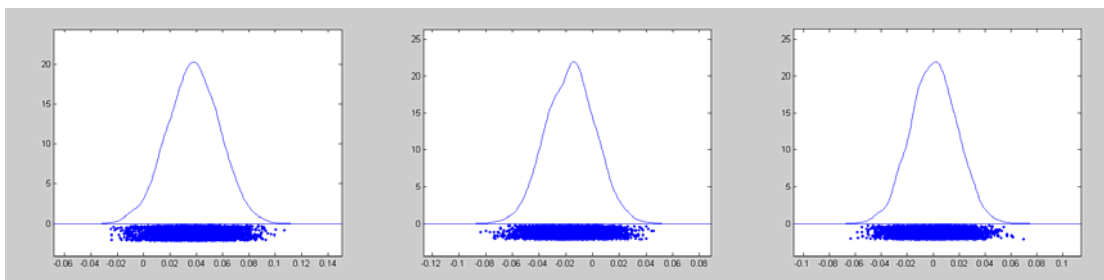
(38) c_{RJ}



5. Resultados para a equação que tem variação da produção industrial no estado da Bahia como variável dependente

Dependent Variable sp			
R-squared	0.5319		
Rbar-squared	-0.334		
sige	3.177		
Nobs, Nvars	63, 8		
ndraws,nomit	5000, 100		
r-value	4		
Variável	Moda do Coef.	Variável	Moda do Coef.
ba lag1	0.037442	sp lag4	-0.078433
ba lag2	-0.016801	pr lag1	-0.018277
ba lag3	0.000224	pr lag2	-0.082578
ba lag4	-0.002659	pr lag3	0.015831
mg lag1	-0.036209	pr lag4	0.00411
mg lag2	-0.273968	sc lag1	0.035903
mg lag3	0.198544	sc lag2	-0.008651
mg lag4	0.010751	sc lag3	-0.008627
es lag1	0.005223	sc lag4	-0.003246
es lag2	-0.013732	rs lag1	0.003855
es lag3	-0.003038	rs lag2	-0.016567
es lag4	-0.015518	rs lag3	0.000293
rj lag1	0.32956	rs lag4	-0.008516
rj lag2	0.012902	ec term ba	-0.151486
rj lag3	0.208823	ec term mg	0.017529
rj lag4	-0.089306	ec term es	-0.273906

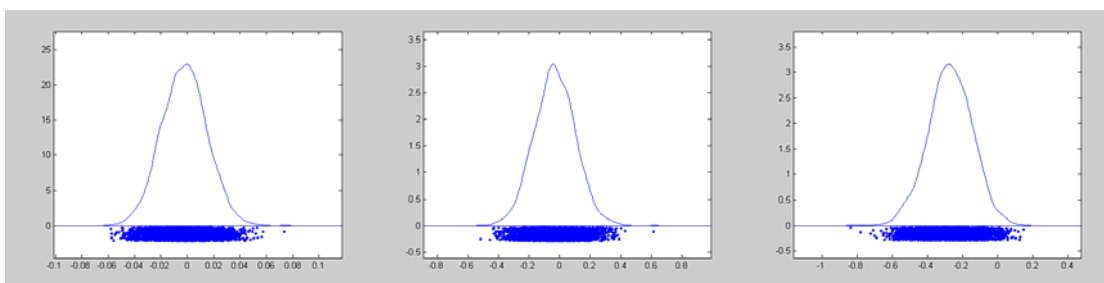
sp lag1	0.185687	ec term rj	0.037346
sp lag2	0.058572	ec term sp	0.411908
sp lag3	-0.017595	constant	6.596459



(1) $\beta_{BA,t-1}$

(2) $\beta_{BA,t-2}$

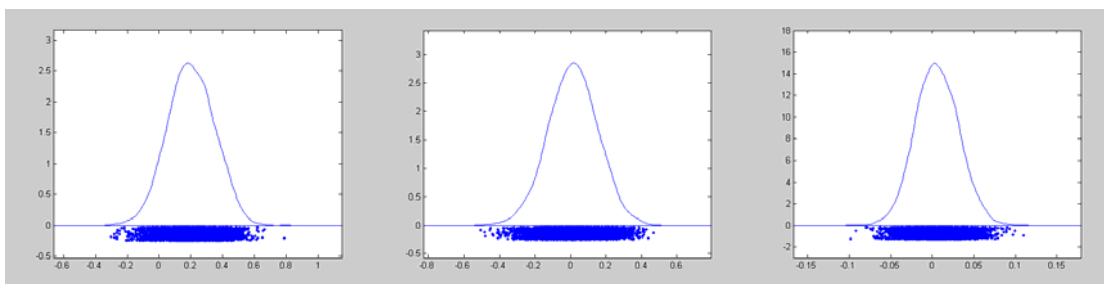
(3) $\beta_{BA,t-3}$



(4) $\beta_{BA,t-4}$

(5) $\beta_{MG,t-1}$

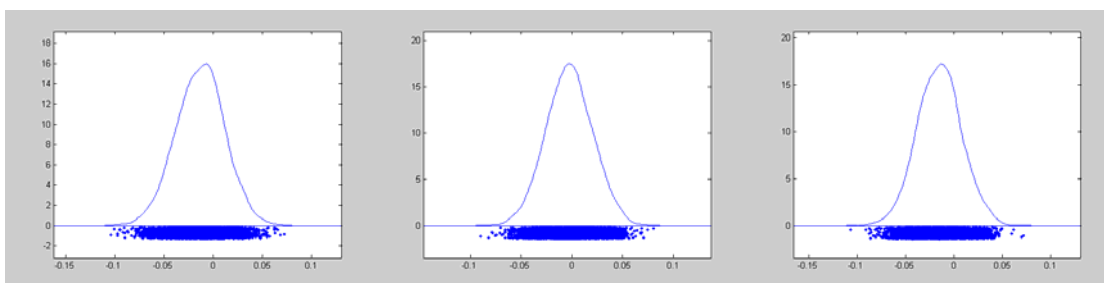
(6) $\beta_{MG,t-2}$



(7) $\beta_{MG,t-3}$

(8) $\beta_{MG,t-4}$

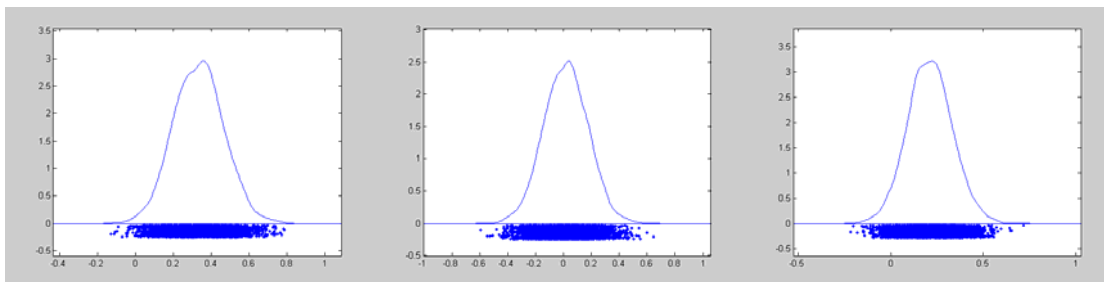
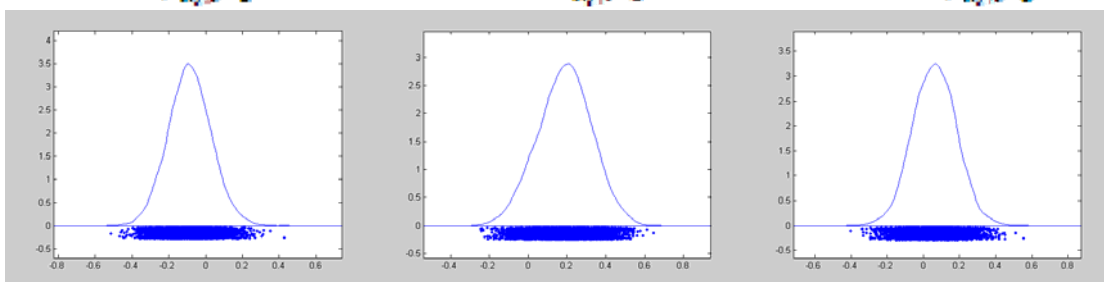
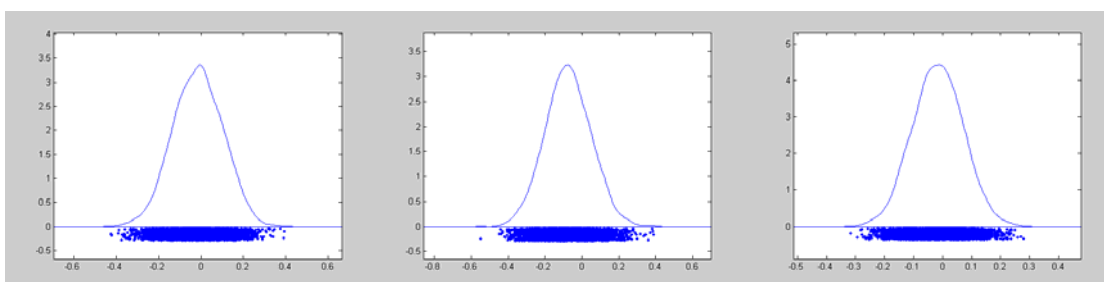
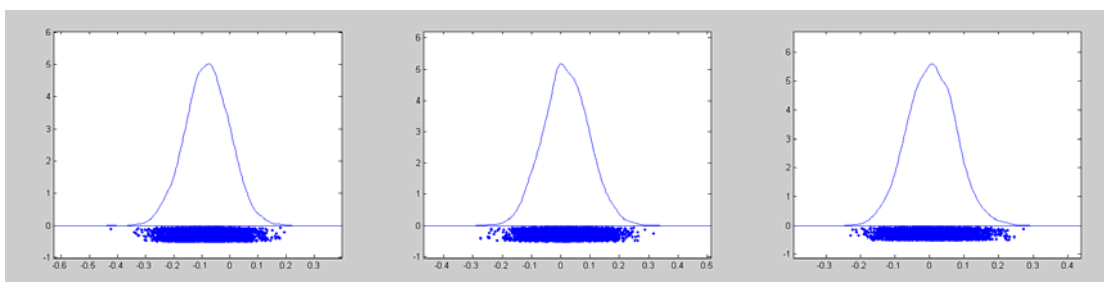
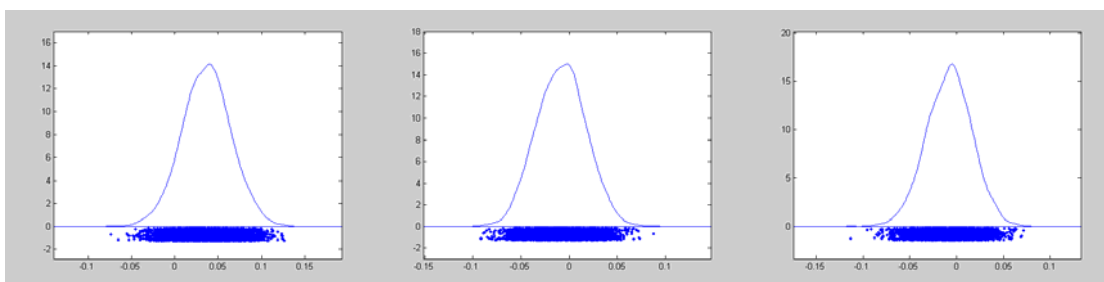
(9) $\beta_{ES,t-1}$

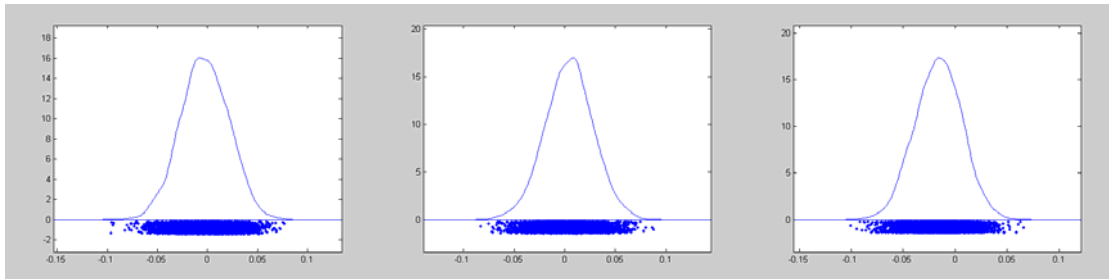
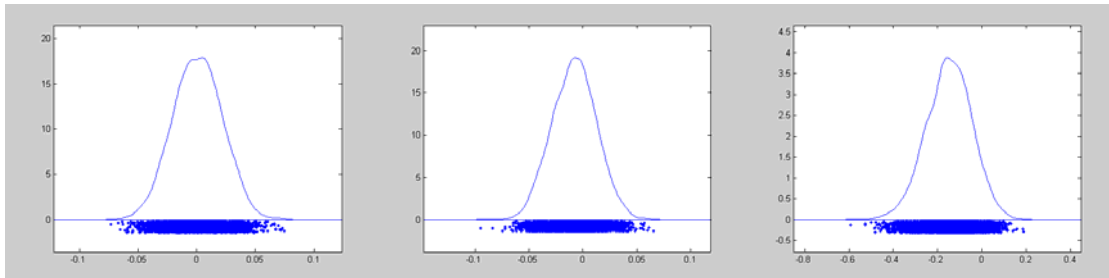


(10) $\beta_{ES,t-2}$

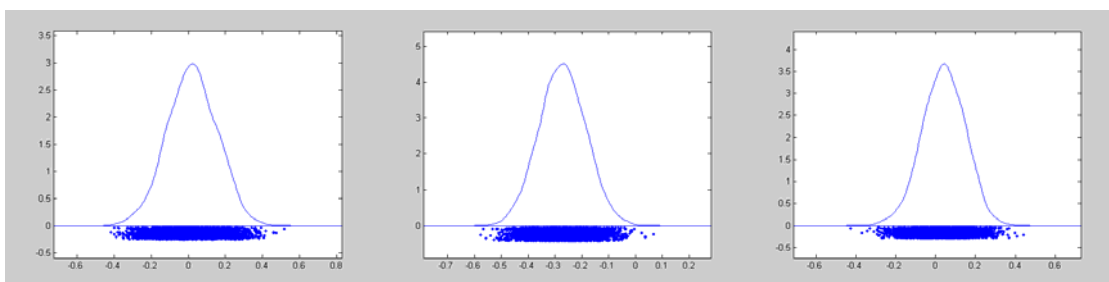
(11) $\beta_{ES,t-3}$

(12) $\beta_{ES,t-4}$

(13) $\beta_{R|S-1}$ (14) $\beta_{R|S-2}$ (15) $\beta_{R|S-3}$ (16) $\beta_{R|S-1}$ (17) $\beta_{SP,S-1}$ (18) $\beta_{SP,S-2}$ (19) $\beta_{SP,S-2}$ (20) $\beta_{SP,S-3}$ (21) $\beta_{SP,S-4}$ (22) $\beta_{FHS,S-2}$ (23) $\beta_{FHS,S-3}$ (24) $\beta_{FHS,S-4}$ (25) $\beta_{SUC,S-1}$ (26) $\beta_{SUC,S-2}$ (27) $\beta_{SUC,S-3}$

(28) $\beta_{SC,t-1}$ (29) $\beta_{RF,t-1}$ (30) $\beta_{RF,t-2}$ (31) $\beta_{RF,t-2}$ (32) $\beta_{RF,t-1}$

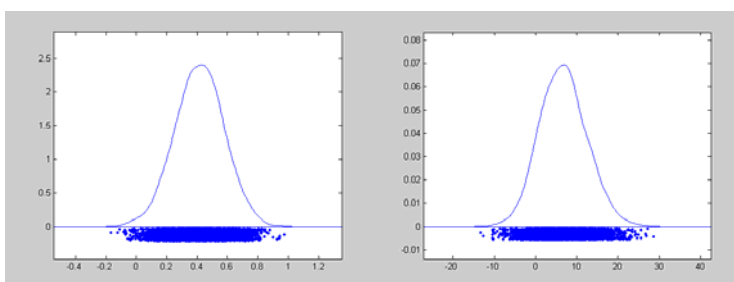
(33) ec term ba



(34) Ec term mg

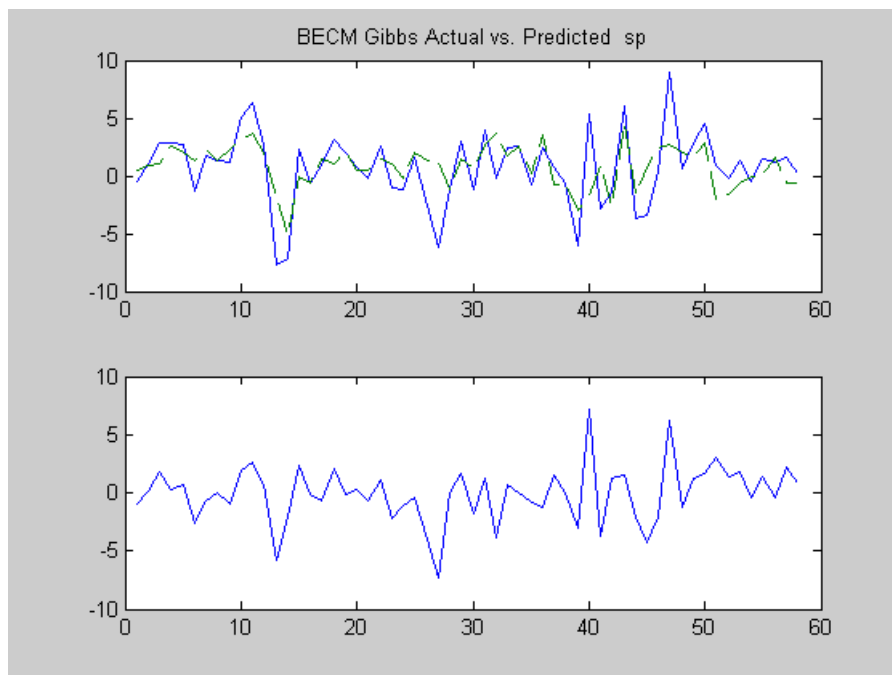
(35) ec term es

(36) ec term rj



(37) ec term sp

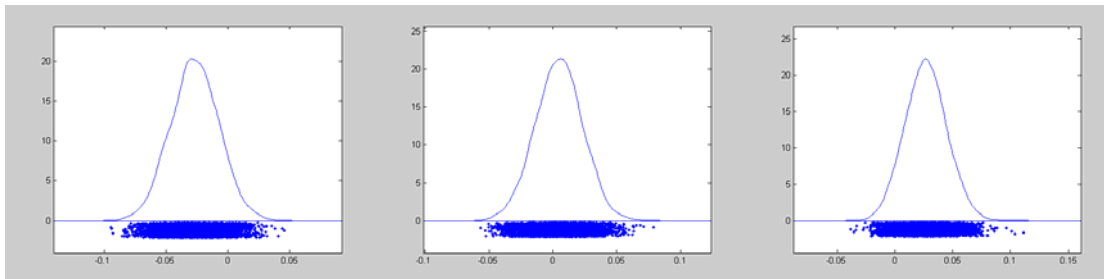
(38) C_{sp}



6. Resultados para a equação que tem variação da produção industrial no estado do Paraná como variável dependente

Dependent Variable pr			
R-squared	0.484		
Rbar-squared	-0.489		
sige	4.6288		
Nobs, Nvars	63, 8		
ndraws,nomit	5000, 100		
r-value	4		
Variável	Moda do Coef.	Variável	Moda do Coef.
ba lag1	-0.027206	sp lag4	-0.046358
ba lag2	0.004304	pr lag1	-0.035249
ba lag3	0.026394	pr lag2	0.35672
ba lag4	-0.015028	pr lag3	0.262272
mg lag1	0.002085	pr lag4	0.058763
mg lag2	-0.024586	sc lag1	0.364889
mg lag3	-0.023016	sc lag2	0.133552
mg lag4	0.007361	sc lag3	-0.037274
es lag1	-0.004111	sc lag4	0.101307
es lag2	-0.015445	rs lag1	0.00667
es lag3	-0.00925	rs lag2	-0.002402
es lag4	-0.002296	rs lag3	0.013533
rj lag1	0.007983	rs lag4	0.004326
rj lag2	0.000523	ec term ba	-0.376901
rj lag3	0.000043	ec term mg	0.779628

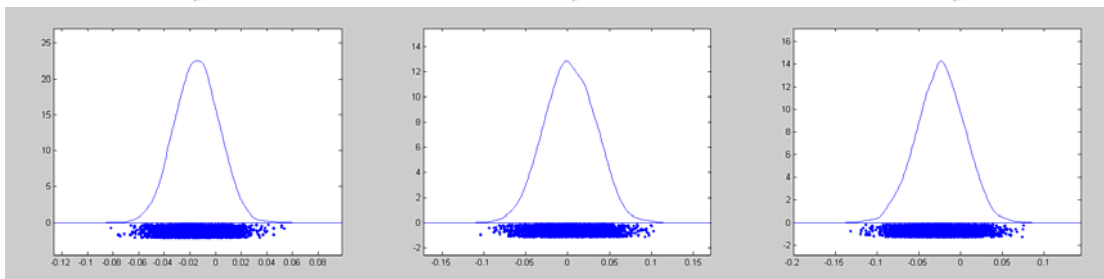
rj lag4	-0.031566	ec term es	0.119355
sp lag1	-0.155381	ec term rj	0.059273
sp lag2	-0.082388	ec term sp	-0.674582
sp lag3	-0.011134	constant	15.050907



(1) β_{BAT-1}

(2) β_{BAT-2}

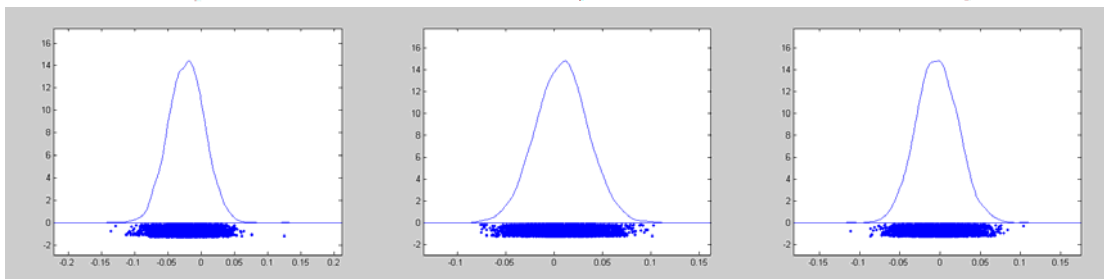
(3) β_{BAT-3}



(4) β_{BAT-4}

(5) β_{MGR-1}

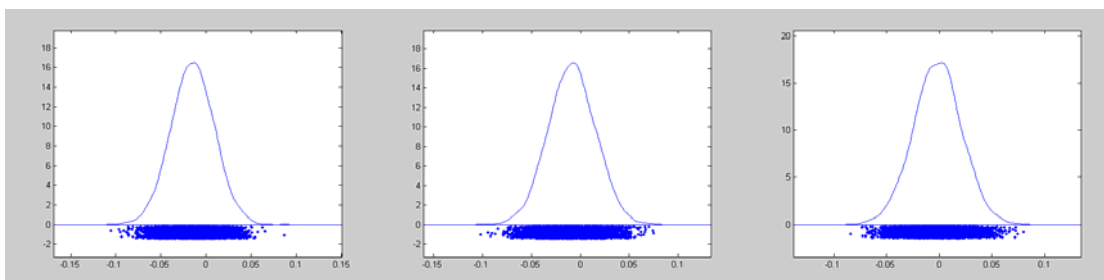
(6) β_{MGR-2}



(7) β_{MGR-3}

(8) β_{MGR-4}

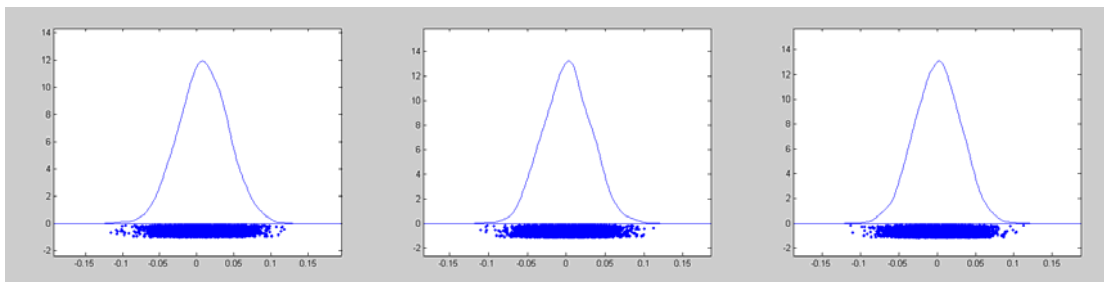
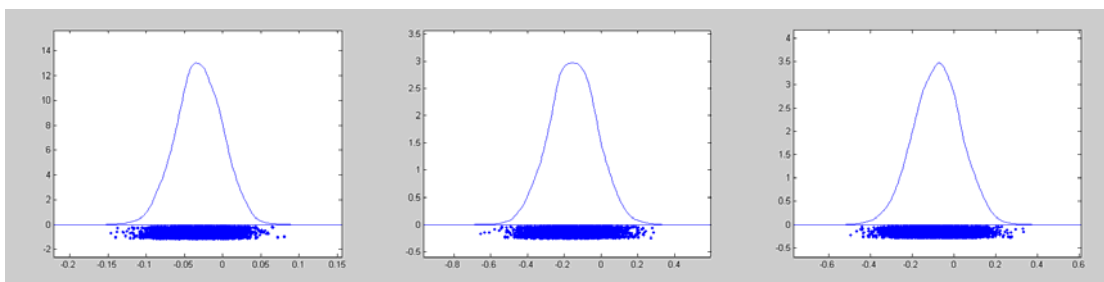
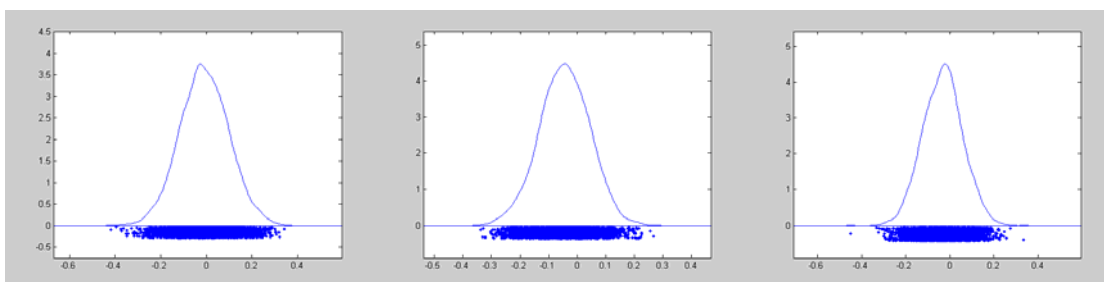
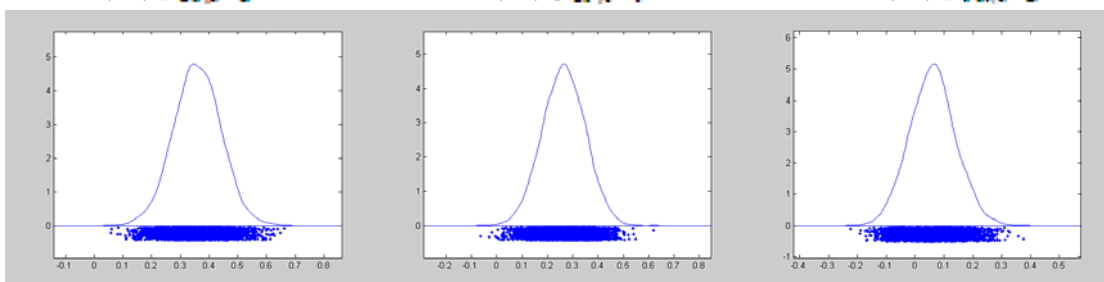
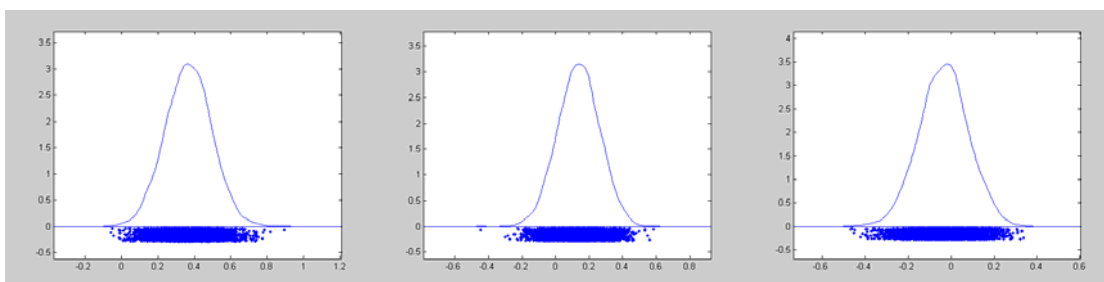
(9) β_{ESL-1}

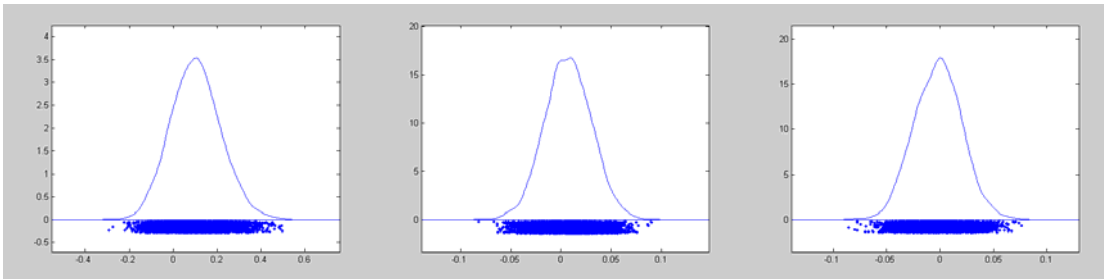


(10) β_{ESL-2}

(11) β_{ESL-3}

(12) β_{ESL-4}

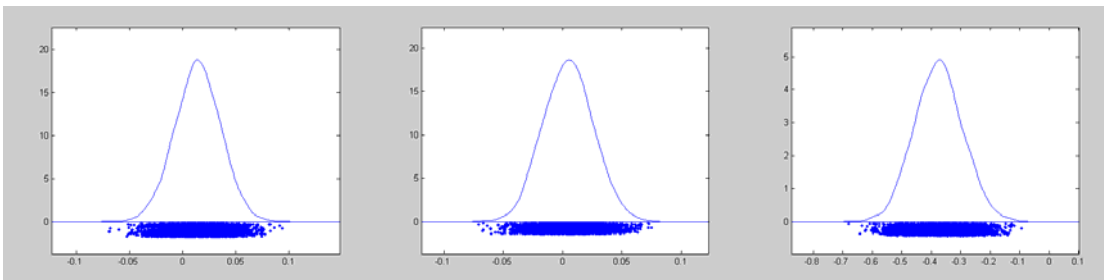
(13) $\beta_{RJT,t-1}$ (14) $\beta_{RJT,t-2}$ (15) $\beta_{RJT,t-3}$ (16) $\beta_{RJT,t-4}$ (17) $\beta_{SP,t-1}$ (18) $\beta_{SP,t-2}$ (19) $\beta_{SP,t-3}$ (20) $\beta_{SP,t-4}$ (21) $\beta_{PR,t-1}$ (22) $\beta_{HKC,t-2}$ (23) $\beta_{HKC,t-3}$ (24) $\beta_{HKC,t-4}$ (25) $\beta_{SUC,t-1}$ (26) $\beta_{SUC,t-2}$ (27) $\beta_{SUC,t-3}$



(28) $\beta_{SC,t-1}$

(29) $\beta_{RF,t-1}$

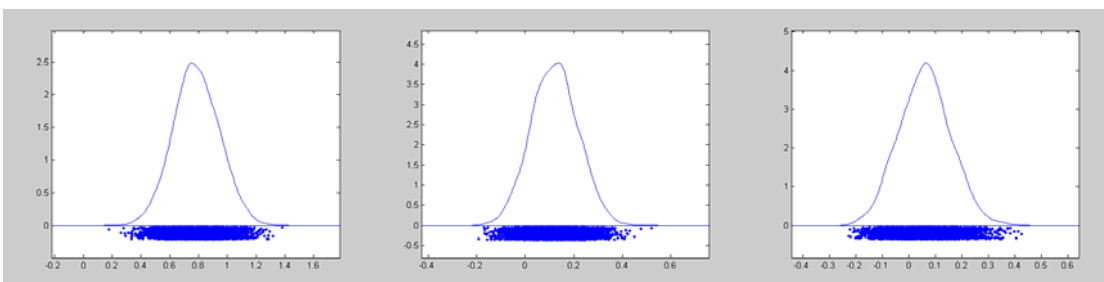
(30) $\beta_{RF,t-2}$



(31) $\beta_{RS,t-3}$

(32) $\beta_{RS,t-4}$

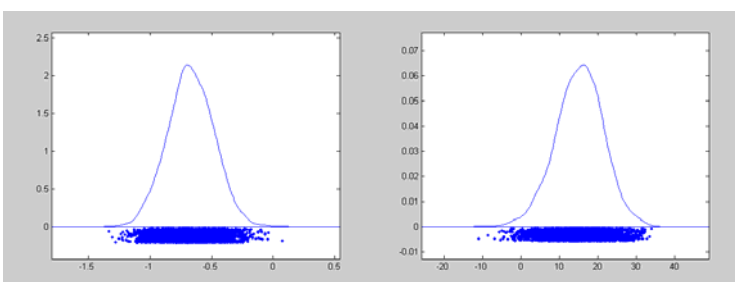
(33) ec term ba



(34) Ec term mg

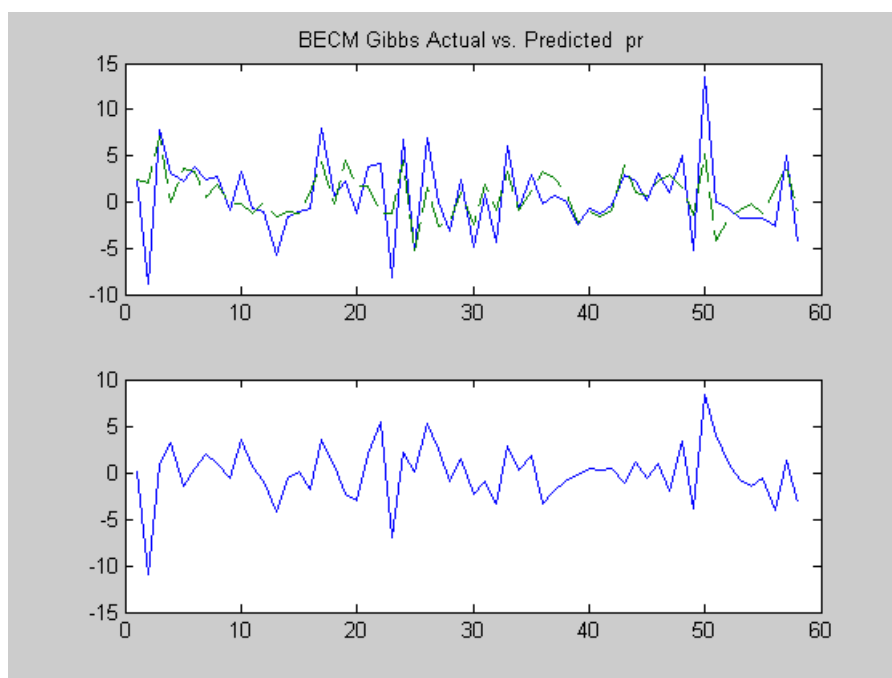
(35) ec term es

(36) ec term rj



(37) ec term sp

(38) C_{FR}



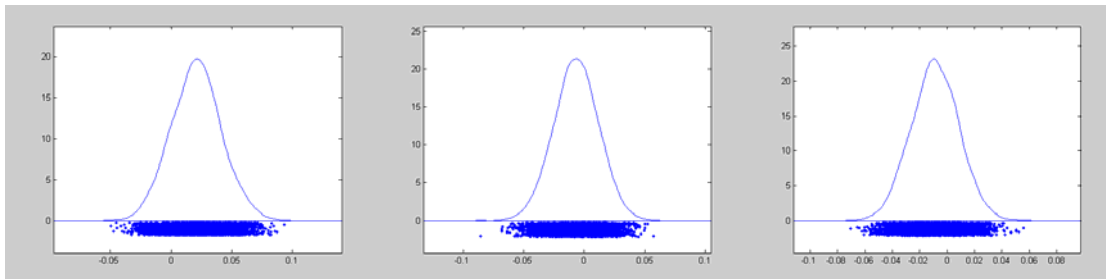
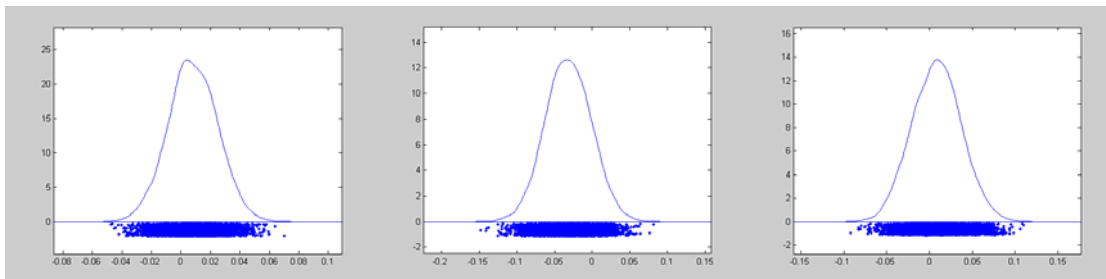
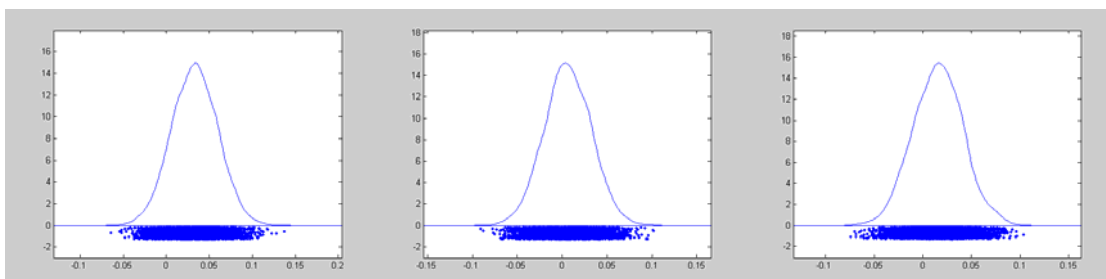
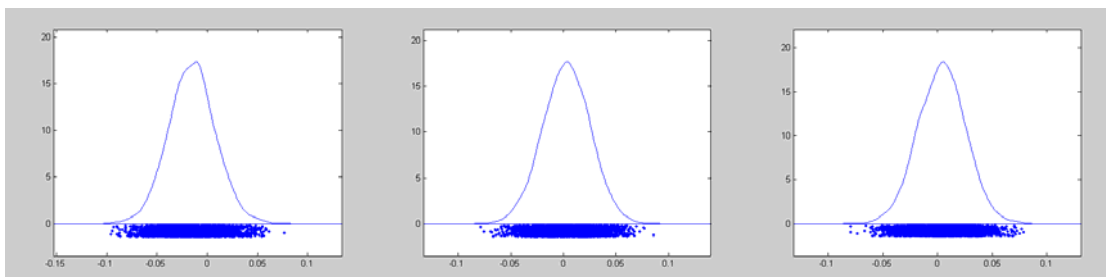
7. Resultados para a equação que tem variação da produção industrial no estado da Bahia como variável dependente

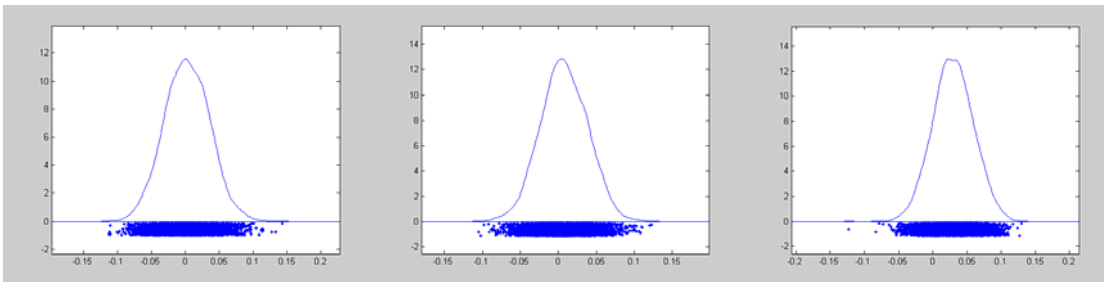
Dependent Variable: sc

R-squared	0.5002
Rbar-squared	-0.4245
sige	2.9214
Nobs, Nvars	63, 8
ndraws, nomit	5000, 100
r-value	4

Variável	Moda do Coef.	Variável	Moda do Coef.
ba lag1	0.020444	sp lag4	-0.008669
ba lag2	-0.007111	pr lag1	-0.016759
ba lag3	-0.009315	pr lag2	-0.023605
ba lag4	0.007787	pr lag3	0.004986
mg lag1	-0.032416	pr lag4	-0.01218
mg lag2	0.007801	sc lag1	0.584735
mg lag3	0.032383	sc lag2	-0.150017
mg lag4	0.005555	sc lag3	0.221529
es lag1	0.014966	sc lag4	-0.105377
es lag2	-0.015617	rs lag1	-0.12203
es lag3	0.002314	rs lag2	-0.133561
es lag4	0.00347	rs lag3	-0.024079
rj lag1	0.001903	rs lag4	-0.176737
rj lag2	0.006825	ec term ba	-0.018277

rj lag3	0.028687	ec term mg	-0.096267
rj lag4	-0.026162	ec term es	-0.434788
sp lag1	-0.002223	ec term rj	0.288435
sp lag2	0.012611	ec term sp	0.262169
sp lag3	0.007566	constant	1.921066

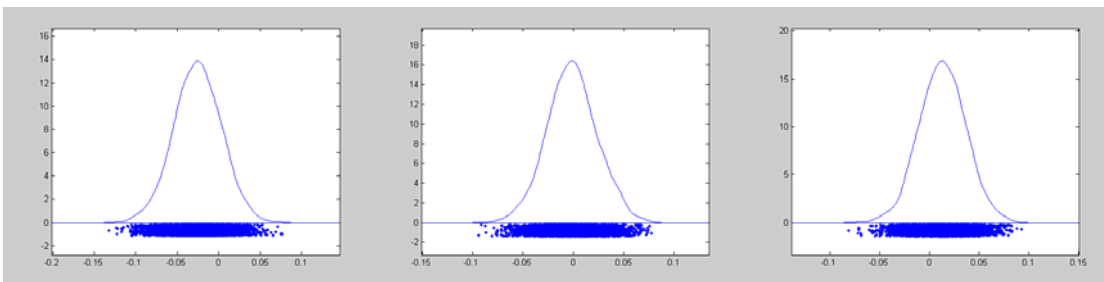
(1) $\beta_{EA,t-1}$ (2) $\beta_{EA,t-2}$ (3) $\beta_{EA,t-3}$ (4) $\beta_{EA,t-4}$ (5) $\beta_{MG,t-1}$ (6) $\beta_{MG,t-2}$ (7) $\beta_{MG,t-3}$ (8) $\beta_{MG,t-4}$ (9) $\beta_{ES,t-1}$ (10) $\beta_{ES,t-2}$ (11) $\beta_{ES,t-3}$ (12) $\beta_{ES,t-4}$



(13) $\beta_{RJ,t-1}$

(14) $\beta_{RJ,t-2}$

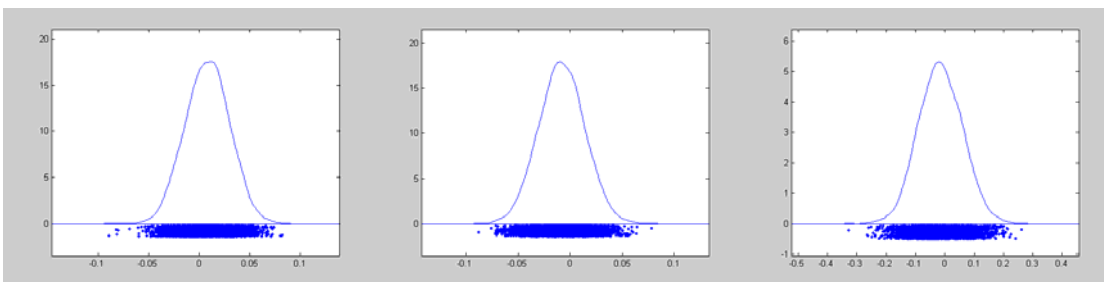
(15) $\beta_{RJ,t-3}$



(16) $\beta_{RJ,t-4}$

(17) $\beta_{SP,t-1}$

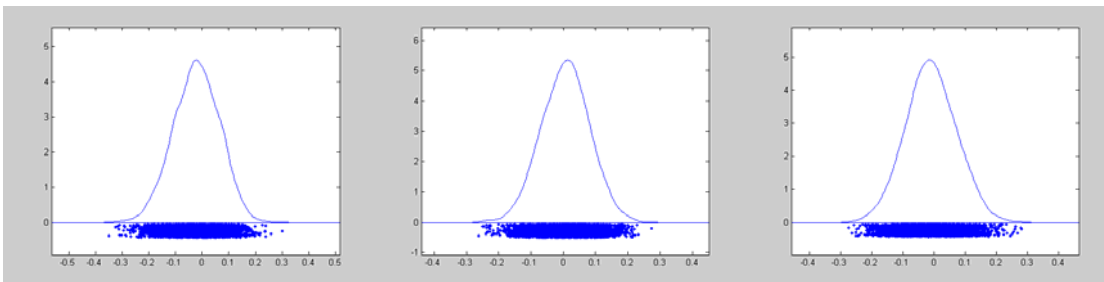
(18) $\beta_{SP,t-2}$



(19) $\beta_{SP,t-3}$

(20) $\beta_{SP,t-4}$

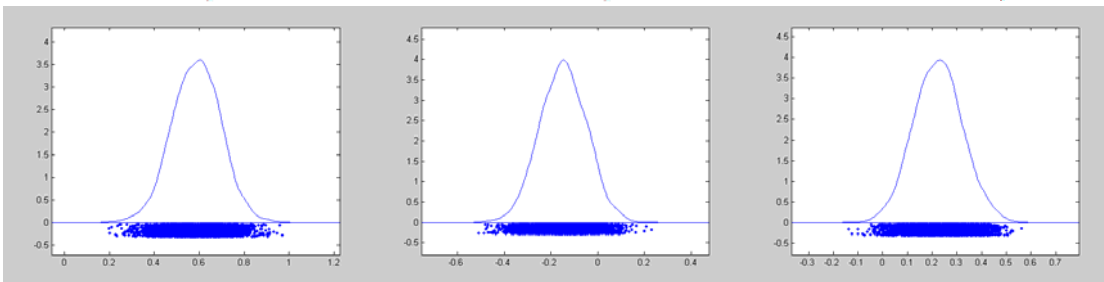
(21) $\beta_{PP,t-1}$



(22) $\beta_{PP,t-2}$

(23) $\beta_{PP,t-3}$

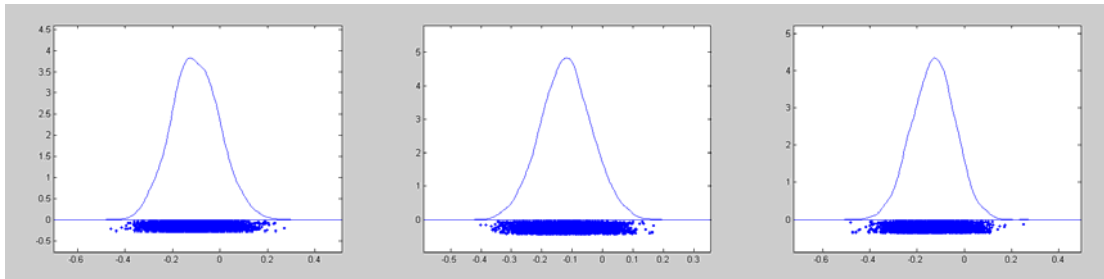
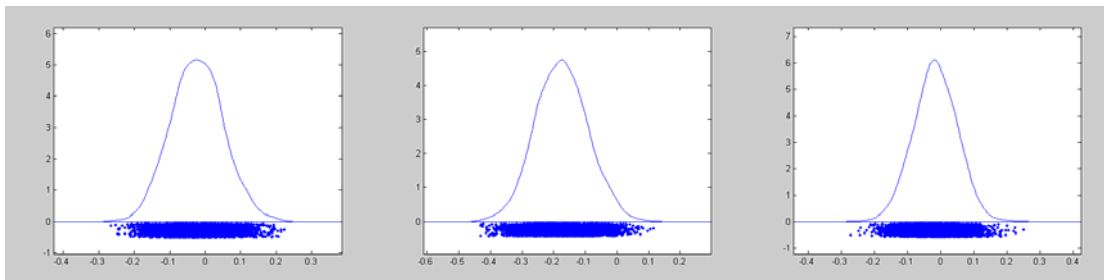
(24) $\beta_{PP,t-4}$



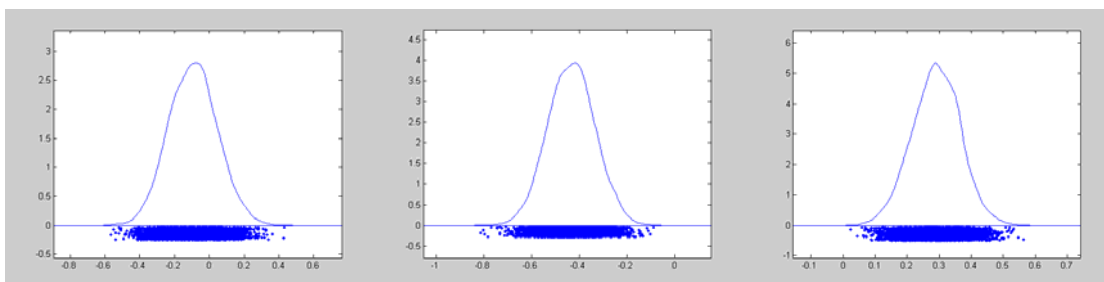
(25) $\beta_{SVC,t-1}$

(26) $\beta_{SVC,t-2}$

(27) $\beta_{SVC,t-3}$

(28) $\beta_{SC,t-1}$ (29) $\beta_{RF,t-1}$ (30) $\beta_{RF,t-2}$ (31) $\beta_{RF,t-3}$ (32) $\beta_{RF,t-4}$

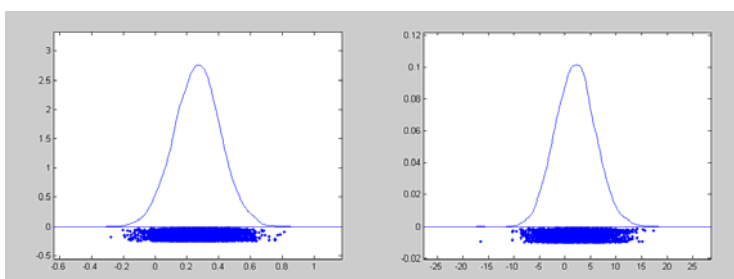
(33) ec term ba



(34) Ec term mg

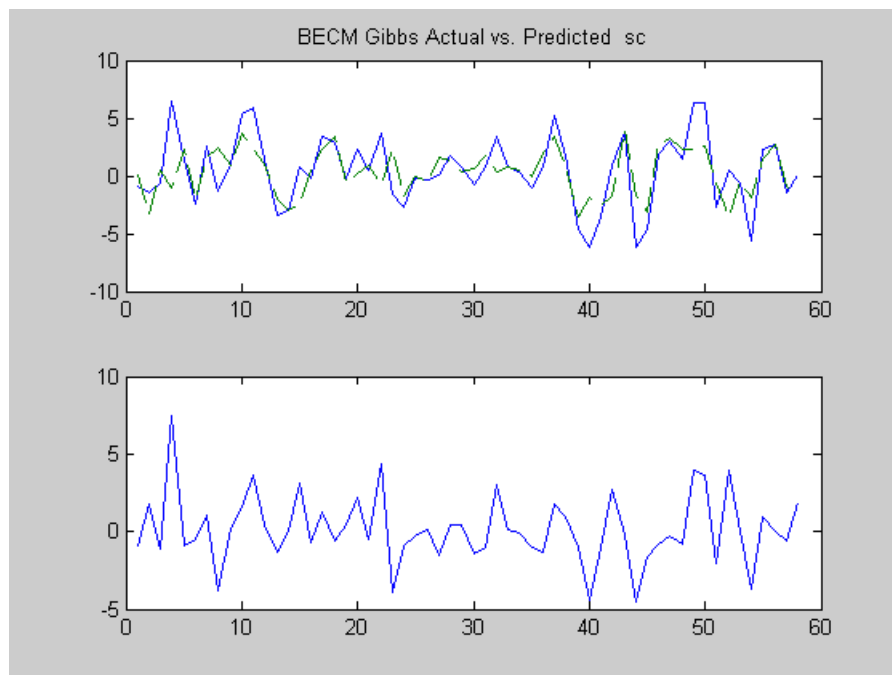
(35) ec term es

(36) ec term rj



(37) ec term sp

(38) C_{SC}



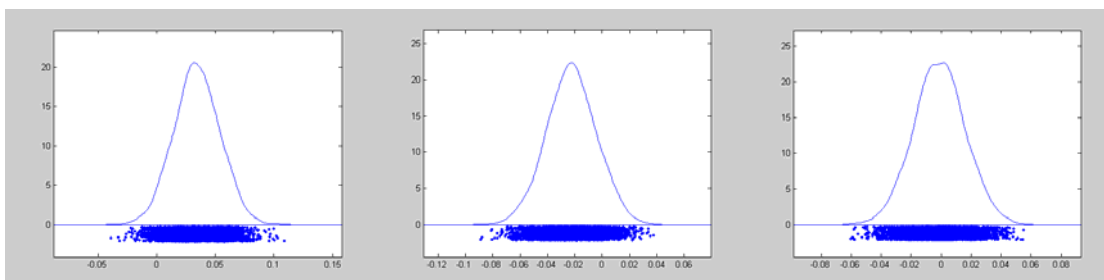
8. Resultados para a equação que tem variação da produção industrial no estado do Rio Grande do Sul como variável dependente

Dependent Variable rs

R-squared	0.4311
Rbar-squared	-0.6212
sige	3.5104
Nobs, Nvars	63, 8
ndraws,nomit	5000, 100
r-value	4

Variável	Moda do Coef.	Variável	Moda do Coef.
ba lag1	0.033787	sp lag4	-0.017792
ba lag2	-0.022843	pr lag1	0.010843
ba lag3	-0.00156	pr lag2	-0.006325
ba lag4	-0.004021	pr lag3	0.013553
mg lag1	-0.02995	pr lag4	-0.002655
mg lag2	0.022018	sc lag1	-0.012877
mg lag3	0.01971	sc lag2	-0.071178
mg lag4	-0.021887	sc lag3	-0.220603
es lag1	0.027346	sc lag4	-0.050774
es lag2	-0.019397	rs lag1	0.262694
es lag3	0.003922	rs lag2	-0.021504
es lag4	-0.009854	rs lag3	-0.00908
rj lag1	0.044244	rs lag4	0.016935
rj lag2	-0.02668	ec term ba	0.150608
rj lag3	0.009882	ec term mg	0.122971
rj lag4	-0.02899	ec term es	-0.112392

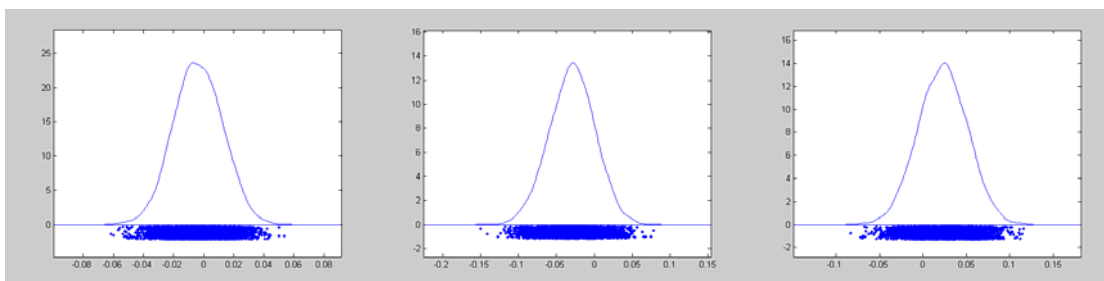
sp lag1	0.005111	ec term rj	-0.142584
sp lag2	0.015819	ec term sp	0.276112
sp lag3	0.006919	constant	9.012482



(1) $\beta_{EA,t-1}$

(2) $\beta_{EA,t-2}$

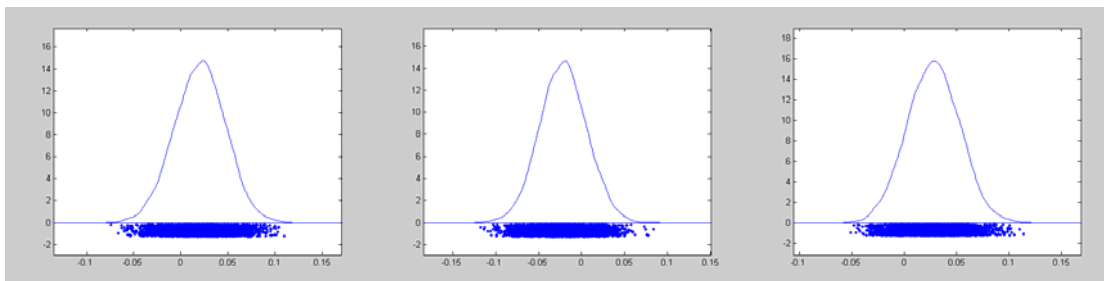
(3) $\beta_{EA,t-3}$



(4) $\beta_{EA,t-4}$

(5) $\beta_{MG,t-1}$

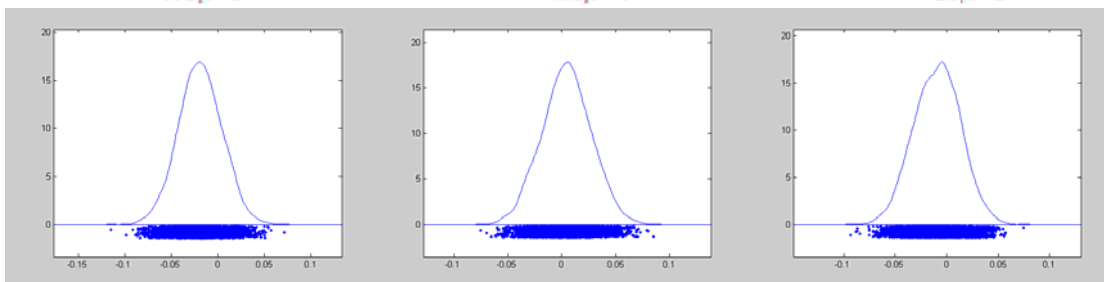
(6) $\beta_{MG,t-2}$



(7) $\beta_{MG,t-3}$

(8) $\beta_{MG,t-4}$

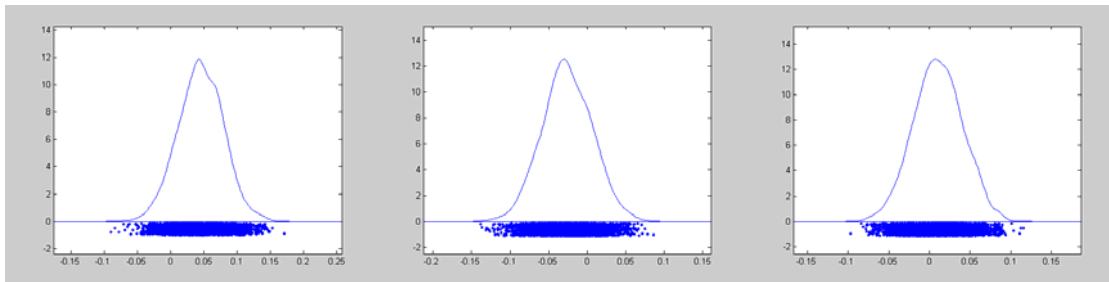
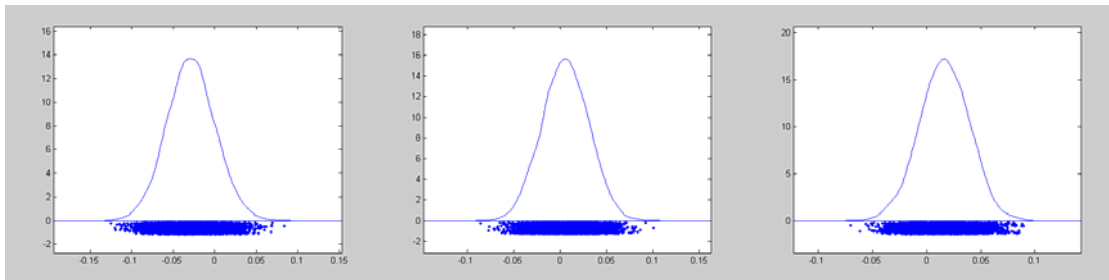
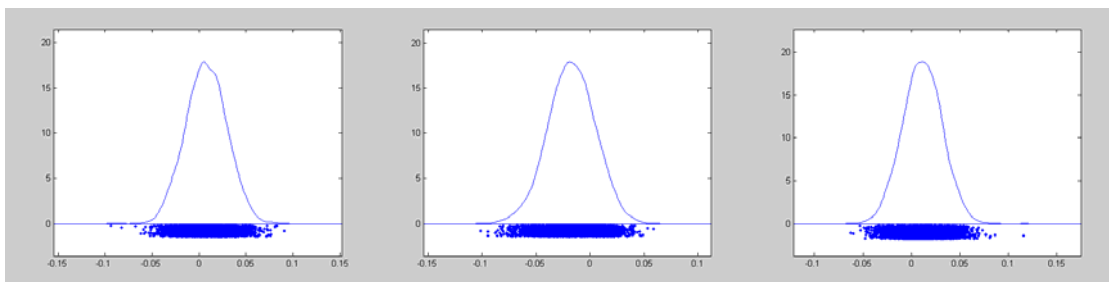
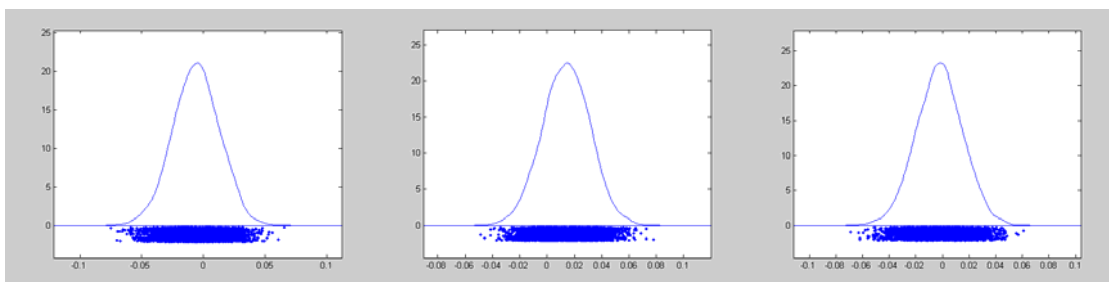
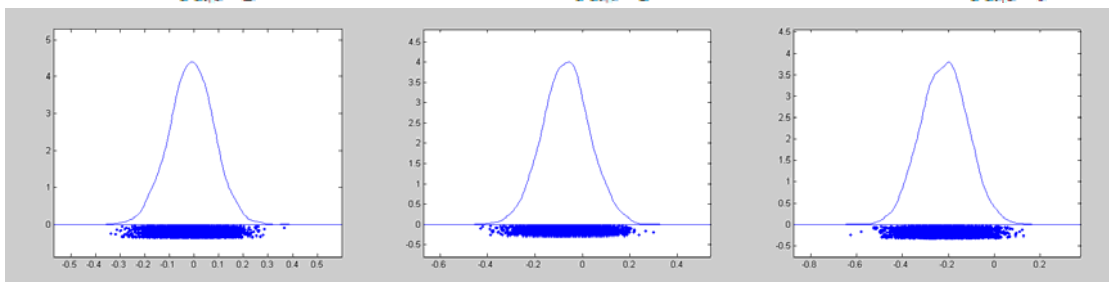
(9) $\beta_{ES,t-1}$

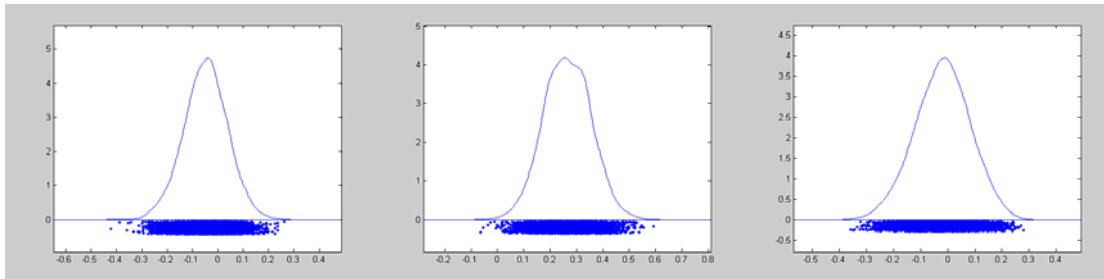
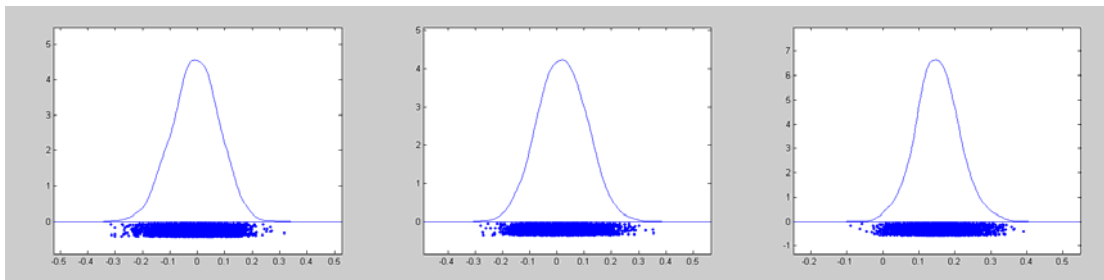


(10) $\beta_{ES,t-2}$

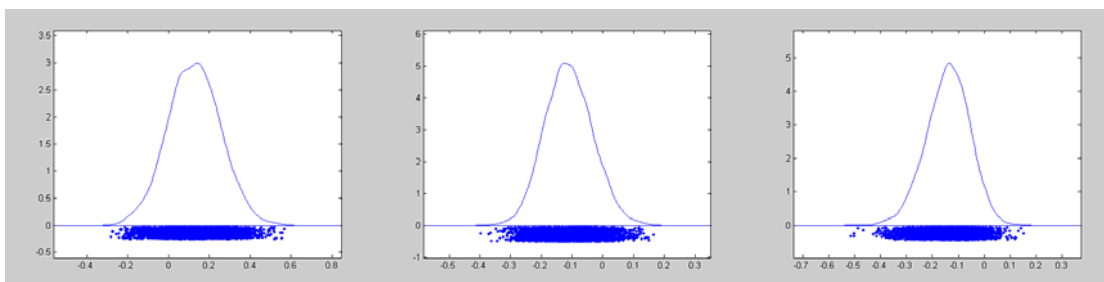
(11) $\beta_{ES,t-3}$

(12) $\beta_{ES,t-4}$

(13) β_{RJT-1} (14) β_{RJT-2} (15) β_{RJT-3} (16) β_{RJT-4} (17) $\beta_{SP,t-1}$ (18) $\beta_{SP,t-2}$ (19) $\beta_{SP,t-3}$ (20) $\beta_{SP,t-4}$ (21) $\beta_{PR,t-1}$ (22) $\beta_{PR,t-2}$ (23) $\beta_{PR,t-3}$ (24) $\beta_{PR,t-4}$ (25) $\beta_{SUC,t-1}$ (26) $\beta_{SUC,t-2}$ (27) $\beta_{SUC,t-3}$

(28) $\beta_{SC,t-1}$ (29) $\beta_{RF,t-1}$ (30) $\beta_{RF,t-2}$ (31) $\beta_{RS,t-3}$ (32) $\beta_{RS,t-4}$

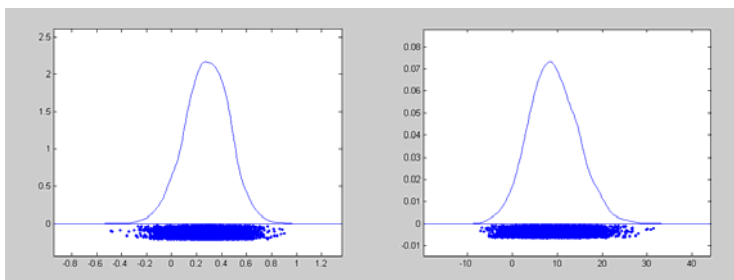
(33) ec term ba



(34) Ec term mg

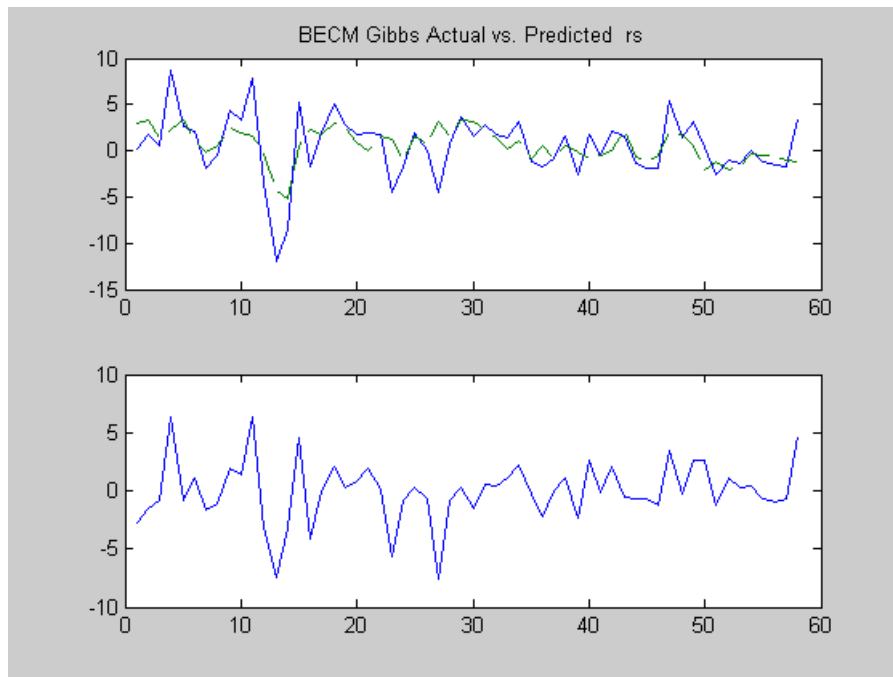
(35) ec term es

(36) ec term rj



(37) ec term sp

(38) c_{RJ}



Apêndice D Tabelas usadas para o cálculo do erro percentual de previsão dentro da amostra no período de setembro/2003 a setembro/2006.

Valores observados das séries

Dates	ba	mg	es	rj	sp	pr	sc	rs
Q4-03	95.82	105.11	106.28	99.21	105.21	108.14	96.48	102.28
Q1-04	106.55	103.13	108.93	99.73	105.9	113.1	97.89	103.74
Q2-04	112.19	105.54	110.8	101.86	108.73	107.92	104.15	106.91
Q3-04	107.27	109.76	110.65	101.7	113.26	121.51	110.42	107.34
Q4-04	111.27	110.68	115.99	102.17	114.22	121.43	107.66	104.77
Q1-05	110.92	111.53	115.46	100.7	113.99	120.65	108.16	103.71
Q2-05	112.62	114.31	112.12	103.51	115.28	118.88	107.58	102.33
Q3-05	116.13	115.64	110.19	103.85	114.86	117.13	102.02	102.39
Q4-05	116.27	115.88	115.97	105.63	116.28	115.38	104.36	101.26
Q1-06	117.8	117.8	116.95	105.64	117.42	112.7	107.02	99.68
Q2-06	118.6	118.38	121.57	105.29	119.04	117.78	105.6	98.02
Q3-06	117.16	119.63	122.16	104.95	119.27	113.45	105.73	101.44

Previsão pelo método de estimação Gibbs Sampling utilizando o modelo VEC Bayesiano com informação *a priori* de Minnesota modificada

Dates	ba	mg	es	rj	sp	pr	sc	rs
Q4-03	103.637	99.198	101.919	98.8	93.794	108.77	99.002	97.665
Q1-04	106.145	98.955	104.192	97.817	94.317	107.499	103.819	98.416
Q2-04	107.344	99.63	102.307	99.432	93.113	108.044	103.711	98.31
Q3-04	108.889	98.984	103.774	99.08	93.043	108.692	105.991	99.094
Q4-04	109.836	99.147	102.721	100.264	92.048	109.657	105.94	99.455
Q1-05	111.688	97.816	106.194	98.795	92.752	110.184	109.81	100.742
Q2-05	111.409	99.986	99.156	103.572	89.529	111.704	104.499	99.895
Q3-05	115.613	94.373	115.305	94.695	94.99	111.213	119.089	103.952
Q4-05	109.912	106.15	79.947	116.267	81.194	115.041	89.508	97.048
Q1-06	125.93	79.2	158.966	70.324	110.115	109.54	157.565	114.689
Q2-06	93.93	138.408	-15.863	174.377	44.336	124.414	9.152	77.705
Q3-06	168.119	6.094	372.898	-54.533	188.614	94.506	340.669	162.224

Medida do erro percentual de previsão para o modelo VEC Bayesiano utilizando informação *a priori* de Minnesota modificada

Dates	ba	mg	es	rj	sp	pr	sc	rs
Q4-03	-8.182	5.632	4.078	0.328	10.93	-0.566	-2.569	4.535
Q1-04	0.355	4.103	4.273	1.861	11.001	4.977	-6.082	5.14
Q2-04	4.262	5.648	7.652	2.278	14.453	-0.109	0.47	8.045
Q3-04	-1.576	9.91	6.143	2.504	17.954	10.551	4.009	7.666
Q4-04	1.196	10.509	11.422	1.749	19.541	9.675	1.652	5.054
Q1-05	-0.777	12.468	7.856	1.877	18.76	8.668	-1.617	2.811
Q2-05	0.923	12.58	11.739	-0.322	22.535	5.957	3.114	2.369
Q3-05	0.386	18.771	-5.293	9.103	17.388	5.093	-17.292	-1.667
Q4-05	5.174	8.074	32.274	-11.055	30.588	0.014	15.561	4.298
Q1-06	-6.759	34.099	-38.917	35.345	5.905	3.219	-50.146	-15.623
Q2-06	20.003	-19.305	119.489	-70.424	64.223	-6.787	98.224	21.808
Q3-06	-42.386	101.031	-220.223	162.651	-60.863	19.103	-237.929	-62.662

Previsão pelo método de estimação Gibbs Sampling utilizando o modelo VEC Bayesiano com informação *a priori* espacial

Dates	ba	mg	es	rj	sp	pr	sc	rs
Q4-03	102.383	180.337	123.42	81.143	113.39	181.91	138.791	134.609
Q1-04	82.657	209.357	152.601	112.422	138.767	146.766	124.615	174.989
Q2-04	72.929	180.599	132.887	128.365	112.284	121.758	114.235	152.031
Q3-04	90.573	192.332	139.129	99.923	128.394	127.26	119.889	147.877
Q4-04	82.615	315.475	152.353	101.26	124.658	214.477	167.82	193.34
Q1-05	47.705	473.389	271.185	177.997	248.196	31.005	140.731	359.484
Q2-05	-2.339	117.642	118.082	351.284	37.984	-124.272	86.136	141.213
Q3-05	136.545	306.532	117.087	-43.024	201.144	126.256	167.472	132.771
Q4-05	29.073	361.809	170.461	228.146	61.346	116.505	146.548	238.102
Q1-06	27.558	1038.347	496.663	243.532	557.297	-28.648	154.244	735.924
Q2-06	3.736	-123.564	50.611	1936.357	-32.437	612.681	2.575	37.927
Q3-06	1210.43	4027.242	573.819	748.805	3080.947	757.294	1046.777	850.369

Medida do erro percentual de previsão para o modelo VEC Bayesiano utilizando informação *a priori* espacial

Dates	ba	mg	es	rj	sp	pr	sc	rs
Q4-03	-6.849	-71.57	-16.127	18.211	-7.775	-68.217	-43.854	-31.609
Q1-04	22.425	-103.003	-40.091	-12.727	-31.036	-29.766	-27.301	-68.681
Q2-04	34.995	-71.119	-19.934	-26.021	-3.269	-12.823	-9.684	-42.204
Q3-04	15.566	-75.23	-25.738	1.747	-13.362	-4.732	-8.575	-37.765
Q4-04	25.753	-185.034	-31.35	0.89	-9.138	-76.626	-55.88	-84.538
Q1-05	56.991	-324.45	-134.873	-76.759	-117.735	74.302	-30.114	-246.624
Q2-05	102.077	-2.915	-5.318	-239.372	67.051	204.535	19.933	-37.998
Q3-05	-17.579	-165.075	-6.259	141.429	-75.121	-7.792	-64.156	-29.672
Q4-05	74.995	-212.227	-46.988	-115.986	47.243	-0.975	-40.425	-135.139
Q1-06	76.606	-781.449	-324.68	-130.53	-374.619	125.42	-44.126	-638.287
Q2-06	96.85	204.379	58.369	-1739.07	127.249	-420.191	97.561	61.307
Q3-06	-933.143	-3266.42	-369.728	-613.487	-2483.17	-567.513	-890.047	-738.298