

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

LEANDRO MANZOLI TROVATI

ANÁLISE BAYESIANA DA DEPENDÊNCIA TEMPORAL DE SÉRIES  
DE AÇÕES NO MERCADO BRASILEIRO

Porto Alegre  
2015

LEANDRO MANZOLI TROVATI

ANÁLISE BAYESIANA DA DEPENDÊNCIA TEMPORAL DE SÉRIES  
DE AÇÕES NO MERCADO BRASILEIRO

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título de Mestre em Economia, com ênfase em Economia Aplicada.

Orientador: Prof. Dr. João Frois Caldeira

Porto Alegre

2015

### CIP - Catalogação na Publicação

Manzoli Trovati, Leandro  
Análise Bayesiana da dependência temporal de  
séries de ações no mercado brasileiro / Leandro  
Manzoli Trovati. -- 2015.  
163 f.

Orientador: João Frois Caldeira.

Dissertação (Mestrado) -- Universidade Federal do  
Rio Grande do Sul, Faculdade de Ciências Econômicas,  
Programa de Pós-Graduação em Economia, Porto Alegre,  
BR-RS, 2015.

1. Dependência temporal. 2. Mercado acionário. 3.  
VAR Bayesiano. 4. GARCH. I. Frois Caldeira, João,  
orient. II. Título.

LEANDRO MANZOLI TROVATI

**ANÁLISE BAYESIANA DA DEPENDÊNCIA TEMPORAL DE SÉRIES  
DE AÇÕES NO MERCADO BRASILEIRO**

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título de Mestre em Economia, com ênfase em Economia Aplicada.

Aprovada em: Porto Alegre, 9 de junho de 2015.

BANCA EXAMINADORA:

---

**Prof. Dr. André Alves Portela Santos**  
UFSC

---

**Prof. Dr. Flávio Augusto Ziegelmann**  
UFRGS

---

**Prof. Dr. Marcelo de Carvalho Griebeler**  
UFRGS

## RESUMO

As pesquisas em finanças tem focado nos últimos anos a modelagem da variância devido à dificuldade de se obter boa previsão dos retornos na média, negligenciando o uso deste último, quanto informação necessária, no estabelecimento de estratégias de alocação de ativos em carteiras de renda variável. Seguindo DeMiguel, Nogales e Uppal (2014), esse trabalho adota o VAR como meio de se obter previsões dos retornos um passo à frente, e então, usá-las na alocação dos ativos. Para a inserção do VAR em finanças, foi permitido que os parâmetros variassem no tempo, o que conseguiu captar com sucesso a dinâmica volátil do mercado de ações. Ainda foram incorporadas técnicas bayesianas de estimação, a fim de driblar a sobreparametrização e obter estimações mais suavizadas, evitando deste modo que a variância das carteiras fossem muito altas. O método teve sucesso na aplicação e mostrou que o uso da previsão um passo à frente para os retornos pode ser usada como uma boa estratégia, expressa nos altos índices de Sharpe encontrados.

**Palavras-chave:** BVAR. Volatilidade. Carteira.

## **ABSTRACT**

Research in finance has focused in recent years in variance modelling due to the difficulty of obtaining good forecasts of mean returns, neglecting the use of this latter on the establishment of asset allocation strategies in the equity portfolios. Following DeMiguel, Nogales e Uppal (2014), this work adopts the VAR as a way of obtaining forecasts of returns one step ahead and to get use of them in the allocation of assets. For insertion of the VAR in finance, the parameters was allowed to vary over time, which successfully captured the volatile dynamics of the stock market. Bayesian estimation techniques was incorporated in order to surmount overparameterization and to get more smoothed estimates, thereby preventing the variance of the portfolio to be very high. The method was successful in this application and showed that the making of one step ahead prediction of returns can be used as a good strategy, which can be expressed in high levels of Sharpe.

**Key-words:** BVAR. Volatility. Portfolio.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Coeficientes IFNC e IEE . . . . .	126
Figura 2 – Coeficientes IFNC e IMAT . . . . .	126
Figura 3 – Coeficientes IFNC e UTIL . . . . .	127
Figura 4 – Coeficientes INDX e IEE . . . . .	127
Figura 5 – Coeficientes IMAT e UTIL . . . . .	128
Figura 6 – Coeficientes dos componentes autorregressivos . . . . .	128
Figura 7 – Coeficientes dos índices IFNC-IEE . . . . .	129
Figura 8 – Coeficientes dos índices INDX-ICON . . . . .	129
Figura 9 – Coeficientes dos índices INDX-UTIL . . . . .	130
Figura 10 – Coeficientes dos índices IEE-ICON . . . . .	130
Figura 11 – Coeficientes dos índices IEE-UTIL . . . . .	131
Figura 12 – Todos os coeficientes fora da diagonal principal . . . . .	131
Figura 13 – Coeficientes dos componentes autorregressivos . . . . .	132
Figura 14 – Coeficientes dos componentes fora da diagonal principal . . . . .	133
Figura 15 – Coeficientes dos componentes autorregressivos . . . . .	135
Figura 16 – Coeficientes dos componentes fora da diagonal principal . . . . .	135
Figura 17 – Carteiras Grande-neutro e Grande-baixo . . . . .	137
Figura 18 – Carteiras Pequeno-neutro e Pequeno-baixo . . . . .	137
Figura 19 – Todos os parâmetros dos componentes autorregressivos das carteiras . . . . .	138
Figura 20 – Carteiras Grande-alto/Pequeno-baixo . . . . .	139
Figura 21 – Carteiras Grande-neutro/Grande-baixo . . . . .	140
Figura 22 – Carteiras Grande-neutro/Pequeno-alto . . . . .	140
Figura 23 – Carteiras Grande-pequeno/Pequeno-neutro . . . . .	141
Figura 24 – Coeficientes extremos . . . . .	141
Figura 25 – Retornos acumulados - Carteiras de arbitragem . . . . .	148
Figura 26 – Retornos acumulados - Retorno acumulado - Carteiras de média-variância com venda a descoberto . . . . .	152
Figura 27 – Retornos acumulados - Retorno acumulado - Carteiras de média-variância sem venda a descoberto . . . . .	153

## LISTA DE TABELAS

Tabela 1	–	Coefficientes da mistura de distribuições normais para a aproximação $\ln \chi_1^2$	. . . .	108
Tabela 2	–	Medidas de desempenho para as carteiras de arbitragem	. . . . .	149
Tabela 3	–	Medidas de desempenho para as carteiras média-variância com venda a descoberto		152
Tabela 4	–	Medidas de desempenho para as carteiras média-variância sem venda a descoberto		154

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>10</b>
<b>2</b>	<b>MÉTODOS EM FINANÇAS</b>	<b>15</b>
2.1	<i>Capital Asset Pricing Model (CAPM) e carteiras de média-variância</i>	15
2.2	Modelos fatoriais	16
2.3	Carteiras de arbitragem	17
2.4	O uso do Vetor Autorregressivo em finanças	18
2.4.1	Interpretação dos coeficientes	19
2.4.2	Construção de carteiras com base no VAR	20
<b>3</b>	<b>DISTRIBUIÇÕES DE PROBABILIDADE</b>	<b>31</b>
<b>3.1</b>	<b>Distribuições univariadas</b>	<b>31</b>
3.1.1	Distribuição normal	31
3.1.2	Distribuição Gama	32
3.1.3	Distribuição qui-quadrado	35
3.1.4	Distribuição gama inversa	37
3.1.5	Distribuição t	39
<b>3.2</b>	<b>Distribuições multivariadas</b>	<b>40</b>
3.2.1	Distribuição normal	41
3.2.2	Distribuição Wishart	51
3.2.3	Distribuição t	52
<b>4</b>	<b>VETOR AUTORREGRESSIVO</b>	<b>55</b>
<b>4.1</b>	<b>Modelo Autorregressivo de Média Móvel</b>	<b>55</b>
4.1.1	Estimação	56
<b>4.2</b>	<b>O Vetor Autorregressivo</b>	<b>57</b>
4.2.1	Estimação	59
<b>4.3</b>	<b>VAR Bayesiano</b>	<b>61</b>
4.3.1	A abordagem Bayesiana	61
4.3.2	<i>Markov Chain Monte Carlo</i>	62
<b>4.4</b>	<b>Escolha das <i>Priori</i></b>	<b>65</b>
4.4.1	Normal-Wishart	65
<b>5</b>	<b>MODELAGEM DA VARIÂNCIA</b>	<b>74</b>
<b>5.1</b>	<b>Modelos univariados</b>	<b>75</b>
5.1.1	GARCH	75
5.1.2	Volatilidade estocástica	77
5.1.3	O modelo	78

5.1.4	Estimação . . . . .	78
<b>5.2</b>	<b>Modelos multivariados . . . . .</b>	<b>82</b>
5.2.1	MGARCH . . . . .	82
5.2.2	MSV . . . . .	89
<b>6</b>	<b>MODELOS ESTADO-ESPAÇO . . . . .</b>	<b>91</b>
<b>6.1</b>	<b>O filtro de Kalman . . . . .</b>	<b>91</b>
6.1.1	Estimação dos parâmetros . . . . .	96
<b>6.2</b>	<b>Suavizador . . . . .</b>	<b>97</b>
<b>6.3</b>	<b>Abordagem Bayesiana . . . . .</b>	<b>99</b>
6.3.1	TVP-VAR . . . . .	100
6.3.2	TVP-VAR com volatilidade estocástica . . . . .	104
<b>7</b>	<b>TÓPICOS ADICIONAIS . . . . .</b>	<b>113</b>
<b>7.1</b>	<b><i>Priori</i> de Minnesota . . . . .</b>	<b>113</b>
<b>7.2</b>	<b>Fatores de desconto . . . . .</b>	<b>116</b>
<b>8</b>	<b>CONSTRUÇÃO DOS MODELOS E INFERÊNCIA . . . . .</b>	<b>121</b>
8.0.1	Metodologia . . . . .	122
<b>8.1</b>	<b>Índices setoriais . . . . .</b>	<b>124</b>
8.1.1	Análise dos componentes autorregressivos . . . . .	125
8.1.2	Análise dos demais componentes . . . . .	128
<b>8.2</b>	<b>Carteiras formadas com base nos fatores de Fama-French . . . . .</b>	<b>131</b>
8.2.1	Carteiras formadas pelo tamanho de mercado . . . . .	132
8.2.2	Análise dos componentes autorregressivos . . . . .	132
8.2.3	Análise dos demais componentes . . . . .	133
8.2.4	Carteiras formadas pelo índice book-to-market . . . . .	134
8.2.5	Análise dos componentes autorregressivos . . . . .	134
8.2.6	Análise dos demais componentes . . . . .	135
8.2.7	Todas as carteiras formadas pelo valor de mercado e índice book-to-market . . . . .	136
8.2.8	Análise dos componentes autorregressivos . . . . .	137
8.2.9	Análise dos demais componentes . . . . .	139
<b>9</b>	<b>CONSTRUÇÃO DAS CARTEIRAS E MEDIDAS DE DESEMPENHO</b>	<b>142</b>
<b>9.1</b>	<b>Metodologia . . . . .</b>	<b>142</b>
9.1.1	Medidas de desempenho . . . . .	145
<b>9.2</b>	<b>Carteiras de arbitragem . . . . .</b>	<b>146</b>
9.2.1	Resultados . . . . .	147
<b>9.3</b>	<b>Carteiras baseadas na média-variância . . . . .</b>	<b>149</b>
9.3.1	Resultados . . . . .	151

**10      CONSIDERAÇÕES FINAIS . . . . . 155**

**Referências . . . . . 157**

## 1 INTRODUÇÃO

A moderna teoria do portfólio tem como uma de suas raízes o trabalho de Markowitz (1952), o qual é responsável por dar base teórica aos estudos em finanças através de um modelo que incorpora um *trade-off* entre risco e retorno. Sob essa ótica, a natureza do modelo e todas as pesquisas que nele se baseiam podem ser vistas como uma questão de média e variância. Esses modelos, também conhecidos como de média-variância, como o próprio nome propõe, esbarram em problemas de ordem estatística na busca de estimar os dois primeiros momentos com a maior acurácia possível. Dentro desse contexto, então, surgiram diversos modelos econométricos.

A literatura, em geral, evita trabalhar com a média dos retornos, sob a justificativa de que há dificuldade de encontrar um padrão que justifique o uso desta para fins de obter previsões. Deste modo, grande parte da literatura especializou-se na modelagem econométrica e construção de estratégias de investimento das séries de retornos no que tange à sua variância, popularmente tida como uma medida de risco, *e.g.* temos o caso das carteiras de variância mínima, amplamente utilizada em trabalhos na área. Estas pesquisas então partem do pressuposto que a média dos retornos incondicional é 0 e invariante no tempo, a qual pode ser ignorada sem muita perda de informação, buscando-se desta forma simplificar os modelos.

As bases teóricas que sustentam a hipótese de média 0 podem ser encontradas na Hipótese dos Mercados Eficientes (HEM). Se um mercado é eficiente e as informações que nele transcorrem são completas (não há assimetria informacional), a HEM diz que os agentes com expectativas racionais, operantes no mercado de ativos, são incapazes de obterem lucro econômico (lucro anormal) com base na disponibilidade de informações. Como corroborado pelo trabalho de Jensen (1978), um mercado é considerado eficiente se os preços dos ativos nele negociados refletem o consenso geral sobre a economia, o mercado financeiro e a empresa específica envolvida. Com base nessas premissas, a expectativa é que o retorno deve aproximar-se de 0 na média, logo não há perdas relevantes de informação ao ignorá-la no modelo. Sob essa justificativa, os modelos de variância mínima, que pressupõem um investidor totalmente avesso ao risco, ganharam destaque nas pesquisas ao longo das últimas décadas.

Dentre os modelos econométricos que visam obter estimativas para a variância, por vezes chamada de volatilidade, encontram-se os da família *Generalized Autoregressive Conditional Heteroscedasticity* (GARCH) e *Stochastic Volatility* (SV), cujo desenvolvimento ocorre ao longo das décadas de 80 e 90 e pode ser atribuído à diversos autores. O uso destes tem mostrado grande êxito no meio acadêmico ao longo das últimas décadas, tanto quanto pela aplicação prática destes no mercado. Esses métodos baseiam-se na incorporação de um dos fato estilizado também conhecido como agrupamento da vo-

latilidade. Em outras palavras, a variância dos retornos apresenta correlação serial no tempo, períodos de grande instabilidade nos retornos tendem a ser seguidos por períodos de grande instabilidade, ao passo que períodos mais estáveis tendem a ser seguidos por períodos de baixa volatilidade.

Motivados pela dificuldade de modelar a média das séries de retornos financeiros, podem ser encontrados trabalhos que especializaram-se em identificar fatores comuns que promovem a variabilidade desses tipos de dados a apresentar determinados co-movimentos, ou seja, padrões comuns. Dentro desta área, também chamada de análise fatorial, o popular modelo unifatorial CAPM (*Capital Asset Pricing Model*) de Sharpe (1964) e Lintner (1965) modelam os retornos de acordo com os movimentos do mercado. O proeminente trabalho de Fama e French (1996) estende as possibilidades de inclusão de novos fatores ao CAPM, desde então, a literatura ganhou força na busca de mais fatores comuns aos retornos de ativos. Dentre esses modelos, vale citar Carhart (1997), o autor propôs que ativos cujos retornos passados apresentaram valores positivos, tendem a ter uma performance melhor no futuro, tal como aqueles que apresentaram movimentos de baixa no mercado, tendem a permanecer assim por um tempo. Ele chamou esse fator de momento, em outras palavras, pode-se dizer que há uma persistência temporal na média dessas séries. Recentemente, DeMiguel, Nogales e Uppal (2014) demonstrou que há persistência temporal na média condicionada dos retornos à informação passada contida nelas mesmas. Para tal, ele fez uso do já tradicional modelo *Vector Autoregression* (VAR) nestas séries. O autor ainda modificou alguns modelos de apreçamento de ativos já consagrados afim de incorporar o VAR, e os resultados foram satisfatórios.

O uso do VAR em economia pode ser atribuído ao trabalho de Sims (1980), cuja motivação inicial é a modelagem e previsão em dados macroeconômicos. O grande trunfo no uso desta metodologia decorre da sua simplicidade e generalidade. A grande quantidade de parâmetros no VAR pode ser considerada a fonte da habilidade deste de se ajustar aos dados, e explica sua popularidade na literatura em macroeconomia. Por outro lado, não raramente os economistas empenham-se em incorporar muitas variáveis no modelo para que este explique melhor o processo gerador dos dados, tal como para fins de previsão. Então, essa parametrização passa a ser considerada uma desvantagem no que tange ao número de coeficientes, os quais crescem muito à medida que a dimensionalidade aumenta. A imprecisão da inferência estatística e incertezas a respeito das previsões, ocasionadas pelo excesso de parâmetros do modelo, leva a abordagem clássica a perder espaço frente à alternativa: a inferência Bayesiana.

Os métodos Bayesianos ganharam força devido ao desenvolvimento de novas técnicas mais eficientes de tratar as distribuições *posteriori*. Não havendo mais a necessidade de derivar uma forma analítica para essas distribuições, a modelagem Bayesiana ganha um carácter muito mais flexível no que tange à distribuição dos dados e crenças a respeito do comportamento dos parâmetros. Sua estimação é feita através de métodos de reamos-

tagem, uma classe de estimadores chamada de *Monte Carlo Markov Chain* (MCMC).

Os métodos baseados em MCMC baseiam-se na construção de uma cadeia de Markov para os parâmetros, cuja distribuição *posteriori* é tida como única e estacionária. Para construir a *posteriori*, parte-se do princípio de que é possível gerar valores aleatórios de acordo com a distribuição condicional dos parâmetros. Se a forma da *posteriori* marginal é sabida, poder-se-ia simular os valores desta distribuição a partir do valor inicial dos parâmetros, também chamada de *priori*, mas deste modo, não haveria vantagens no uso dos métodos em MCMC, visto que sua principal motivação deixaria de existir (porém não ocorre grande em grande parte das modelagens econométricas na área). Geralmente, o valor inicial dos parâmetro parte de um ponto arbitrário no espaço paramétrico e a velocidade de convergência à distribuição estacionária *posteriori* é um aspecto a ser olhado com atenção pelo pesquisador.

Uma das distribuições *priori* mais consagradas nesta abordagem é a de Minnesota, a qual foi originalmente concebida por Litterman (1979) como forma de reduzir a quantidade de parâmetros no VAR através da imposição de restrições na distribuição *priori*. A ideia central é que o VAR pode ser encolhido para uma forma em que cada variável possa ser explicada por um processo univariado.

Dentre os mais recentes trabalhos na área de VAR Bayesiano, vale citar Cogley e Sargent (2005) e Primiceri (2005) construíram um TVP-VAR com volatilidade estocástica (TVPSV-VAR). Esses últimos modelos VAR Bayesiano (BVAR), apesar de serem concebidos para dados macro, mostram-se pertinentes, pelas suas propriedades, à aplicação em dados financeiros, ajustando-se bem quanto à recente abordagem proposta por DeMiguel, Nogales e Uppal (2014). Essas propriedades referem-se à heteroscedasticidade dos erros do modelo, as quais são auto regressivas condicionadas pelo modelo SV, e pela modelagem da média, a qual acontece conjuntamente.

Tendo em vista toda a evolução dos modelos em finanças apresentada e seu panorama atual, tal como a evolução dos métodos econométricos disponíveis, indispensáveis à área no que tange à montagem de estratégias de investimento, então este trabalho tem um objetivo bifocal. O primeiro deles é dar uma base teórica econométrica para os BVAR, motivando o seu uso em finanças quando possível. O segundo foco então encontra-se na reestruturação das estratégias de investimento existentes de forma a incorporar a informação contida no BVAR.

Para o segundo foco do trabalho, foi escolhida a abordagem apresentada em DeMiguel, Nogales e Uppal (2014), sua simplicidade e incorporação do VAR à montagem de estratégias em finanças já consagradas ocorre de forma natural e intuitiva. O primeiro foco citado então entra de forma direta através da escolha de métodos de estimação mais modernos, motivando seu uso não apenas em macroeconomia, mas também em finanças.

Além das vantagens do uso da abordagem Bayesiana já citadas, é motivado o uso

de parâmetros variantes no tempo (TVP-VAR), tal como em Primiceri (2005). A flexibilidade do método em adaptar-se aos dados mostra-se atrativa, sobretudo para o mercado financeiro. Mudanças no cenário econômico, tal como da divulgação de informações de empresas, as quais ocorrem a todo instante, o torna instável. Ademais, métodos do tipo TVP-VAR Bayesianos já fornecem estimativas suavizadas de modo natural, adequadas para a inferência e avaliação de mercado.

Das desvantagens do uso de métodos Bayesianos TVP-VAR, pode-se citar os altos custos computacionais envolvidos. Quando do aumento do número de parâmetros, a estimação pode se tornar infactível, principalmente sobre o ponto de vista de que um analista de mercado estará interessado em atualizar toda a série de parâmetros da chegada de novas informações diariamente. Para montar carteiras grandes e avaliar estratégias de investimento que envolvam muitos ativos, ou seja, alta dimensão do modelo, foi escolhida a abordagem encontrada em Koop e Korobilis (2013), a qual visa a redução dos custos computacionais via fator de desconto.

A estrutura de trabalho segue da seguinte forma, a dissertação foi dividida em duas partes e os capítulos realocados nelas de acordo com o seu perfil. A primeira parte trata da teoria em finanças e métodos Bayesianos para VAR. A segunda então trata da aplicação empírica dos modelos apresentados, visando avaliar a capacidade desses métodos em ajustar-se aos dados de retornos financeiros em séries de ações negociadas na BM&FBovespa.

O primeiro capítulo introduz os modelos em finanças no que tange à estratégias de investimento e suas bases na moderna teoria em finanças. O segundo capítulo, considerado mais técnico, introduz as distribuições de probabilidade usadas nesse trabalho, e necessárias à compreensão básica dos métodos Bayesianos aqui utilizados. O terceiro capítulo introduz os modelos VAR, responsáveis pela modelagem da média condicional, necessárias à previsão um passo à frente, as quais entram na construção das carteiras, tão bem quanto na avaliação do perfil temporal apresentado pelo mercado financeiro. O quarto capítulo aborda os métodos de modelagem da variância, focando nos grupos GARCH e SV, igualmente necessários às estratégias de investimento. O quinto capítulo introduz a variação dos parâmetros dos modelos no tempo via modelos estado-espço e filtro de kalman, e motiva seu uso na abordagem Bayesiana. O sexto capítulo trata de alguns tópicos extras para completar os modelos, como a metodologia proposta por Koop e Korobilis (2013) e a não menos importante *priori* de Minnesota. O sexto capítulo aplica então os métodos Bayesianos ao caso do mercado acionário brasileiro, fazendo uma avaliação dos resultados encontrados via gráficos. Por último, o sétimo capítulo monta estratégias de investimento com base nas informações contidas no BVAR.

Para extrair as informações contidas neste trabalho de forma adequada, espera-se que o leitor tenha alguma familiaridade com conceitos matemáticos e estatísticos, tão

bem quanto econometria de séries temporais. Ademais espera-se de forma igual alguma familiaridade com a teoria em finanças, promovendo a correta motivação no uso dos métodos apresentados.

## 2 MÉTODOS EM FINANÇAS

Este capítulo aborda alguns métodos em finanças relativos a montagem de carteiras. Para tal, é feita uma breve dissertação a respeito da moderna teoria em finanças e suas implicações, de modo que é dada uma base teórica que sustente as estratégias de investimento. A segunda parte do capítulo trata da abordagem feita em DeMiguel, Nogales e Uppal (2014) sobre a incorporação da dependência temporal nos retornos de ativos na estruturação dessas carteiras, de forma que a informação disponível no VAR, modelo econométrico capaz de captar essa dependência temporal, possa ser utilizada de forma ótima. Referências adicionais podem ser encontradas em Campbell et al. (1997).

### 2.1 *Capital Asset Pricing Model (CAPM) e carteiras de média-variância*

A moderna teoria em finanças gira em torno do problema da quantificação do *tradeoff* existente entre risco e retorno, o qual o CAPM dá base teórica para que essa quantificação seja feita.

O trabalho seminal na área é atribuído à Markowitz (1952), cujo trabalho empenha-se na construção de uma carteira que leve em consideração o retorno esperado e sua variância. Desta forma, o investidor estará interessado em alocar de forma ótima seus recursos de forma a maximizar o retorno esperado tolerando um certo nível de variabilidade neste, ou em outros termos, o retorno é maximizado dado um nível de risco aceitável para ele.

Sharpe (1964) e Lintner (1965) mostram que se um investidor tem expectativas homogêneas e mostram preferências na alocação do portfólio do tipo Markowitz, então todos os investimentos conjuntamente, ou também chamado de carteira de mercado, será uma carteira do tipo média-variância. Deste modo, qualquer retorno pode ser explicado pelo retorno da carteira de mercado, e nasce o CAPM:

$$E(y_i) = y_f + \beta_i(E(y_m) - y_f)$$

Onde  $y_i$  representa o retorno do ativo ou alguma carteira montada pelo investidor,  $y_f$  representa o ativo livre de risco e  $y_m$ , o retorno da carteira de mercado. Os testes empíricos desse modelo focam em três implicações, a primeira diz que o termo de intercepto é zero, a segunda diz que o  $\beta_i$  captura completamente a variação do excesso de retorno, e por último, que o valor esperado do prêmio de risco é positivo.

Retornando ao caso das estratégias de investimento, a alocação ótima do portfólio de um investidor então ganha bases teóricas fortes que sustentam o ganho obtido pela diversificação em termos de risco diluído. Sob outro ponto de vista, o que ocorre na prática é minimizar o risco esperado da carteira, uma vez que prever os retornos nem sempre é uma tarefa fácil. Na forma algébrica, a alocação é dada como:

$$\begin{aligned} \min_w \mathbf{w}'\Sigma\mathbf{w} \\ s.t. \mathbf{w}'\boldsymbol{\mu} = \mu_c \\ \mathbf{w}'\mathbf{1} = 1 \end{aligned}$$

Fica claro então que o risco da carteira, dado pelo termo  $\min_w \mathbf{w}'\Sigma\mathbf{w}$  é minimizado. A primeira das restrições envolve estabelecer um retorno esperado para a carteira, dado pelo termo  $\mu_c$ . Uma vez que essa condição não entra na restrição, a carteira é chamada de variância mínima, um *benchmark* na literatura na área. A segunda restrição diz que 100% dos recursos serão alocados no portfólio.

Voltamos a atenção para à carteira de mínima variância, suas propriedades podem ser vistas de forma mais agradável quando um investidor tenta alocar seus recursos entre dois ativos, dessa forma:

$$\begin{aligned} \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\ = w_1^2\sigma_{11} + 2w_1w_2\sigma_{12} + w_2^2\sigma_{22} \end{aligned}$$

Onde foi usado o fato que a matriz de covariâncias é simétrica. Os termos da variância sempre serão positivos, então essa estratégia atribuirá menor peso àqueles ativos que apresentarem maior  $\sigma_{ii}$ . Já para o termo de covariâncias, se positivo, então essa estratégia tenderá a dar pesos mais negativos para um dos ativos. Caso ambos os ativos sejam positivamente correlacionados, então os pesos de ambos serão positivos.

## 2.2 Modelos fatoriais

Há amplas evidências na literatura de que o  $\beta$  do CAPM, por si só, não é capaz de explicar todas as variações no retorno esperado dos ativos. Deste modo, abre-se caminho à incorporação de fatores adicionais capazes de explicar o comportamento do retorno esperado.

Um dos modelos que surgiram na busca de melhores resultados foi o *Arbitrage Pricing Theory* (APT), de Ross (1976). O APT é um modelo mais genérico que o CAPM e permite a incorporação de múltiplos fatores. Sua generalidade ainda reside no fato de que o método não restringe a escolha dos fatores. De modo geral, ele é dado como:

$$y_i = a_i + \mathbf{b}'_i\mathbf{f} + e_i \qquad E(e_i|\mathbf{f}) = 0 \quad E(e_i^2) = \sigma_i^2$$

Onde  $f$  são os fatores utilizados no modelo.

Um modelo fatorial, já amplamente aceito, é o proposto por Fama e French (1996), o qual estende o tradicional CAPM em duas direções. A primeira, então, visa acrescentar um fator que capte o tamanho de mercado, dado pela diferença de retornos das empresas pequenas pelos os retornos das grandes. Uma vez que as empresas menores são menos consolidadas no mercado, geralmente seus papéis trazem um risco maior, logo um retorno maior é esperado. A segunda direção do modelo trata do índice *book-to-market*, o qual é responsável por dizer em que grau uma determinada empresa está valorizada no mercado. Ele é dado pela razão entre o valor contábil da empresa dividido pelo seu valor de mercado, geralmente dado pelo número de ações que circulam vezes o seu valor. Quando maior esse índice, sinal de que a empresa está subvalorizada, então seu preço tende a subir no mercado, ao passo que para o contrário, a empresa já estará valorizada, e seus papéis não são de crescimento. Esse fator, então, busca captar a diferença de retorno entre aquelas que possuem um alto índice *book-to-market* por aquelas que apresentam um baixo índice.

Na forma algébrica, o modelo é escrito como:

$$y_i = \alpha_i + \beta_i^m (y_m - y_f) + \beta_i^{SMB} SMB + \beta_i^{HML} HML + e_i$$

Onde *SMB* é o fator do tamanho de mercado, ou *Small minus Big*, e *HML* é o fator de crescimento de mercado, ou *High minus low*.

### 2.3 Carteiras de arbitragem

Carteiras de arbitragem são comumente chamadas de carteiras de custo zero. Sua estratégia baseia-se em entrar comprado no mesmo montante que se entra vendido, ou seja, a soma dos pesos dos ativos nesta carteira somam zero.

Duas carteiras do tipo arbitragem são consideradas aqui, a primeira delas segue a proposta em Lo e MacKinlay (1990). O autor justifica que se há uma relação, mais conhecida do inglês como *lead-lag*, entre as séries de retornos, ou seja, os valores apresentados por uma podem explicar os retornos da outra, então essa informação deve ser utilizada. Os autores então justificam a montagem de uma carteira que vende aqueles ativos que mostraram um bom desempenho e compram aqueles que não o fizeram, de forma que possa ser capaz de obter retornos esperados positivos. Essa estratégia baseia-se no pressuposto de que os mercados apresentam autocorrelação negativa. O autor ainda justifica que os retornos semanais do portfólio são muito auto correlacionados, resultado das autocorrelações entre os mercados. De modo compacto então, a estratégia da carteira proposta pode ser escrita como:

$$\mathbf{w}_{t+1}^c = -\frac{1}{n}(\mathbf{y}_t - \bar{y}_t \mathbf{1}) \quad (2.1)$$

$$\bar{y}_t = \frac{\mathbf{1}'\mathbf{y}_t}{n} \quad (2.2)$$

Onde  $\mathbf{1}$  é um vetor  $n$ -dimensional de 1 e  $\bar{y}_t$  é a média de todos os retornos individuais em  $t$ . Como visto, os retornos deste portfólio somam zero, ou em termos práticos, o volume de ativos que se entra na forma vendida, também o é na forma comprada. A interpretação dos pesos dessa estratégia então é direta, se um determinado ativo apresenta retorno positivo em  $t$ , então a carteira de arbitragem contrária garantirá uma posição vendida, ou negativa para este em  $t + 1$ .

A segunda estratégia considerada segue a proposta de Pan, Liano e Huang (2004), também chamada de momento, obtida similarmente à carteira de arbitragem contrária, porém com o sinal oposto:

$$\mathbf{w}_{t+1}^m = \frac{1}{n}(\mathbf{y}_t - \bar{y}_t \mathbf{1}) \quad (2.3)$$

$$\bar{y}_t = \frac{\mathbf{1}'\mathbf{y}_t}{n} \quad (2.4)$$

A diferença crucial dessas estratégias pode ser explicada pela relação entre os retornos dos ativos. Para a carteira contrária, o retorno esperado será positivo se a auto correlação for negativa, uma vez que, se um determinado ativo  $i$  apresenta bom desempenho em  $t$ , é esperado uma reversão nesse desempenho em  $t + 1$ . Por outro lado, se um outro determinado ativo, chamamos de  $j$ , apresentar retorno positivo em  $t$ , e ele for positivamente correlacionado com  $i$ , então essa estratégia também proporcionará retornos positivos.

Uma vez que há a possibilidade de não haver o cenário descrito, a carteira momento trabalha de forma contrária, esta considera que há alguma persistência temporal nos retornos de  $t$  para  $t + 1$  quando assume um sinal contrário à carteira de arbitragem contrária.

## 2.4 O uso do Vetor Autorregressivo em finanças

Há evidências empíricas sobre a dependência temporal nas séries de retornos no mercado acionário, contudo essas evidências não são consenso geral, tal como a volatilidade auto condicionada, e é negligenciada por grande parte dos autores.

Essa seção tem como base o trabalho de DeMiguel, Nogales e Uppal (2014) que se propõem a elaborar formas de acrescentar a dependência temporal dessas séries através do VAR, modelo escolhido pela sua característica de ser amplo quanto a modelagem das

relações, pressupostamente lineares, entre os retornos e seus valores passados. O modelo VAR aplicado neste trabalho é explorado num maior nível de detalhamento no capítulo Vetor Autorregressivo.

Os autores partem da premissa que o modelo VAR com um grau de defasagem é suficiente para captar as relações necessárias para a montagem das carteiras, o que torna a estimação dos parâmetros parcimoniosa do ponto de vista estatístico.

Seguindo seu trabalho, a previsão dos retornos é sempre feita um passo à frente, não havendo necessidade de previsão  $h > 1$  passos à frente, uma vez que as carteiras são rebalanceadas diariamente. De forma compacta, a previsão feita pelo VAR é feita como:

$$\mathbf{y}_{t+1} = \mathbf{a} + \mathbf{B}\mathbf{y}_t + \mathbf{e}_{t+1}$$

Onde  $\mathbf{y}_t \in \mathbb{R}^n$  é vetor de retornos para o período  $t$ ,  $\mathbf{a} \in \mathbb{R}^n$  é um vetor de constantes,  $\mathbf{B} \in \mathbb{R}^{n \times n}$  é a matriz de coeficientes angulares e  $\mathbf{e}_t \in \mathbb{R}^n$  é o vetor de erros, responsável pela aleatoriedade no modelo.

Esse modelo diz que a predição de um dado retorno de ativo amanhã é uma função seu valor assumido hoje e também dos valores assumidos pelos retornos de outros ativos hoje. todas as relações são lineares e dada pela matriz de coeficientes  $\mathbf{B}$ .

$$y_{it+1} = a_i + B_{ii}y_{it} + \sum_{j=1, j \neq i}^n B_{ij}y_{jt} + e_{it+1}$$

Onde  $B_{ij}$  representa o  $i$ -ésimo elemento da  $j$ -ésima coluna, os quais assumem os valores  $i, j = 1, \dots, n$ .

Em outras palavras, o VAR é capaz de dizer o quão bem os retornos de hoje podem prever os retornos de amanhã e, com base nessa informação, montar carteiras que explorem essa dependência temporal, a fim de que o investidor tenha um portfólio que apresente um desempenho melhor fora da amostra.

#### 2.4.1 Interpretação dos coeficientes

Para facilitar a exposição, suponha que seja montada uma carteira com dois ativos, chamados aqui de ativo 1 e ativo 2. Suponha ainda que seus respectivos retornos diários sejam dados pelo vetor  $\mathbf{y}_t$ . Então, ao aplicar o modelo VAR sobre essas séries, somos capazes de avaliar quaisquer relações lineares existentes entre esses dois ativos temporalmente. De forma compacta temos:

$$\hat{\mathbf{y}}_{t+1} = \hat{\mathbf{a}} + \hat{\mathbf{B}}\mathbf{y}_t$$

O símbolo sobre as variáveis e coeficientes ( $\hat{\cdot}$ ) aqui quer dizer que, uma vez que o termo de erro aleatório foi suprimido, estamos trabalhando com os valores preditos, obtidos pela estimação dos parâmetros  $\mathbf{a}$  e  $\mathbf{B}$ . A expressão acima é melhor compreendida se aberta da seguinte forma:

$$\begin{bmatrix} \mathbf{y}_{(1)\hat{t}+1} \\ \mathbf{y}_{(2)\hat{t}+1} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{(1)\hat{t}} \\ \mathbf{y}_{(2)\hat{t}} \end{bmatrix}$$

Ou então:

$$\mathbf{y}_{(1)\hat{t}+1} = a_1 + \beta_{11}\mathbf{y}_{(1)\hat{t}} + \beta_{12}\mathbf{y}_{(2)\hat{t}} \quad = a_2 + \beta_{21}\mathbf{y}_{(1)\hat{t}} + \beta_{22}\mathbf{y}_{(2)\hat{t}}$$

Percebe-se então que todo termo fora da diagonal principal  $\beta_{ij}$  para todo  $i \neq j$  em  $i, j = 1, \dots, n$  relaciona o retorno de um dado ativo aos valores assumidos pelos demais ativos hoje, ao passo que todos os parâmetros da diagonal principal  $\beta_{ii}$ , para todo  $i = 1, \dots, n$  relaciona os retornos de amanhã com os de hoje para o mesmo ativo.

Se há uma persistência temporal para esses retornos, ou seja, ativos que mostraram um bom desempenho recentemente tendem a apresentar também um bom desempenho num futuro próximo, então esses parâmetros tem que ser positivos e estatisticamente significantes. Intuitivamente, geralmente os parâmetros fora da diagonal tendem a ser menores em módulo, não sendo uma regra. Eles podem apresentar valores negativos também, ou seja, se o ativo  $i$  teve um bom desempenho hoje, pode acarretar um mau desempenho de outro ativo  $j$  amanhã.

Voltemos ao nosso exemplo acima, onde a carteira é composta por apenas dois ativos. Se  $\beta_{12}$  e  $\beta_{21}$  são ambos positivos e ainda, se  $\beta_{12} > \beta_{21}$ , há um indício de que o ativo 2 lidera o ativo 1. Essa relação de liderança está intrinsecamente relacionada à precedência temporal de 1 sobre o 2. Podemos dizer, em outras palavras, que 2 causa 1, ou ainda que este tem uma influência maior sobre o 1 do que vice-versa. Essa análise é importante ao montar as carteiras, uma vez que tendo o resultado de 2 em mãos, podemos esperar com maior certeza o resultado que teremos em 1 amanhã, porém o contrário não pode ser feito com tamanha precisão.

## 2.4.2 Construção de carteiras com base no VAR

Nas próximas seções serão apresentados duas estratégias de alocação de carteiras baseadas no modelo VAR.

### 2.4.2.1 Carteira de arbitragem com VAR

Essa carteira é similar às já propostas carteiras de arbitragem contrária e de momento e surge como alternativa direta à elas ao incorporar a informação contida no modelo

VAR para os retornos. Algebricamente:

$$\mathbf{w}_{t+1}^{VAR} = \frac{1}{n}(\hat{\mathbf{y}}_{t+1} - \bar{y}_t^{VAR} \mathbf{1}) \quad (2.5)$$

$$= \frac{1}{n}(\mathbf{a} + \mathbf{B}\mathbf{y}_t - \bar{y}_t^{VAR} \mathbf{1}) \quad (2.6)$$

Sendo  $\mathbf{1}$  um vetor  $n$ -dimensional de 1, então  $\bar{y}_t^{VAR} = \frac{1}{n} \hat{\mathbf{y}}_{t+1}^{VAR'} \mathbf{1} = \frac{1}{n} (\mathbf{a} + \mathbf{B}\mathbf{y}_t)' \mathbf{1}$  é a média dos valores preditos de  $\mathbf{y}$  em  $t+1$ , em outras palavras,  $\bar{y}_t$  é a predição dos retornos de uma carteira ingênua baseada no modelo VAR.

Note ainda que os pesos de cada ativo na carteira somam zero. Outro ponto a ser notado é que  $w_{i(t+1)}^{VAR}$  só terá um valor positivo para todo ativo  $i$  que apresentar um valor esperado acima da média, e negativo caso contrário.

Abrindo a expressão 2.6 chegamos na expressão:

$$\begin{bmatrix} \mathbf{w}_{(1)t+1} \\ \mathbf{w}_{(2)t+1} \\ \vdots \\ \mathbf{w}_{(n)t+1} \end{bmatrix}^{VAR} = \frac{1}{n} \begin{bmatrix} \hat{y}_{(1)t+1} - \bar{y}_t \\ \hat{y}_{(2)t+1} - \bar{y}_t \\ \vdots \\ \hat{y}_{(n)t+1} - \bar{y}_t \end{bmatrix}^{VAR}$$

### 1ª proposição

Assumindo que  $\mathbf{y}_t$  é um processo estacionário na matriz de covariâncias, cujo vetor de médias, matriz de covariâncias e matriz de covariâncias cruzadas são dados respectivamente pelas expressões:

$$\boldsymbol{\mu} = E[\mathbf{y}_t]$$

$$\boldsymbol{\Gamma}_0 = E[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_t - \boldsymbol{\mu})']$$

$$\boldsymbol{\Gamma} = E[(\mathbf{y}_{t-1} - \boldsymbol{\mu})(\mathbf{y}_t - \boldsymbol{\mu})']$$

Onde  $\boldsymbol{\Gamma}_0$  é positiva definida e  $\mathbf{B} = \boldsymbol{\Gamma}'_1 \boldsymbol{\Gamma}_0^{-1}$  e  $\mathbf{A} = \boldsymbol{\mu} - \mathbf{B}\boldsymbol{\mu} = (\mathbf{I} - \mathbf{B})\boldsymbol{\mu}$ , em que  $\mathbf{I}$  é uma matriz identidade em  $\mathbb{R}^{n \times n}$ .

- a) A carteira de arbitragem com VAR pode ser decomposta na soma de duas outras carteiras, uma relacionada à dependência temporal capturada pela matriz  $\mathbf{B}$ , chamamos esse portfólio de  $\mathbf{w}_{(\Gamma)t}$ , e outra carteira relacionada à covariância cruzada na média dos retornos, chamamos esse portfólio de  $\mathbf{w}_{(\mu)t}$ , algebricamente a relação pode ser escrita como:

$$\mathbf{w}_t^{VAR} = \mathbf{w}_{(\Gamma)t} + \mathbf{w}_{(\mu)t}$$

Onde

$$\mathbf{w}_{(\Gamma)t} = \frac{1}{n} \left( \mathbf{B}(\mathbf{y}_{t-1} - \boldsymbol{\mu}) - \frac{\mathbf{1}'\mathbf{B}(\mathbf{y}_{t-1} - \boldsymbol{\mu})\mathbf{1}}{n} \right) \quad \mathbf{w}_{(\mu)t} = \frac{1}{n} \left( \boldsymbol{\mu} - \frac{\mathbf{1}'\boldsymbol{\mu}\mathbf{1}}{n} \right)$$

PROVA

Da definição para os pesos da carteira de arbitragem com VAR, tem-se que:

$$\mathbf{w}_t^{VAR} = \frac{1}{n} (\mathbf{a} + \mathbf{B}\mathbf{y}_t - \mathbf{1}'(\mathbf{a} + \mathbf{B}\mathbf{y}_{t-1})\mathbf{1})$$

Usando-se do fato que  $\mathbf{a} = (\mathbf{I} - \mathbf{B})\boldsymbol{\mu}$ , a substituição pode ser feita:

$$\mathbf{w}_t^{VAR} = \frac{1}{n} ((\mathbf{I} - \mathbf{B})\boldsymbol{\mu} + \mathbf{B}\mathbf{y}_t - \mathbf{1}'((\mathbf{I} - \mathbf{B})\boldsymbol{\mu} + \mathbf{B}\mathbf{y}_{t-1})\mathbf{1})$$

Isolando os termos que contém  $\mathbf{B}$ :

$$\begin{aligned} \mathbf{w}_t^{VAR} &= \frac{1}{n} \left( \mathbf{B}(\mathbf{y}_{t-1} - \boldsymbol{\mu}) - \frac{\mathbf{1}'\mathbf{B}(\mathbf{y}_{t-1} - \boldsymbol{\mu})\mathbf{1}}{n} + \boldsymbol{\mu} - \frac{\mathbf{1}'\boldsymbol{\mu}\mathbf{1}}{n} \right) \\ &= \frac{1}{n} \left( \mathbf{B}(\mathbf{y}_{t-1} - \boldsymbol{\mu}) - \frac{\mathbf{1}'\mathbf{B}(\mathbf{y}_{t-1} - \boldsymbol{\mu})\mathbf{1}}{n} \right) + \frac{1}{n} \left( \boldsymbol{\mu} - \frac{\mathbf{1}'\boldsymbol{\mu}\mathbf{1}}{n} \right) \\ &= \mathbf{w}_{(\Gamma)t} + \mathbf{w}_{(\mu)t} \end{aligned}$$

b) O valor esperado da carteira pode ser escrita como:

$$E[\mathbf{w}_t^{VAR'} \mathbf{y}_{t-1}] = E[(\mathbf{w}_{(\Gamma)t} + \mathbf{w}_{(\mu)t})' \mathbf{y}_t] = E[\mathbf{w}'_{(\Gamma)t} \mathbf{y}_t] + E[\mathbf{w}'_{(\mu)t} \mathbf{y}_t] \quad (2.7)$$

Onde

$$E[\mathbf{w}'_{(\Gamma)t} \mathbf{y}_t] = \frac{\text{tr}(\boldsymbol{\Gamma}'_1 \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Gamma}_1)}{n} - \frac{\mathbf{1}' \boldsymbol{\Gamma}'_1 \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Gamma}_1 \mathbf{1}}{n^2} \geq 0 \quad E[\mathbf{w}'_{(\mu)t} \mathbf{y}_t] = \sigma^2(\boldsymbol{\mu}) \geq 0$$

PROVA

A primeira parte da prova compreende achar a forma analítica de  $E[\mathbf{w}'_{(\Gamma)_t} \mathbf{y}_t]$ , para tal, é usado o fato que  $E[\mathbf{w}'_{(\Gamma)_t} \mathbf{y}_t] = E[\mathbf{w}'_{(\Gamma)_t} (\mathbf{y}_t - \boldsymbol{\mu})]$ , então:

$$\begin{aligned}
E[\mathbf{w}'_{(\Gamma)_t} (\mathbf{y}_t - \boldsymbol{\mu})] &= E \left[ \frac{1}{n} \left( \mathbf{B}(\mathbf{y}_{t-1} - \boldsymbol{\mu}) - \frac{\mathbf{1}' \mathbf{B}(\mathbf{y}_{t-1} - \boldsymbol{\mu}) \mathbf{1}}{n} \right) (\mathbf{y}_t - \boldsymbol{\mu}) \right] \\
&= \frac{1}{n} E \left[ (\mathbf{y}_{t-1} - \boldsymbol{\mu})' \mathbf{B}' (\mathbf{y}_t - \boldsymbol{\mu}) \right] + \frac{1}{n^2} E \left[ \mathbf{1}' \left( (\mathbf{y}_{t-1} - \boldsymbol{\mu})' \mathbf{B}' \mathbf{1} \right) (\mathbf{y}_t - \boldsymbol{\mu}) \right] \\
&= \frac{1}{n} E \left[ \text{tr} \left( (\mathbf{y}_{t-1} - \boldsymbol{\mu})' \mathbf{B}' (\mathbf{y}_t - \boldsymbol{\mu}) \right) \right] + \frac{1}{n^2} \text{tr} \left( \mathbf{B}' \mathbf{1} \mathbf{1}' E \right) \\
&= \frac{1}{n} E \left[ \text{tr} \left( \mathbf{B}' (\mathbf{y}_t - \boldsymbol{\mu}) (\mathbf{y}_{t-1} - \boldsymbol{\mu})' \right) \right] + \frac{1}{n^2} \text{tr} (\mathbf{1}' \Gamma_1' \mathbf{B}' \mathbf{1}) \\
&= \frac{\text{tr} (\Gamma_1' \Gamma_0^{-1} \Gamma_1)}{n} - \frac{\mathbf{1}' \Gamma_1' \Gamma_0^{-1} \Gamma_1 \mathbf{1}}{n^2}
\end{aligned}$$

Usando do fato que  $\mathbf{B} = \Gamma_1' \Gamma_0^{-1}$ , então podemos concluir que  $E[\mathbf{w}'_{(\Gamma)_t} (\mathbf{y}_t - \boldsymbol{\mu})] \geq 0$

A segunda parte da demonstração consiste em calcular:

$$\begin{aligned}
E[\mathbf{w}'_{(\mu)_t} \mathbf{y}_t] &= E \left[ \frac{1}{n} \left( \boldsymbol{\mu} - \frac{\mathbf{1}' \boldsymbol{\mu} \mathbf{1}}{n} \right)' \mathbf{y}_t \right] \\
&= \frac{1}{n} \left( \boldsymbol{\mu} - \frac{\mathbf{1}' \boldsymbol{\mu} \mathbf{1}}{n} \right)' \boldsymbol{\mu} \\
&= \sigma^2(\mu) \geq 0
\end{aligned}$$

A primeira parte desta proposição diz que o carteira de arbitragem com VAR pode ser decomposta em outras duas, igualmente de arbitragem. A primeira, chamada de  $\mathbf{w}_{(\Gamma)_t}$ , é responsável por dar os pesos relacionados à persistência temporal, a qual é relacionada à matriz de coeficientes  $\mathbf{B}$ . Esse portfólio dá pesos positivos àqueles ativos que mostraram um bom desempenho em excesso recentemente frente ao excesso de retorno na média incondicional, descrito então pela subtração do termo  $\mathbf{B}(\mathbf{y}_{t-1} - \boldsymbol{\mu})$  pelo termo  $\mathbf{1}' \mathbf{B}(\mathbf{y}_{t-1} - \boldsymbol{\mu}) \mathbf{1}/n$ , pesos negativos ocorrem para o contrário.

O segundo portfólio, chamado de  $\mathbf{w}_{(\mu)_t}$ , explora a variância cruzada existente nos retornos incondicionais. Essa carteira dá pesos positivos para aqueles ativos cuja média incondicional do retorno está acima da média de uma carteira ingênua, e pesos negativos caso contrário.

Já a segunda parte da proposição diz que o valor esperado da carteira  $\mathbf{w}_{(\Gamma)_t}$  depende apenas das matrizes de covariância e covariância cruzada  $\Gamma_0$  e  $\Gamma_1$ . Essa parte da proposição é importante no que assegura um retorno estritamente positivo para a carteira  $\mathbf{w}_t^{VAR}$ , exceto quando o retorno incondicional de todos os ativos são idênticos, o que iguala a carteira ao retorno de uma ingênua, para esse caso, o retorno esperado de  $\mathbf{w}_t^{VAR}$  é 0.

Em contraste com as carteiras de arbitragem contrária e baseada no momento, a carteira de arbitragem com VAR assegura um valor positivo esperado do retorno. Esse fato advém da vantagem de que esse portfólio explora a autocorrelação, tanto quanto a correlação cruzada entre o retorno dos diferentes ativos, independentemente se ele foi positivo ou negativo. Já para a carteira de arbitragem contrária, o valor esperado será apenas positivo quando a autocorrelação for negativa e a correlação cruzada for positiva. Ao passo que, para a carteira de arbitragem em momento, ocorre quando o contrário.

A relação entre a carteira de arbitragem baseada no VAR e as outras duas está justamente na estrutura completa da dependência temporal entre os retornos que a matriz de coeficientes  $\mathbf{B}$  dá. Enquanto o VAR garante que a carteira possa trabalhar com todos os elementos dessa matriz, ou seja, todas as relações lineares existentes entre os ativos, as outras duas carteiras assumem que o  $\mathbf{B}$  é na verdade uma matriz diagonal onde todos os elementos não nulos são iguais, algebricamente, elas assumem que  $\mathbf{B} = \alpha \mathbf{I}$ , onde  $\mathbf{I}$  é uma matriz identidade  $n$ -dimensional e  $\alpha$  é um escalar.

### 2ª proposição

Sob as mesmas suposições que embasam a primeira proposição, a segunda diz que a carteira  $\mathbf{w}_{(\Gamma)t}$  ainda pode ser decomposta na soma de outras duas, uma relacionada às auto covariâncias, chamamo-la de  $\mathbf{w}_{(\Gamma)t}^O$ , e outra relacionada às covariâncias cruzadas, aqui descrita por  $\mathbf{w}_{(\Gamma)t}^C$ . Então:

$$\mathbf{w}_{(\Gamma)t} = \mathbf{w}_{(\Gamma)t}^O + \mathbf{w}_{(\Gamma)t}^C$$

Onde

$$\mathbf{w}_{(\Gamma)t}^O = \frac{1}{n} \left( \mathbf{B}_O(\mathbf{y}_{t-1} - \boldsymbol{\mu}) - \frac{\mathbf{1}' \mathbf{B}_O(\mathbf{y}_{t-1} - \boldsymbol{\mu})}{n} \mathbf{1} \right) \quad (2.8)$$

$$\mathbf{w}_{(\Gamma)t}^C = \frac{1}{n} \left( \mathbf{B}_C(\mathbf{y}_{t-1} - \boldsymbol{\mu}) - \frac{\mathbf{1}' \mathbf{B}_C(\mathbf{y}_{t-1} - \boldsymbol{\mu})}{n} \mathbf{1} \right) \quad (2.9)$$

### PROVA

Abrindo a matriz de coeficientes  $\mathbf{B}$ , podemos decompô-la na soma de outras duas, chamadas de  $\mathbf{B}_O$  e  $\mathbf{B}_C$ . A primeira do lado direito explica as auto covariâncias e a segunda explica as covariâncias cruzadas. Para achá-las, basta decompor a matriz  $\mathbf{\Gamma}_1$  na soma de outras duas, uma com elementos diagonais e outra contendo apenas elementos de fora das diagonais

$$\begin{aligned}
\mathbf{\Gamma} &= \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn} \end{bmatrix} \\
&= \begin{bmatrix} \gamma_{11} & 0 & \cdots & 0 \\ 0 & \gamma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_{nn} \end{bmatrix} + \begin{bmatrix} 0 & \gamma_{12} & \cdots & \gamma_{1n} \\ \gamma_{21} & 0 & \cdots & \gamma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \cdots & 0 \end{bmatrix} \\
&= \text{diag}(\mathbf{\Gamma}_1) + (\mathbf{\Gamma} - \text{diag}(\mathbf{\Gamma}_1))
\end{aligned}$$

Então a matriz de coeficientes deve ser decomposta da seguinte forma:

$$\begin{aligned}
\mathbf{B}' &= \mathbf{\Gamma}_O^{-1}(\text{diag}(\mathbf{\Gamma}_1) + (\mathbf{\Gamma} - \text{diag}(\mathbf{\Gamma}_1))) \\
&= \mathbf{\Gamma}_O^{-1}\text{diag}(\mathbf{\Gamma}_1) + \mathbf{\Gamma}_O^{-1}(\mathbf{\Gamma} - \text{diag}(\mathbf{\Gamma}_1)) \\
&= \mathbf{B}'_O + \mathbf{B}'_C
\end{aligned}$$

Tendo em vistas a definição da carteira  $\mathbf{w}_{(\Gamma)t}$ :

$$\mathbf{w}_{(\Gamma)t} = \frac{1}{n} \left( \mathbf{B}(\mathbf{y}_{t-1} - \boldsymbol{\mu}) - \frac{\mathbf{1}'\mathbf{B}(\mathbf{y}_{t-1} - \boldsymbol{\mu})\mathbf{1}}{n} \right)$$

Então segue que:

$$\begin{aligned}
\mathbf{w}_{(\Gamma)t} &= \frac{1}{n} \left( \mathbf{B}(\mathbf{y}_{t-1} - \boldsymbol{\mu}) - \frac{\mathbf{1}'\mathbf{B}(\mathbf{y}_{t-1} - \boldsymbol{\mu})\mathbf{1}}{n} \right) \\
&= \frac{1}{n} \left( (\mathbf{B}_O + \mathbf{B}_C)(\mathbf{y}_{t-1} - \boldsymbol{\mu}) - \frac{\mathbf{1}'(\mathbf{B}_O + \mathbf{B}_C)(\mathbf{y}_{t-1} - \boldsymbol{\mu})\mathbf{1}}{n} \right) \\
&= \frac{1}{n} \left( \mathbf{B}_O(\mathbf{y}_{t-1} - \boldsymbol{\mu}) - \frac{\mathbf{1}'\mathbf{B}_O(\mathbf{y}_{t-1} - \boldsymbol{\mu})\mathbf{1}}{n} \right) + \frac{1}{n} \left( \mathbf{B}_C(\mathbf{y}_{t-1} - \boldsymbol{\mu}) - \frac{\mathbf{1}'\mathbf{B}_C(\mathbf{y}_{t-1} - \boldsymbol{\mu})\mathbf{1}}{n} \right) \\
&= \mathbf{w}_{(\Gamma)t}^O + \mathbf{w}_{(\Gamma)t}^C
\end{aligned}$$

#### 2.4.2.2 Modelos fatoriais

A análise de componentes principais, como já definida anteriormente, busca identificar quais componentes podem explicar as variações que ocorrem num dado conjunto de retornos, para o caso presente. Especificadamente, a habilidade do VAR em gerar valores positivos esperados para a carteira nesse modelo baseada, pode ser explicada através da

análise de componentes principais, os quais são capazes de prever quais ativos mostrarão desempenhos bons e ruins no dia de amanhã.

A parte algébrica dos componentes principais será estendida mais criteriosamente no capítulo tópicos adicionais, por ora, para compreender o funcionamento, basta supor que a matriz de covariâncias dos retornos pode ser decomposta da seguinte forma:

$$\Gamma_0 = \mathbf{Q}\Lambda_0\mathbf{Q}'$$

Em que  $\mathbf{Q}$  é uma matriz ortogonal, cuja definição de ortogonalidade segue o critério  $\mathbf{Q}\mathbf{Q}' = \mathbf{I}$ . As colunas de  $\mathbf{Q}$  também são chamadas de componentes principais. Já  $\Lambda_0$  é uma matriz diagonal cujos elementos representam as variâncias dos componentes principais correspondentes. Não há um consenso sobre a escolha desses componentes principais, existem várias abordagens a respeito, algumas mais utilizadas em finanças. Matricialmente temos:

$$\mathbf{Q} = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{n1} & q_{n2} & \cdots & q_{nn} \end{bmatrix} \quad \Lambda = \begin{bmatrix} \lambda_{11} & 0 & \cdots & 0 \\ 0 & \lambda_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{nn} \end{bmatrix}$$

Usando-se desse resultado, o modelo VAR trabalhado até então pode ser reescrito de forma que:

$$\mathbf{y}_{t+1} = \mathbf{A} + \mathbf{B}\mathbf{Q}\mathbf{Q}'\mathbf{y}_t + \mathbf{E}_{t+1} \quad (2.10)$$

$$= \mathbf{A} + \hat{\mathbf{B}}\mathbf{p}_t + \mathbf{E}_{t+1} \quad (2.11)$$

Onde:

$$\mathbf{p}_t = \mathbf{Q}'\mathbf{y}_t\hat{\mathbf{B}} = \mathbf{B}\mathbf{Q}$$

O vetor  $n$ -dimensional  $\mathbf{p}_t$  corresponde então ao retorno dos componentes principais, e  $\hat{\mathbf{B}}$ , a matriz de coeficientes expressa pela matriz de covariâncias desses componentes principais. Segue então a terceira proposição:

### 3ª proposição

Sob as mesmas hipóteses que sustentam a primeira proposição, o valor esperado da carteira de arbitragem com VAR pode ser escrita na forma de componentes principais como:

$$E[\mathbf{w}_t^{VAR'} \mathbf{y}_t] = \frac{n-1}{n} \sum_j^n \lambda_j \text{var}(\hat{\mathbf{B}}_j) + \sigma^2(\boldsymbol{\mu})$$

Onde o índice  $j$  indica a  $j$ -ésima coluna em  $\hat{\mathbf{B}}$  e o  $j$ -ésimo elemento da diagonal de  $\boldsymbol{\Lambda}$ .

### PROVA

Da definição do valor esperado para a carteira de arbitragem com base no VAR temos:

$$E[\mathbf{w}_t^{VAR'} \mathbf{y}_t] = \frac{\text{tr}(\boldsymbol{\Gamma}'_1 \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Gamma}_1)}{n} - \frac{\mathbf{1}' \boldsymbol{\Gamma}'_1 \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Gamma}_1 \mathbf{1}}{n^2} + \sigma^2(\boldsymbol{\mu})$$

Substituindo  $\mathbf{B} = \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Gamma}_1$  e  $\boldsymbol{\Gamma}'_1 = \mathbf{B} \boldsymbol{\Gamma}_0$  chegamos na forma:

$$E[\mathbf{w}_t^{VAR'} \mathbf{y}_t] = \frac{\text{tr}(\mathbf{B} \boldsymbol{\Gamma}_0 \mathbf{B})}{n} - \frac{\mathbf{1}' \mathbf{B} \boldsymbol{\Gamma}_0 \mathbf{B}' \mathbf{1}}{n^2} + \sigma^2(\boldsymbol{\mu})$$

Substituindo  $\boldsymbol{\Gamma}_0 = \mathbf{Q} \boldsymbol{\Lambda}_0 \mathbf{Q}$  na expressão acima obtemos:

$$\begin{aligned} E[\mathbf{w}_t^{VAR'} \mathbf{y}_t] &= \frac{\text{tr}(\mathbf{B} \mathbf{Q} \boldsymbol{\Lambda}_0 \mathbf{Q} \mathbf{B})}{n} - \frac{\mathbf{1}' \mathbf{B} \mathbf{Q} \boldsymbol{\Lambda}_0 \mathbf{Q} \mathbf{B}' \mathbf{1}}{n^2} + \sigma^2(\boldsymbol{\mu}) \\ &= \frac{\text{tr}(\hat{\mathbf{B}} \boldsymbol{\Lambda}_0 \mathbf{Q} \mathbf{B})}{n} - \frac{\mathbf{1}' \mathbf{B} \mathbf{Q} \boldsymbol{\Lambda}_0 \mathbf{Q} \mathbf{B}' \mathbf{1}}{n^2} + \sigma^2(\boldsymbol{\mu}) \\ &= \frac{\sum_{i,j,k} \lambda_j \hat{\mathbf{B}}_{ij}^2}{n} - \frac{\sum_{i,j} \lambda_j \hat{\mathbf{B}}_{ij} \hat{\mathbf{B}}_{kj}}{n^2} + \sigma^2(\boldsymbol{\mu}) \\ &= \frac{1}{n^2} \sum_{i,j,k} \lambda_j \hat{\mathbf{B}}_{ij} (\hat{\mathbf{B}}_{ij} - \hat{\mathbf{B}}_{kj}) + \sigma^2(\boldsymbol{\mu}) \\ &= \frac{1}{N} \sum_{i,j} \lambda_j \hat{\mathbf{B}}_{i,j} (\hat{\mathbf{B}}_{i,j} - \sum_i \hat{\mathbf{B}}_{ij}/N) + \sigma^2(\boldsymbol{\mu}) \end{aligned}$$

Usando do resultado  $\sum_i (\sum_i \hat{\mathbf{B}}_{ij}/n) (\hat{\mathbf{B}}_{ij} - (\sum_i \hat{\mathbf{B}}_{ij}/n))$ , então:

$$\begin{aligned} E[\mathbf{w}_t^{VAR'} \mathbf{y}_t] &= \frac{1}{n} \sum_j^n \lambda_j \sum_i^n (\hat{\mathbf{B}}_{ij} - (\sum_i \hat{\mathbf{B}}_{ij}/n))^2 + \sigma^2(\boldsymbol{\mu}) \\ &= \frac{n-1}{n} \sum_j^n \lambda_j \text{var}(\hat{\mathbf{B}}_j) + \sigma^2(\boldsymbol{\mu}) \end{aligned}$$

Intuitivamente essa terceira proposição mostra que altos valores assumidos pelas variâncias, correspondidos pelos valores  $\lambda_j$ , multiplicados pelas correspondentes colunas de  $\hat{\mathbf{B}}$ , produzem altos valores esperados para a carteira  $\mathbf{w}_t^{VAR'}$ . Em outras palavras,

quanto mais a variação do  $j$ -ésimo componente principal, maior será sua predição sobre os retornos de amanhã, e mais será o retorno esperado para o portfólio. Logo, a variação dos componentes principais nos servem de indícios para quais ativos devem ser selecionados a fim de que se tome uma posição comprada ou vendida, para o primeiro caso, os maiores valores esperado do retorno de amanhã, e os menores para o segundo caso.

Há ainda outras abordagens dentro da análise de componentes principais que não o VAR. Uma alternativa é o modelo de regressão fatorial múltiplo, como exemplo, temos extensivamente utilizados na literatura os fatores elaborados por Fama-French já abordados. Algebricamente a relação é escrita da seguinte forma:

$$\mathbf{y}_{t+1} = \mathbf{A}^f + \mathbf{B}^f \mathbf{f}_t + \mathbf{E}_{t+1}^f$$

Onde  $\mathbf{A}^f$  é um vetor  $n$ -dimensional, distinguindo-se do anteriormente utilizado pelo sobrescrito  $f$ , indicando que se trata de um modelo fatorial agora. Igualmente temos essa relação para a matriz de coeficientes  $\mathbf{B}^f$  e o termo de erro aleatório  $\mathbf{E}_{t+1}^f$ .

Com base nesses fatores, (DeMiguel et al.) mostra que é possível também construir carteiras de arbitragem baseadas nestes. Considere agora a seguinte carteira, cujos pesos são dados pela relação algébrica:

$$\mathbf{w}_{t+1}^f = \frac{1}{n}(\mathbf{A}^f + \mathbf{B}^f \mathbf{f}_t - \bar{y}_t^f \mathbf{1})$$

Onde  $\mathbf{A}^f + \mathbf{B}^f \mathbf{f}_t = \hat{\mathbf{y}}_{t+1}^f$  é o valor predito dos retornos dos ativos através do modelo fatorial, e  $\bar{y}_t^f = (\mathbf{A}^f + \mathbf{B}^f \mathbf{f}_t)' \mathbf{1} / n$  é a média dos valores preditos pelos fatores.

#### 4ª proposição

Essa proposição equivale a terceira, porém agora a matriz de covariâncias  $\Gamma_0^f = E[(\mathbf{f}_t = \boldsymbol{\mu}^f)(\mathbf{f}_t = \boldsymbol{\mu}^f)']$  refere-se aos fatores e não mais ao retornos dos ativos como outrora. A terceira proposição pode então ser facilmente modificada de forma a obter:

$$E[\mathbf{w}_t^{VAR'} \mathbf{y}_t] = \frac{n-1}{n} \sum_j \lambda_j^f \text{var}(\hat{\mathbf{B}}_j^f) + \sigma^2(\mu)$$

Onde o índice  $j$  indica a  $j$ -ésima coluna em  $\hat{\mathbf{B}}^f = \mathbf{B}^f \mathbf{Q}$  e o  $j$ -ésimo elemento da diagonal de  $\boldsymbol{\Lambda}^f$ . Com todas as propriedades e demonstrações mantendo-se constantes.

A interpretação mantém-se igual, com ressalvas de que agora trata-se da capacidade dos fatores, quanto a sua variação, em predizer o retorno dos ativos no período seguinte. De outro modo, a variação dos fatores, sendo estes não correlacionados, tem que ser boa o suficiente em predizer quais ativos terão um bom desempenho amanhã, dada

a variação hoje dos fatores, e assim construir a carteira dando peso positivo àqueles que apresentaram um bom desempenho hoje e negativo caso contrário.

#### 2.4.2.3 Carteira de média-variância com VAR

Essa carteira é similar à carteira tradicional de média-variância, com ressalva de que agora se utiliza das informações disponibilizadas pelo modelo paramétrico VAR, o qual é capaz de oferecer estatisticamente a média dos retornos de amanhã condicionada aos retornos apresentados hoje. Sob a suposição de que o retorno dos ativos segue um processo estocástico linear do tipo VAR(1), então a carteira de média-variância pode ser escrita como:

$$\min_{\mathbf{w}} \quad \mathbf{w}'\Sigma^{VAR}\mathbf{w} - \frac{1}{\gamma}\mathbf{w}'\boldsymbol{\mu}^{VAR}$$

$$s.a \quad \mathbf{w}'\mathbf{1} = 1$$

Onde novamente  $\gamma$  é o grau de aversão ao risco do investidor. Agora,  $\boldsymbol{\mu}^{VAR}$  é o vetor de médias condicionais do retorno dos ativos do período seguinte, dado por  $\boldsymbol{\mu}^{VAR} = \mathbf{A} + \mathbf{B}\mathbf{y}_t$ . A matriz de covariâncias  $\Sigma^{VAR}$  é dada pela expressão:

$$\Sigma^{VAR} = \frac{1}{\tau} \sum_{t=T-\tau+1}^T (\mathbf{y}_t - \mathbf{A} - \mathbf{B}\mathbf{y}_{t-1})(\mathbf{y}_t - \mathbf{A} - \mathbf{B}\mathbf{y}_{t-1})'$$

#### 2.4.2.4 Carteira de média-variância com NAR

Um dos pressupostos que embasam o uso do modelo VAR é que a relação temporal entre o retorno dos ativos é linear. O modelo não paramétrico autorregressivo utilizado em (DeMiguel et al.), chamado de NAR, é uma alternativa à esse pressuposto. De modo geral, modelos paramétricos exigem que sejam impostas diversas restrições sobre o comportamento dos dados no que tange ao seu processo estocástico. Por exemplo, para o caso exposto até então, foram impostas restrições de que o processo é linear e segue o modelo VAR(1). Para que haja a estimação do conjunto de parâmetros  $\boldsymbol{\theta} = \mathbf{A}, \mathbf{B}, \Sigma$ , é necessário saber de antemão a distribuição do termo de erros aleatórios  $\mathbf{E}$ . Os resultados apresentados até então tem como base uma distribuição de probabilidade normal, o que se aproxima bem do que é observado empiricamente.

Essas suposições podem ser, por vezes, demasiadamente restritivas, então os modelos não-paramétricos surgem como uma alternativa mais flexível, em que os dados "falam por si próprios", não havendo a necessidade de estimação dos parâmetros.

O modelo não-paramétrico em questão é o NAR com a técnica de regressão conhecida como regressão de vizinhança mais perto. Na prática, essa técnica trabalha com janelas móveis, ou seja, através da coleta do retorno dos ativos num período de  $k$  dias de

cotação passados, são calculadas ambas média e matriz de covariância amostrais e, com base nestas, a carteira de média-variância pode ser calculada.

A média amostral do retorno com base nos  $k$  dias passados é escrita algebricamente como:

$$\boldsymbol{\mu}^{NAR} = \frac{1}{k} \sum_{i=1}^k \mathbf{y}_{t-i}$$

E então a matriz de covariâncias pode ser calculada como:

$$\boldsymbol{\Sigma}^{NAR} = \frac{1}{k-1} \sum_{i=1}^k (\mathbf{y}_{t-i} - \boldsymbol{\mu}^{NAR})(\mathbf{y}_{t-i} - \boldsymbol{\mu}^{NAR})'$$

Com esses dois conjuntos de dados, montar a carteira de média-variância torna-se direta através da substituição dos mesmos na expressão dada anteriormente, a saber:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}' \boldsymbol{\Sigma}^{NAR} \mathbf{w} - \frac{1}{\gamma} \mathbf{w}' \boldsymbol{\mu}^{NAR} \\ \text{s.a} \quad & \mathbf{w}' \mathbf{1} = 1 \end{aligned}$$

### 3 DISTRIBUIÇÕES DE PROBABILIDADE

Esse capítulo tem um caráter mais técnico estatístico e objetiva introduzir as distribuições de probabilidade utilizadas nesse trabalho, discorrendo sobre algumas de suas propriedades, visando a correta motivação de seu uso na inferência Bayesiana, a ser dissertada com maior grau de detalhamento nos capítulos posteriores.

Com um foco didático, então inicialmente são abordadas as distribuições densidade probabilidade, ou f.d.p. univariadas, ou seja, quando é atribuída uma distribuição de probabilidade à apenas uma variável aleatória, ou v.a.. Essa forma foi escolhida pois, uma vez que trabalhar com as f.d.p. para apenas uma v.a. é mais intuitivo, a relação existente entre elas pode ser verificada de forma mais fácil. O segundo momento do capítulo, então, estende essas mesmas distribuições para suas respectivas versões multivariadas, quando as v.a.s não são mais modeladas individualmente, porém conjuntamente, e passam a ser vetores aleatórios.

#### 3.1 Distribuições univariadas

A seguir serão abordados diferentes tipos de distribuição de probabilidade que serão utilizadas ao longo do trabalho. Por simplicidade inicialmente serão vistas na sua forma univariada.

##### 3.1.1 Distribuição normal

Essa é a mais simples dentre as versões da normal gaussiana, é de interesse apresentá-la pois fornece uma visão básica da forma funcional das normais, uma vez que as demais formas são meramente uma extensão desta. Esse tipo de distribuição depende de dois tipos de parâmetros, também associados aos dois primeiros momentos populacionais, o primeiro deles é a média  $\mu$  e o segundo é a variância em torno da média  $\sigma^2$ , sendo ambos finitos e o segundo estritamente positivo. Supondo um valor  $y \in \mathbb{R}$  e, sabendo que este advém de uma população com distribuição normal, a probabilidade de que  $y$  assumira um valor dado é então representado pela f.d.p.:

$$f(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right) \quad (3.1)$$

Por vezes, é de interesse assumir que a média e a variância tenham valores 0 e 1 respectivamente. Quando tal, chamamos a distribuição normal de padrão e ela assume a forma:

$$f(y, \mu, \sigma) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \quad (3.2)$$

### 3.1.2 Distribuição Gama

A base da distribuição de probabilidade Gama é a função de mesmo nome, definida como:

$$\Gamma(\alpha) = \int_0^{\infty} u^{\alpha-1} e^{-u} du$$

Essa função possui algumas propriedades que podem ser de interesse ao se trabalhar com essa distribuição:

**Propriedade 1**  $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$

PROVA<sup>1</sup>

$$\begin{aligned} \Gamma(\alpha + 1) &= \int_0^{\infty} u^{\alpha} e^{-u} du && \text{integrando por partes,} \\ &= -u^{\alpha} e^{-u} \Big|_0^{\infty} + \alpha \int_0^{\infty} u^{\alpha-1} e^{-u} du \\ &= \alpha\Gamma(\alpha) \end{aligned}$$

**Propriedade 2**  $\Gamma(n + 1) = n!$  para todo número inteiro positivo  $n$ .

PROVA<sup>2</sup>

$$\begin{aligned} \Gamma(n + 1) &= n\Gamma(n) \\ &= n(n - 1)\Gamma(n - 1) \\ &= n(n - 1)(n - 2)\Gamma(n - 2) \\ &= n(n - 1)(n - 2) \dots 2\Gamma(1) = n! \end{aligned}$$

**Propriedade 3**  $\Gamma(1/2) = \sqrt{\pi}$

PROVA

a) Inicialmente, deve-se substituir  $u = v^2$  e depois  $\alpha$  por  $(1/2)$ :

$$\begin{aligned} \Gamma(\alpha) &= 2 \int_0^{\infty} v^{2\alpha-1} e^{-v^2} dv \\ \Gamma\left(\frac{1}{2}\right) &= 2 \int_0^{\infty} e^{-v^2} dv \end{aligned}$$

<sup>1</sup> Foi usado o fato que  $-u^{\alpha} e^{-u} \Big|_0^{\infty}$  é 0, uma vez que quando  $u \rightarrow \infty$ ,  $e^{-u}$  converge mais rapidamente a 0 que  $u^{\alpha}$

<sup>2</sup> Foi usada a integral  $\Gamma(1) = \int_0^{\infty} e^{-u} du = -e^{-u} \Big|_0^{\infty} = 1$

b) Trabalhar na forma quadrática da função:

$$\begin{aligned} \left[ \Gamma\left(\frac{1}{2}\right) \right]^2 &= \left[ \Gamma\left(\frac{1}{2}\right) \right] \left[ \Gamma\left(\frac{1}{2}\right) \right] \\ &= \left[ 2 \int_0^\infty e^{-v^2} dv \right] \left[ 2 \int_0^\infty e^{-w^2} dw \right] \\ &= 4 \int_0^\infty \int_0^\infty e^{-(v^2+w^2)} dv dw \end{aligned}$$

c) Fazer a transformação do cartesiano em coordenadas polares  $(r, \theta)$ . Para tal, substituímos  $v = r \cos \theta$  e  $w = r \sin \theta$ , com  $r \in [0, \infty($  e  $\theta \in [0, \frac{\pi}{2}]$ . O Jacobiano então tem a forma:

$$|\mathbf{J}| = \begin{vmatrix} \frac{\partial v}{\partial r} & \frac{\partial v}{\partial \theta} \\ \frac{\partial w}{\partial r} & \frac{\partial w}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r(\cos^2 \theta + \sin^2 \theta) = r$$

Então reescrevemos como:

$$\begin{aligned} 4 \int_0^\infty \int_0^\infty e^{-(v^2+w^2)} dv dw &= 4 \int_0^{\frac{\pi}{2}} \int_0^\infty r e^{-r^2} dr d\theta \\ &= 4 \left[ \int_0^{\frac{\pi}{2}} d\theta \right] \left[ \int_0^\infty r e^{-r^2} dr \right] \\ &= 4 \left[ \frac{\pi}{2} \right] \left[ -\frac{1}{2} \int_0^\infty e^{-z} dz \right] \\ &= -\pi[0 - 1] = \pi \end{aligned}$$

d) Como o termo  $\Gamma(\frac{1}{2})$  foi elevado ao quadrado, tiramos a raiz do resultado acima e a prova está completa.

Dividindo o integrando da função gama pela própria função gama, garantimos que a integral no intervalo  $[0, \infty($  seja 1. Deste modo, a função de probabilidade gama pode ser expressa algebricamente:

$$f(u) = \begin{cases} \frac{u^{\alpha-1} e^{-u}}{\Gamma(\alpha)}, & 0 \leq u \leq \infty \\ 0 & u < 0 \end{cases} \quad (3.3)$$

A distribuição gama, como apresentada, possui um parâmetro  $\alpha$  que é responsável por dar a forma à distribuição. Ao acrescentar outro parâmetro, fazendo  $y = \beta u$  e substituindo essa parametrização da equação dada, a f.d.p. gama fica com a forma:

$$f(y|\alpha, \beta) = \frac{y^{\alpha-1} e^{-\frac{y}{\beta}}}{\Gamma(\alpha)\beta^\alpha}, \quad 0 \leq y \leq \infty, \quad \alpha \geq 0, \quad \beta \geq 0 \quad (3.4)$$

O novo parâmetro  $\beta$  é conhecido como parâmetro de escala, uma vez que este é responsável pela dispersão da distribuição.

A função geradora de momentos da distribuição gama é importante para relacionar os momentos desta distribuição com os de outras advindas desta. Algebricamente, ela toma a seguinte forma:

$$M_Y(t) = (1 - \beta t)^{-\alpha} \quad (3.5)$$

PROVA

$$\begin{aligned} M_Y(t) &= \int_0^{\infty} e^{tY} \frac{y^{\alpha-1} e^{-\frac{y}{\beta}}}{\Gamma(\alpha)\beta^\alpha} dy \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^{\infty} y^{\alpha-1} e^{-y(\frac{1}{\beta}-t)} dy \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^{\infty} y^{\alpha-1} \exp \left\{ -y \underbrace{\left( \frac{1 - \beta t}{\beta} \right)}_{1/c} \right\} dy \end{aligned}$$

Considerando que a integral acima faz parte de uma distribuição gama, a manipulamos de forma que:

$$\begin{aligned} \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^{\infty} y^{\alpha-1} e^{-\frac{y}{c}} dy &= 1 \\ \int_0^{\infty} y^{\alpha-1} e^{-\frac{y}{c}} dy &= \Gamma(\alpha)\beta^\alpha \end{aligned}$$

Fazendo a devida substituição, encontramos:

$$M_Y(t) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \Gamma(\alpha)\beta^\alpha \left( \frac{1}{1 - \beta t} \right)^\alpha = (1 - \beta t)^{-\alpha}$$

A f.d.p. gama é muito útil porque é uma distribuição flexível, capaz de gerar diversas outras distribuições populares na estatística. Não é muito difícil ver que, se  $(x - \mu) = y \sim N(0, \sigma^2)$  então a distribuição de  $y^2$  pode ser obtida através da gama quando colocamos os parâmetros de forma  $\alpha = 1/2$  e o de escala  $\beta = 2$ :

$$\Gamma(y|\alpha = \frac{1}{2}, \beta = 2\sigma^2) = \frac{y^{-\frac{1}{2}}}{\sqrt{\pi}\sqrt{2\sigma}} e^{-\frac{y}{2\sigma^2}} = N^2 \sim (\mu, \sigma^2) \quad (3.6)$$

### 3.1.3 Distribuição qui-quadrado

A distribuição qui-quadrado com  $n$  graus de liberdade ( $\chi_n^2$ ) é importante por relacionar a distribuição gama com a normal. Considerando-se que  $y \sim N(0, 1)$  é uma normal padrão, a distribuição de  $Y^2$  é então derivada como<sup>3</sup>:

$$\begin{aligned}
 F_{Y^2} &= p(Y^2 \leq y) = p(-\sqrt{y} \leq Y \leq \sqrt{y}) \\
 &= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{y}}^{\sqrt{y}} e^{-\frac{t^2}{2}} dt \\
 f_{Y^2} &= \frac{\partial}{\partial y} F_{Y^2} \\
 &= \frac{1}{\sqrt{2\pi}} \frac{\partial}{\partial y} \int_{-\sqrt{y}}^{\sqrt{y}} e^{-\frac{t^2}{2}} dt \\
 &= \frac{1}{\sqrt{2\pi}} \left[ \frac{1}{2\sqrt{y}} e^{-\frac{\sqrt{y}^2}{2}} - \left( -\frac{1}{2\sqrt{y}} \right) e^{-\frac{(-\sqrt{y})^2}{2}} \right] \\
 &= \frac{2}{\sqrt{2\pi}} \frac{1}{2\sqrt{y}} e^{-\frac{y}{2}} = \frac{y^{(1-\frac{1}{2})}}{\sqrt{2\pi}} e^{-\frac{y}{2}} \tag{3.7}
 \end{aligned}$$

Através de 3.7, chegamos à distribuição  $\chi_1^2$ , o que equivale a dizer que a distribuição qui-quadrado com 1 grau de liberdade é uma gama com parâmetros  $\Gamma(\frac{1}{2}, 2)$ .

Se somarmos  $n$  variáveis  $y$  independentes e identicamente distribuídas pela normal padrão,  $y_i \sim n(0, 1)$  e, seja  $W = \sum_{i=1}^n y_i^2$ , a distribuição de  $W$  é dada então pela qui-quadrado com  $n$  graus de liberdade,  $W \sim \chi_n^2$ .

#### PROVA<sup>4</sup>

$$M_W(t) = M_{y_1^2}(t) \times M_{y_2^2}(t) \times \dots \times M_{y_n^2}(t)$$

Da distribuição gama, sabemos que sua função geradora de momentos tem a forma  $M_x = (1 - \beta t)^{-\alpha}$ , substituindo os parâmetros pelos da distribuição  $\chi^2$ , temos que  $M_{y^2} = (1 - 2t)^{-\frac{1}{2}}$ , então a f.g.m. de  $W$  pode ser obtida diretamente:

<sup>3</sup> Para a demonstração, é utilizada a propriedade que, se  $F(x) = \int_{i(x)}^{s(x)} f(t)dt$ , então  $F'(x) = s'(x)f[s(x)] - i'(x)f[i(x)]$

<sup>4</sup> Usa-se da propriedade que, sendo  $x$  e  $y$  variáveis aleatórias independentes com suas respectivas f.g.m.  $M_x(t)$  e  $M_y(t)$  e sabendo que  $E[f(x)g(y)] = E[f(x)]E[g(y)]$ , a f.g.m. da soma  $x + y$  é obtida como:

$$M_{x+y}(t) = E[e^{t(x+y)}] = E[e^{tx}e^{ty}] = E[e^{tx}]E[e^{ty}] = M_x(t).M_y(t)$$

$$M_W(t) = \underbrace{(1-2t)^{-\frac{1}{2}}(1-2t)^{-\frac{1}{2}} \dots (1-2t)^{-\frac{1}{2}}}_{\times n} = (1-2t)^{-\frac{n}{2}}$$

O resultado obtido em 3.1.3 equivale dizer que a  $W \sim \Gamma(\frac{n}{2}, 2)$ , o que na verdade é uma distribuição  $\chi_n^2$  com  $n$  graus de liberdade:

$$f(W) = \frac{W^{\frac{n}{2}-1} e^{-\frac{W}{2}}}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}}}$$

Uma importante aplicação da distribuição  $\chi_n^2$  na inferência bayesiana ocorre da distribuição a priori da variância. Supõe-se que  $y_i$  sejam variáveis aleatórias i.i.d.  $N \sim (\mu, \sigma^2)$  para todo  $i$ . Padronizando-as como  $z_i = (y_i - \mu)/\sigma$  temos que  $z \sim N(0, 1)$  cuja distribuição de  $\sum_{i=1}^n z_i^2$  é uma  $\chi_n^2$ . Dois resultados interessantes podem ser retirados dessa relação, o primeiro trata-se da relação entre a variância e a distribuição  $\chi_n^2$ , essa relação pode ser descrita algebricamente:

$$\frac{(y_1 - \mu)^2}{\sigma^2} + \frac{(y_2 - \mu)^2}{\sigma^2} + \dots + \frac{(y_n - \mu)^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n z_i^2$$

O que leva diretamente ao resultado que  $w = \sum_{i=1}^n z_i^2 \sim \sigma^2 \chi^2(n)$ .

O segundo resultado relaciona a variância  $\sigma^2$  à distribuição  $\Gamma(\alpha, \beta)$ , aqui, novamente temos que:

$$\frac{(y_1 - \mu)^2}{\sigma^2} + \frac{(y_2 - \mu)^2}{\sigma^2} + \dots + \frac{(y_n - \mu)^2}{\sigma^2}$$

Pode-se rearranjar a expressão acima fazendo  $1/\sigma^2 = \tau^2$ . O novo parâmetro  $\tau^2$  é também conhecido como parâmetro de precisão, uma vez que ele é o inverso da variância. A relação fica clara quando atribui-se à  $\sigma^2$  valores extremos, quando a variância é baixa,  $\sigma^2 \rightarrow 0$ , então a precisão quanto ao valor de  $\mu$  é alto, alternativamente, quando a variância é alta,  $\sigma^2 \rightarrow \infty$ , a precisão é baixa:

$$w = \tau^2 \sum_{i=1}^n z_i^2$$

Sabe-se que  $w = \sum_{i=1}^n z_i^2$  tem f.d.p.  $\chi^2(n)$ , cuja f.g.m. é  $M_w = (1-2t)^{-\frac{n}{2}}$ . Como na estatística é comum atribuir média 0 ao termo de resíduos  $\epsilon$ , podemos simplificar o problema de modo que  $w = s$ , em que  $s$  é a soma dos resíduos de modo que:

$$s = \tau^2 \sum_{i=1}^n \epsilon_i^2$$

Através da expressão acima, chegamos que  $s \sim \Gamma(\frac{n}{2}, \frac{\tau^2}{2})$ .

PROVA<sup>5</sup>

$$\begin{aligned}
M_s(t) &= M_{\tau\epsilon_1^2} \times M_{\tau\epsilon_2^2} \times \dots \times M_{\tau\epsilon_n^2} \\
&= (1 - 2\tau^2 t)^{-\frac{1}{2}} (1 - 2\tau^2 t)^{-\frac{1}{2}} \dots (1 - 2\tau^2 t)^{-\frac{1}{2}} \\
&= (1 - (2\tau^2)t)^{-\frac{n}{2}}
\end{aligned}$$

Da f.g.m. temos que  $M_y(t) = (1 - \beta t)^{-\alpha}$ , então concluímos que a expressão acima tem uma distribuição  $\Gamma \sim (\alpha = \frac{n}{2}, \beta = 2\tau^2)$ :

$$f(s) = \frac{s^{\frac{n}{2}-1} e^{-\frac{s}{2\tau^2}}}{\Gamma(\frac{n}{2})(2\tau^2)^{\frac{n}{2}}}$$

De acordo com os resultados anteriores,  $s$  pode ser considerada uma amostra, pela qual é feita a inferência estatística sobre a real variância, dada pelo parâmetro  $\sigma^2 = \frac{1}{\tau^2}$ . Na abordagem bayesiana, essa lógica é invertida, uma vez que atribui-se uma distribuição de probabilidade ao real parâmetro do modelo, através da distribuição *priori*. Deste ponto de vista, é natural que, sob a hipótese de normalidade dos resíduos de um modelo arbitrário,  $\sigma^{-2}$  seja distribuído pela  $\Gamma(\frac{\nu}{2}, 2\underline{\delta})$ , onde  $\underline{\nu}$  e  $\underline{\delta}$  são dadas antes da observação da amostra, representando o conhecimento *a priori* do pesquisador sobre o comportamento dos parâmetros:

$$\sigma^{-2} \sim \Gamma(\frac{\underline{\nu}}{2}, 2\underline{\delta})$$

## 3.1.4 Distribuição gama inversa

Os resultados anteriores levam à distribuição de  $\sigma^{-2}$ , eles são importantes para o caso quando reparametrização é feita para  $\tau^2$ , o parâmetro de precisão. Por vezes, é melhor atribuir diretamente uma distribuição à variância  $\sigma^2$ , o que não convém trabalhar com a distribuição  $\Gamma$ . A alternativa então é usar sua inversa de forma que  $\sigma^2 \sim i\Gamma(\frac{\underline{\nu}}{2}, 2\underline{\delta})$ .

Para tal, é necessário usar de algumas propriedades de funções. Seja  $x$  uma variável contínua com f.d.p.  $f_x$  no domínio  $\mathcal{X}$ . Suponha que  $y = g(x)$  seja uma função monótona com inversa e derivada contínua em  $\mathcal{Y}$ , para  $\mathcal{Y} = y : y = g(x)$ , então parra todo  $y \in \mathcal{Y}$ :

$$f_y(y) = f_x(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

<sup>5</sup> É utilizada a propriedade da f.g.m. que, sendo  $x$  uma v.a. e  $y = a + bx$ , então:

$$M_y(t) = E[e^{ty}] = E[e^{t(a+bx)}] = e^{at} E[e^{(bt)x}] = e^{at} M_x(bt)$$

Usando desse resultado, a obtenção da f.d.p. gama inversa é direta, seja a função gama dada pela seguinte expressão:

$$f(x|\alpha, \beta) = \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha}, \quad 0 \leq x \leq \infty, \quad \alpha \geq 0, \quad \beta \geq 0 \quad (3.8)$$

Se  $g(x) = y = 1/x$ , temos que  $g^{-1}(y) = x = 1/y$  e a substituição pode ser feita diretamente em  $f_x$ , a qual é dada pela distribuição gama, a lembrar:

$$\begin{aligned} f_y(y) &= \frac{\left(\frac{1}{y}\right)^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} \exp\left\{-\frac{1}{\beta y}\right\} \frac{1}{y^2} \\ &= \frac{y^{-\alpha+1-2}}{\Gamma(\alpha)\beta^\alpha} \exp\left\{-\frac{1}{\beta y}\right\} \\ &= \frac{((1/\beta)^\alpha)y^{-(\alpha+1)}}{\Gamma(\alpha)} \exp\left\{-\frac{(1/\beta)}{y}\right\} \end{aligned} \quad (3.9)$$

A expressão 3.9 é a distribuição Gama inversa, também escrita de forma compacta como  $y \sim iG(\alpha, 1/\beta)$ . O cálculo dos momentos pode ser feita simples, fazendo  $x \sim iG(\alpha, \beta)$ :

$$\begin{aligned} E(x^n) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^n x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right) dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{n-\alpha-1} \exp\left(-\frac{\beta}{x}\right) dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha-n)}{\beta^{\alpha-n}} \\ &= \frac{\beta^n \Gamma(\alpha-n)}{(\alpha-1) \dots (\alpha-n) \Gamma(\alpha-n)} \\ &= \frac{\beta^n}{\prod_{i=1}^n (\alpha-i)} \end{aligned}$$

A média desta distribuição é dada se  $\alpha > 1$ :

$$E(x) = \frac{\beta}{\alpha-1}$$

se  $\alpha > 2$ , então:

$$E(x^2) = \frac{\beta^2}{(\alpha-1)(\alpha-2)}$$

e então a variância pode ser calculada da seguinte forma:

$$\text{Var}(x) = E(x^2) - E(x)^2 = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)}$$

Para o caso anterior onde estávamos interessados na distribuição *priori* de  $\sigma^2$ , porém foi resolvido para  $\sigma^{-2}$ . Agora podemos reescrever de forma que  $\sigma^2 \sim i\Gamma(\nu/2, 2/\omega)$ :

$$f(\sigma^2) = \frac{((s/2)^{\frac{\nu}{2}})\sigma^{-2(\frac{\nu}{2}+1)}}{\Gamma(\frac{\nu}{2})} \exp\left\{-\frac{(\sigma^2/2)}{s}\right\}$$

### 3.1.5 Distribuição t

Se  $Y_1, Y_2, \dots, Y_n$  são variáveis aleatórias *i.i.d.*  $N \sim (\mu, \sigma^2)$ , e sejam  $\bar{Y} = n^{-1} \sum_{i=1}^n Y_i$  e  $S = (n - 1)^{-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$  estimadores amostrais de  $\mu$  e  $\sigma^2$  respectivamente, então:

$$\frac{\bar{Y} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

#### PROVA

Sabendo que  $Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ ,  $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$  e  $Z$  e  $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$  são independentes, então:

$$T = \frac{\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S}{\sigma^2(n-1)}}} = \frac{\bar{Y} - \mu}{S/\sigma} = \frac{\bar{Y} - \mu}{S/\sqrt{n}}$$

Onde  $T$  é distribuído por  $t$  com  $n - 1$  graus de liberdade.

Para derivar essa forma de distribuição, considere uma variável aleatória normal padronizada  $Z \sim (0, 1)$  e seja  $U$  outra variável aleatória distribuída como  $U \sim \chi_{\nu}^2$ , suas distribuições podem ser escritas respectivamente por:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad g(\nu) = \frac{1}{\Gamma(\frac{\nu}{2})} \left(\frac{1}{2}\right)^{\frac{\nu}{2}} u^{\frac{\nu}{2}-1} e^{-\frac{u}{2}} I_{(0,\infty)}(u) \quad (3.10)$$

A f.d.p. conjunta pode ser escrita multiplicando-se ambas as distribuições  $f(z)g(\nu)$ :

$$f_{Z,U}(z, u) = \frac{1}{\sqrt{2\pi}\Gamma(\frac{\nu}{2})} \left(\frac{1}{2}\right)^{\frac{\nu}{2}} u^{\frac{\nu}{2}-1} e^{-\frac{z^2}{2}} e^{-\frac{u}{2}} I_{(0,\infty)}(u)$$

As seguintes transformações devem ser feitas:

$$Y = U \quad \Rightarrow \quad U = Y = h_2(Y, T)$$

$$T = \frac{Z}{\sqrt{U/\nu}} \quad \Rightarrow \quad Z = T\sqrt{\frac{Y}{\nu}} = h_1(Y, T)$$

Diferentemente da ultima subseção, essa transformação tem mais de uma variável, então é necessário recorrermos à matriz Jacobiana da transformação  $\mathbf{J}$ , cujo determinante é dado por:

$$|\mathbf{J}| = \begin{vmatrix} \frac{\partial h_1}{\partial T} & \frac{\partial h_1}{\partial Y} \\ \frac{\partial h_2}{\partial T} & \frac{\partial h_2}{\partial Y} \end{vmatrix} = \begin{vmatrix} \sqrt{\frac{Y}{\nu}} & \frac{T}{2\sqrt{\nu Y}} \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{Y}{\nu}}$$

Fazendo as substituições diretamente em  $f_{Z,U}(z, u)$  obtemos:

$$f_{T,Y}(t, y) = \sqrt{\frac{y}{\nu}} \frac{1}{2\pi} \frac{1}{\Gamma(\frac{\nu}{2})} \left(\frac{1}{2}\right)^{\frac{\nu}{2}} y^{\frac{\nu}{2}-1} e^{-\frac{y}{2}} e^{-\frac{t^2}{2\nu}y} I_{(0,\infty)}(y)$$

Integrando a expressão acima em relação a  $y$  obtém-se a distribuição marginal de  $t$ :

$$\begin{aligned} f(t) &= \frac{1}{\sqrt{2\nu\pi}\Gamma(\frac{\nu}{2})} \left(\frac{1}{2}\right)^{\frac{\nu}{2}} \int_0^\infty y^{\frac{\nu}{2}+\frac{1}{2}-1} e^{-\frac{1}{2}(1+\frac{t^2}{\nu})y} I_{(0,\infty)}(y) \\ &= \frac{1}{\sqrt{2\nu\pi}\Gamma(\frac{\nu}{2})} \left(\frac{1}{2}\right)^{\frac{\nu}{2}} \frac{1}{\left[\frac{1}{2}\left(1+\frac{t^2}{\nu}\right)\right]^{\frac{\nu+1}{2}}} \int_0^\infty \left(\frac{1}{2}\left(1+\frac{t^2}{\nu}\right)y\right)^{\frac{\nu+1}{2}-1} e^{-\frac{1}{2}\left(1+\frac{t^2}{\nu}\right)y} \left(1+\frac{t^2}{\nu}\right)y \, dy \end{aligned}$$

Fazendo a substituição  $w = \left(1 + \frac{t^2}{\nu}\right)y$  pode-se chegar ao seguinte resultado:

$$\begin{aligned} f(t) &= \frac{1}{\sqrt{2\nu\pi}\Gamma(\frac{\nu}{2})} \left(\frac{1}{2}\right)^{\frac{\nu}{2}} \frac{1}{\frac{1}{2}\left(1+\frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}} \int_0^\infty w^{\frac{\nu+1}{2}-1} e^{-w} \, dw \\ &= \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\sqrt{\nu\pi}} \left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \end{aligned} \quad (3.11)$$

A expressão 3.11 é a distribuição  $t$  com  $\nu - 1$  graus de liberdade.

### 3.2 Distribuições multivariadas

Na seção anterior foram introduzidas as principais distribuições utilizadas nesse trabalho, sendo que todas elas foram apresentadas na sua forma univariada, ou seja,

quando leva-se em consideração a distribuição de uma única variável. Atente que, mesmo para o caso quando estuda-se a distribuição  $\chi_n^2$ , a variável em questão é única e dada pela soma de  $n$  outras variáveis,  $w = \sum_{i=1}^n y_i$ . Em economia, dificilmente são feitas inferências sobre uma única variável independente, é comum que se trabalhe com diversas variáveis conjuntamente, onde uma relaciona-se com a outra através de uma distribuição de probabilidade. Para tal, é necessário que as distribuições univariadas sejam estendidas para suas respectivas versões multivariadas. As próximas seções tem o objetivo de expandir as f.d.p. apresentadas nesse sentido.

### 3.2.1 Distribuição normal

A compreensão do comportamento da f.d.p. normal multivariada fica mais clara partindo de modelos mais simples, como para o caso de processos estocásticos univariados. Para esses processos, é comum modelar a variável  $y_t$  de acordo com os valores assumidos por ela mesma temporalmente, para  $t = [1, 2, \dots, T]$ . Simplificando ainda mais, partimos do pressuposto que  $y_t$  é observado em dois periodos de tempo distintos  $y_1, y_2$ . Podemos então construir uma distribuição normal bivariada de modo que possamos construir uma distribuição conjunta para as duas v.a..

Se  $y_1$  e  $y_2$  são eventos independentes, então a distribuição conjunta é dada por  $f(y_1, y_2) = f(y_1)f(y_2)$ . Para o caso em que não são independentes, devemos subtrair a correlação existente entre os dois eventos de forma que  $f(y_1, y_2) = f(y_1)f(y_2) - cov(y_1, y_2)$ . Aplicando esse conceito às f.d.p. das v.a., derivamos a distribuição conjunta como:

$$\begin{aligned} f(y_1, y_2) &= \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(y_1-\mu_1)^2}{2\sigma_1^2}} \times \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(y_2-\mu_2)^2}{2\sigma_2^2}} - \frac{1}{2\pi\sqrt{\sigma_1\sigma_2}} e^{-\frac{(y_1-\mu_1)(y_2-\mu_2)}{2(2\sigma_1^2)\equiv 2(\sigma_{12}\sigma_{21})}} \\ &= \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2 - \sigma_{12}\sigma_{21}}} \exp\left\{-\frac{\sigma_1^{-2}(y_1 - \mu_1)^2 + \sigma_2^{-2}(y_2 - \mu_2)^2 - \sigma_1^{-1}\sigma_2^{-1}(y_1 - \mu_1)(y_2 - \mu_2)}{2}\right\} \end{aligned}$$

Reescrevendo  $\sigma_1^2 = \sigma_{11}$  e  $\sigma_2^2 = \sigma_{22}$ , e sabendo que  $\sigma_{12} = \sigma_{21}$  a equação acima pode ser colocada matricialmente como:

$$\begin{aligned} f(y_1 \quad y_2) &= \frac{1}{(2\pi)^{\frac{2}{2}} \sqrt{\begin{vmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{vmatrix}}} \exp\left\{\left[ \begin{array}{cc} (y_1 - \mu_1) & (y_2 - \mu_2) \end{array} \right] \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} (y_1 - \mu_1) \\ (y_2 - \mu_2) \end{bmatrix}\right\} \\ &= \frac{1}{(2\pi)^{\frac{2}{2}} \sqrt{|\Sigma|}} \exp\left\{\mathbf{y}'\Sigma^{-1}\mathbf{y}\right\} \end{aligned} \quad (3.12)$$

Tomando o exemplo acima, se não houver correlação temporal entre  $y_1$  e  $y_2$ , substituímos  $\sigma_{12} = \sigma_{21} = 0$  e a matriz  $\Sigma$  torna-se diagonal. Para o caso em que as v.a.  $y_1$  e  $y_2$  são retiradas da mesma população com média  $\mu$  e variância  $\sigma^2$ , então os elementos da diagonal de  $\Sigma$  são idênticos.

A f.d.p. em 3.12 pode ser facilmente estendida para o caso em que são observadas  $T$  v.a.  $y_t$ . Para o caso em que  $T > 2$ , não podemos visualizar a distribuição, porém esta mantém a forma genérica apresentada, sendo que a matriz  $\Sigma$  passa a ter dimensão  $T \times T$  e o vetor  $\mathbf{y}$  dimensão  $T$ .

$$f(y_1, y_2, \dots, y_T) = \frac{1}{(2\pi)^{\frac{T}{2}} \sqrt{|\Sigma|}} \exp \left\{ \mathbf{y}' \Sigma^{-1} \mathbf{y} \right\} \quad (3.13)$$

A mesma análise pode ser feita para o caso em que as  $T$  observações são não correlacionadas temporalmente, neste caso, a matriz  $\Sigma$  toma a forma diagonal com  $\sigma_{ij} = 0$  para todo  $i \neq j$ . Acrescentando o pressuposto de que é um processo estacionário e homoscedástico, ou seja, a variância populacional mantém-se constante ao longo do tempo, tal como sua média, reescrevemos a distribuição de forma que  $\mathbf{y} \sim N(\mu, \sigma^2 \mathbf{I})$ , em que  $\mathbf{I}$  é uma matriz identidade de dimensão  $T \times T$ .

O tipo de distribuição apresentada tem um formato útil para processos estocásticos multivariados. Suponha que agora temos  $n$  v.a.  $\mathbf{y}_n$ , com  $n = 1, 2, \dots, N$ , observadas  $T$  vezes ao longo do tempo  $t$ . Uma forma comum de estimarmos a média de  $\mathbf{y}_n$  é empilhando as  $T$  observações de  $y_{1t}$ , depois as  $T$  observações de  $y_{2t}$  e assim por diante até a  $n$ -ésima variável  $y_{nt}$ :

$$\mathbf{Y}_{(T \times N)} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = \text{vec}(\mathbf{y}_{tn})$$

A f.d.p. conjunta de  $\mathbf{Y}$  é então escrita como:

$$\begin{aligned}
f(\mathbf{Y}) &= \frac{1}{(2\pi)^{\frac{TN}{2}} \sqrt{|\Sigma|}} \exp \left\{ \mathbf{y}' \Sigma^{-1} \mathbf{y} \right\} \\
&= (2\pi)^{\frac{-TN}{2}} \left| \begin{array}{cccc} \Sigma_{11} & \Sigma_{12} & \dots & \Sigma_{1N} \\ \Sigma_{21} & \Sigma_{22} & \dots & \Sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{1N} & \Sigma_{1N} & \dots & \Sigma_{NN} \end{array} \right|^{-\frac{1}{2}} \\
&\times \exp \left\{ \left[ \begin{array}{cccc} (\mathbf{y}'_1 - \boldsymbol{\mu}'_1) & (\mathbf{y}'_2 - \boldsymbol{\mu}'_2) & \dots & (\mathbf{y}'_N - \boldsymbol{\mu}'_N) \end{array} \right] \left[ \begin{array}{cccc} \Sigma_{11} & \Sigma_{12} & \dots & \Sigma_{1N} \\ \Sigma_{21} & \Sigma_{22} & \dots & \Sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{1N} & \Sigma_{1N} & \dots & \Sigma_{NN} \end{array} \right]^{-1} \left[ \begin{array}{c} (\mathbf{y}_1 - \boldsymbol{\mu}_1) \\ (\mathbf{y}_2 - \boldsymbol{\mu}_1) \\ \vdots \\ (\mathbf{y}'_N - \boldsymbol{\mu}_1) \end{array} \right] \right\}
\end{aligned}$$

Essa forma é muito comum na estimação do conjunto de parâmetros  $\boldsymbol{\theta}_n = (\mu_n, \Sigma_{nm})$  para todo  $n = 1, 2, \dots, N$  e  $m = 1, 2, \dots, N$ .

A interpretação das matrizes  $\Sigma_{nn}$  é a mesma para o caso do processo univariado, salvo que agora temos o mesmo arranjo  $N$  vezes para o bloco matriz diagonal, onde as restrições quanto à estacionariedade e homoscedasticidade podem ser mantidas. As demais submatrizes  $\Sigma_{nm}$  para todo  $n \neq m$  podem ser interpretadas como as relações temporais entre as  $N$  variáveis  $y_t$ . Por exemplo, tomando a submatriz  $\Sigma_{13}$ , a  $i$ -ésima é dada pela covariância entre  $y_1$  no tempo  $t = i$  com todos os  $t = 1, \dots, T$  da variável  $y_3$ . De modo similar, a coluna  $j$ -ésima de  $\Sigma_{13}$  contém todas as covariâncias de  $y_1$  para  $t = 1, 2, \dots, T$  de acordo com o tempo  $t = j$  para  $y_3$ .

Em economia, comumente são confrontados dados objetivando verificar relações existentes entre si. Para dar continuidade ao tratamento da f.d.p. normal multivariada, suponha o seguinte modelo de regressão linear univariado  $y_t | \boldsymbol{\beta}, \mathbf{X}_t$ , em que  $\mathbf{X}_t$  é um conjunto de  $n$  variáveis explicativas e  $\boldsymbol{\beta}$  seus respectivos coeficientes angulares:

$$y_t = \mathbf{X}_t \boldsymbol{\beta} + e_t$$

Esse modelo pode ser reescrito como  $Y = \mathbf{X} \boldsymbol{\beta} + E$  onde:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1T} & x_{2T} & \dots & x_{nT} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \quad E = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_T \end{bmatrix}$$

Sob a suposição de normalidade do termo de erro  $E$  a f.d.p. normal multivariada condicionada pode ser escrita como:

$$f(Y|\boldsymbol{\beta}, \mathbf{X}) = \frac{1}{(2\pi)^{\frac{T}{2}} \sqrt{|\sigma^2|}} \exp \left\{ (Y - \mathbf{X}\boldsymbol{\beta})' \sigma^{-2} (Y - \mathbf{X}\boldsymbol{\beta}) \right\}$$

O estimador de M.V. ocorre ao maximizar a expressão acima em relação ao vetor  $\boldsymbol{\beta}$ , o que é similar a dizer que, uma vez maximizado, o termo de erro  $E$  foi minimizado. O estimador de M.V. para  $\boldsymbol{\beta}$  na expressão acima é o mesmo que o de M.Q.O.. Sob a hipótese de que os regressores são não-correlacionados, o posto de  $(\mathbf{X}'\mathbf{X})$  é  $n$  e o E.M.V:

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'Y$$

Onde o termo  $(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$  é conhecido como matriz de projeção<sup>6</sup> de dimensão  $n \times T$ .

Para a inferência estatística, deveríamos coletar diferentes valores de  $y_t$  e  $\mathbf{X}_t$  a cada instante  $t$ . Na prática, em economia, isto é impossível, há apenas uma realização a cada  $t$ . Se o termo de erro  $E$  for normalmente distribuído, os parâmetros amostrais também serão e a inferência pode ser feita sobre o vetor  $\boldsymbol{\beta}$  calculando-se seus dois primeiros momentos amostrais:

$$\begin{aligned} E(\tilde{\boldsymbol{\beta}}) &= E[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'Y] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + u)] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}\boldsymbol{\beta}] + E[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E] \\ &= \boldsymbol{\beta} \end{aligned}$$

$$\begin{aligned} Var(\tilde{\boldsymbol{\beta}}) &= E[(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'] \\ &= E[(\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E + \boldsymbol{\beta})(\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E + \boldsymbol{\beta})'] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'EE' \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\sigma^2 \mathbf{I}) \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

Para o cálculo acima, foi usado o fato que os regressores são não correlacionados com o termo de erro, então  $E(\mathbf{X}'E) = 0$  e que  $E$  é homoscedástico,  $Var(E) = E(EE') = \sigma^2 \mathbf{I}$ , onde  $\mathbf{I}$  é uma matriz identidade de dimensão  $T$ .

Claramente, o E.M.V. é um estimador não viesado para  $\boldsymbol{\beta}$ . Adentrando na análise, se  $(\mathbf{X}'\mathbf{X})$  tem posto menor que  $n$ , então há combinação linear entre os regressores e a

<sup>6</sup> O nome advém do fato que essa matriz projeta o vetor  $Y$  de dimensão  $T$  no espaço paramétrico de dimensão  $n$ .

variância do estimador tenderá a infinito, uma vez que  $(\mathbf{X}'\mathbf{X})^{-1} \rightarrow 0$ . No oposto, quanto maior a variação dos regressores, menor será a variância do vetor  $\tilde{\boldsymbol{\beta}}$ .

A distribuição de probabilidade dos parâmetros pode ser escrita pela normal multivariada e a inferência ser feita com base nessa distribuição:

$$f(\boldsymbol{\beta}) = \frac{1}{(2\pi)^{\frac{T}{2}} \sqrt{|\sigma^2(\mathbf{X}'\mathbf{X})^{-1}|}} \exp \left\{ (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})' [\sigma^{-2}(\mathbf{X}'\mathbf{X})] (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \right\}$$

### 3.2.1.1 Normal matricial

A normal matricial, igualmente popular em processos estocásticos multivariados, é obtida através da normal multivariada, rearranjando seus termos. Para a compreensão da relação entre ambas distribuições normais multivariadas, é necessário introduzir os operadores algébricos Kronecker e traço e suas utilidades nessas distribuições.

Relembrando o operador de Kronecker para multiplicação de matrizes, temos que, se  $\mathbf{A}$  é uma matriz  $m \times n$  e  $\mathbf{B}$  uma matriz  $p \times q$ , o operador é dado como:

$$\begin{aligned} \mathbf{A} \otimes \mathbf{B} &= \begin{bmatrix} a_{11} & a_{22} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \otimes \begin{bmatrix} b_{11} & b_{22} & \dots & b_{1q} \\ b_{21} & b_{22} & \dots & b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \dots & b_{pq} \end{bmatrix} \\ &= \begin{bmatrix} a_{11}b_{11} & \dots & a_{11}b_{1q} & a_{12}b_{11} & \dots & a_{12}b_{1q} & \dots & a_{1n}b_{11} & \dots & a_{1n}b_{1q} \\ a_{11}b_{21} & \dots & a_{11}b_{2q} & a_{12}b_{21} & \dots & a_{12}b_{2q} & \dots & a_{1n}b_{21} & \dots & a_{1n}b_{2q} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{11}b_{p1} & \dots & a_{11}b_{pq} & a_{12}b_{p1} & \dots & a_{12}b_{pq} & \dots & a_{1n}b_{p1} & \dots & a_{1n}b_{pq} \\ a_{21}b_{11} & \dots & a_{21}b_{1q} & a_{22}b_{11} & \dots & a_{22}b_{1q} & \dots & a_{2n}b_{11} & \dots & a_{2n}b_{1q} \\ a_{21}b_{21} & \dots & a_{21}b_{2q} & a_{22}b_{21} & \dots & a_{22}b_{2q} & \dots & a_{2n}b_{21} & \dots & a_{2n}b_{2q} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1}b_{p1} & \dots & a_{m1}b_{pq} & a_{22}b_{p1} & \dots & a_{22}b_{pq} & \dots & a_{2n}b_{p1} & \dots & a_{2n}b_{pq} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1}b_{11} & \dots & a_{m1}b_{1q} & a_{m2}b_{11} & \dots & a_{m2}b_{1q} & \dots & a_{mn}b_{11} & \dots & a_{mn}b_{1q} \\ a_{m1}b_{21} & \dots & a_{m1}b_{2q} & a_{m2}b_{21} & \dots & a_{m2}b_{2q} & \dots & a_{mn}b_{21} & \dots & a_{mn}b_{2q} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1}b_{p1} & \dots & a_{m1}b_{pq} & a_{m2}b_{p1} & \dots & a_{m2}b_{pq} & \dots & a_{mn}b_{p1} & \dots & a_{mn}b_{pq} \end{bmatrix} \\ &= \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \dots & a_{mn}\mathbf{B} \end{bmatrix} = \mathbf{C}_{(mp \times nq)} \end{aligned}$$

A forma kronecker é uma representação matricial com propriedades que são interessantes do ponto de vista da f.d.p. normal. Tomando como exemplo a matriz  $\Sigma$ , comumente são impostas algumas restrições sobre o comportamento dos dados de modo que essa matriz é modificada. A suposição de homoscedasticidade faz com que a diagonal principal de cada submatriz  $\Sigma_{ij}$  para todo  $i = 1, \dots, N$  e  $j = 1, \dots, N$  seja constante. Ao adicionar a suposição de que os dados não se relacionam intertemporalmente, ou seja, não há correlação temporal, cada submatriz  $\Sigma_{ij}$  torna-se diagonal. Essas suposições são importantes na modelagem do termo de erro aleatório pois através delas e outras restrições adicionais, a correta inferência pode ser feita.

Desta forma, a matriz  $\Sigma$  pode ser decomposta na forma kronecker, onde  $\Sigma_N$  e  $\Sigma_T$  representam a covariância entre as  $n$  variáveis do modelo e a covariância intertemporal, respectivamente. Tomando uma submatriz  $\Sigma_{ij}$  arbitrária, esta pode ser reescrita como  $\sigma^{ij}\Sigma_T$ , em que  $\Sigma_T$  toma a forma de uma matriz identidade de ordem  $T$ ,  $\mathbf{I}_T$ :

$$\begin{aligned} \Sigma_{(NT \times NT)} &= \begin{bmatrix} \sigma_{11} & 0 & \dots & 0 & \dots & \sigma_{1N} & 0 & \dots & 0 \\ 0 & \sigma_{11} & \dots & 0 & \dots & 0 & \sigma_{1N} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \dots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{11} & \dots & 0 & 0 & \dots & \sigma_{1N} \\ \vdots & \vdots \\ \sigma_{N1} & 0 & \dots & 0 & \dots & \sigma_{NN} & 0 & \dots & 0 \\ 0 & \sigma_{N1} & \dots & 0 & \dots & 0 & \sigma_{NN} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \dots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{N1} & \dots & 0 & 0 & \dots & \sigma_{NN} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{11}\mathbf{I}_T & \sigma_{12}\mathbf{I}_T & \dots & \sigma_{1N}\mathbf{I}_T \\ \sigma_{21}\mathbf{I}_T & \sigma_{22}\mathbf{I}_T & \dots & \sigma_{2N}\mathbf{I}_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1}\mathbf{I}_T & \sigma_{N2}\mathbf{I}_T & \dots & \sigma_{NN}\mathbf{I}_T \end{bmatrix} \\ &= \Sigma_N \otimes \mathbf{I}_T \end{aligned}$$

Para ir além na representação Kronecker da matriz de covariâncias, retome o modelo univariado do final da seção anterior, a saber:

$$f(Y|\beta, \mathbf{X}) = \frac{1}{(2\pi)^{\frac{T}{2}} \sqrt{|\sigma^2|}} \exp \left\{ (Y - \mathbf{X}\beta)' \sigma^{-2} (Y - \mathbf{X}\beta) \right\}$$

Agora, é possível estendê-lo de forma a incorporar mais variáveis endógenas  $Y_i$ , para  $i = 1, 2, \dots, n$ . Um modelo popular que relaciona essas variáveis à seus valores passados é conhecido como Vetor Autoregressivo (VAR). O VAR busca captar as relações

lineares existentes entre os  $Y_1$  e existem diferentes parametrizações para tal. Visando focar na f.d.p. normal métrica, considere que:

$$\begin{aligned}
 \mathbf{y}'_t &= \begin{bmatrix} y_{1t} & y_{2t} & \cdots & y_{nt} \end{bmatrix} \\
 \mathbf{e}'_t &= \begin{bmatrix} e_{1t} & e_{2t} & \cdots & e_{nt} \end{bmatrix} \\
 \mathbf{x}'_t &= \begin{bmatrix} y_{1(t-1)} & y_{2(t-1)} & \cdots & y_{n(t-1)} & y_{1(t-2)} & y_{2(t-2)} & \cdots & y_{1(t-p)} & y_{2(t-p)} & \cdots & y_{n(t-p)} \end{bmatrix} \\
 &= \begin{bmatrix} \mathbf{y}'_{t-1} & \mathbf{y}'_{t-2} & \cdots & \mathbf{y}'_{t-p} \end{bmatrix} \\
 \mathbf{B} &= \begin{bmatrix} \beta_{11}^{(1)} & \beta_{21}^{(1)} & \cdots & \beta_{n1}^{(1)} \\ \beta_{12}^{(1)} & \beta_{22}^{(1)} & \cdots & \beta_{n2}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{1n}^{(1)} & \beta_{2n}^{(1)} & \cdots & \beta_{nn}^{(1)} \\ \beta_{11}^{(2)} & \beta_{21}^{(2)} & \cdots & \beta_{n1}^{(2)} \\ \beta_{12}^{(2)} & \beta_{22}^{(2)} & \cdots & \beta_{n2}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{1n}^{(2)} & \beta_{2n}^{(2)} & \cdots & \beta_{nn}^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{11}^{(p)} & \beta_{21}^{(p)} & \cdots & \beta_{n1}^{(p)} \\ \beta_{12}^{(p)} & \beta_{22}^{(p)} & \cdots & \beta_{n2}^{(p)} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{1n}^{(p)} & \beta_{2n}^{(p)} & \cdots & \beta_{nn}^{(p)} \end{bmatrix} \\
 &= \begin{bmatrix} \mathbf{B}'^{(1)} \\ \mathbf{B}'^{(2)} \\ \vdots \\ \mathbf{B}'^{(p)} \end{bmatrix}
 \end{aligned}$$

O modelo então pode ser escrito na forma:

$$\mathbf{y}'_t = \mathbf{x}'_t \mathbf{B} + \mathbf{e}'_t$$

Empilhando os dados obtidos em  $t$  chegamos ao modelo na forma compacta:

$$\begin{bmatrix} \mathbf{y}'_1 \\ \mathbf{y}'_2 \\ \vdots \\ \mathbf{y}'_T \end{bmatrix}_{T \times n} = \begin{bmatrix} \mathbf{x}'_0 \\ \mathbf{x}'_1 \\ \cdots \\ \mathbf{x}'_{T-1} \end{bmatrix}_{T \times np} \times \begin{bmatrix} \mathbf{B}'^{(1)} \\ \mathbf{B}'^{(2)} \\ \vdots \\ \mathbf{B}'^{(p)} \end{bmatrix}_{np \times n} + \begin{bmatrix} \mathbf{e}'_1 \\ \mathbf{e}'_2 \\ \vdots \\ \mathbf{e}'_T \end{bmatrix}_{T \times n}$$

$$\mathbf{Y} = \mathbf{X} \mathbf{B} + \mathbf{E}$$

Mantendo o pressuposto de normalidade do termo de perturbação  $\mathbf{E}$ , a estimação dos parâmetros  $\mathbf{B}$  pode ser feita maximizando a f.d.p. normal multivariada em quanto à  $\mathbf{B}$ :

$$f(\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{nT}{2}} |\boldsymbol{\sigma}^2|^{\frac{T}{2}}} \exp \left\{ -\frac{1}{2} \sum_1^T (\mathbf{y}'_t - \mathbf{x}'_t \mathbf{B})' \boldsymbol{\Sigma}^{-2} (\mathbf{y}'_t - \mathbf{x}'_t \mathbf{B})' \right\}$$

Com foco no núcleo da distribuição acima, a expressão neste em notação matricial pode ser aberta de forma mais conveniente para mostrar os resultados:

$$\begin{aligned} & \sum_1^T (\mathbf{y}'_t - \mathbf{x}'_t \mathbf{B})' \boldsymbol{\Sigma}^{-2} (\mathbf{y}'_t - \mathbf{x}'_t \mathbf{B})' \\ &= \sum_1^T \begin{pmatrix} e_{1t} & e_{2t} & \dots & e_{nt} \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{pmatrix}^{-1} \begin{pmatrix} e_{1t} \\ e_{2t} \\ \dots \\ e_{nt} \end{pmatrix} \\ &= \sum_{t=1}^T \sum_{i=1}^n \sum_{j=1}^n e_{it} e_{jt} \sigma_{ij}^- \end{aligned} \quad (3.14)$$

Onde  $\sigma_{ij}^-$  corresponde a cada elemento da inversa da matriz  $\boldsymbol{\Sigma}$  para todo  $i, j = 1, 2, \dots, n$ .

Ao pegar a expressão compacta  $\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}$  e colocá-la na mesma forma quadrática acima, elimina-se a necessidade do operador de somatório, tem-se então que:

$$\begin{aligned} & (\mathbf{Y} - \mathbf{X}\mathbf{B})' \boldsymbol{\Sigma}^{-2} (\mathbf{Y} - \mathbf{X}\mathbf{B}) \\ &= \begin{pmatrix} \mathbf{e}'_1 \\ \mathbf{e}'_2 \\ \vdots \\ \mathbf{e}'_T \end{pmatrix}_{T \times n} \begin{pmatrix} \sigma_{11}^- & \sigma_{12}^- & \dots & \sigma_{1n}^- \\ \sigma_{21}^- & \sigma_{22}^- & \dots & \sigma_{2n}^- \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1}^- & \sigma_{n2}^- & \dots & \sigma_{nn}^- \end{pmatrix} \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_T \end{pmatrix}_{n \times T} \\ &= \begin{pmatrix} \sum_{i=1}^n \sum_{j=1}^n e_{i1} e_{j1} \sigma_{ij}^- & \sum_{i=1}^n \sum_{j=1}^n e_{i1} e_{j2} \sigma_{ij}^- & \dots & \sum_{i=1}^n \sum_{j=1}^n e_{i1} e_{jT} \sigma_{ij}^- \\ \sum_{i=1}^n \sum_{j=1}^n e_{i2} e_{j1} \sigma_{ij}^- & \sum_{i=1}^n \sum_{j=1}^n e_{i2} e_{j2} \sigma_{ij}^- & \dots & \sum_{i=1}^n \sum_{j=1}^n e_{i2} e_{jT} \sigma_{ij}^- \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n \sum_{j=1}^n e_{iT} e_{j1} \sigma_{ij}^- & \sum_{i=1}^n \sum_{j=1}^n e_{iT} e_{j2} \sigma_{ij}^- & \dots & \sum_{i=1}^n \sum_{j=1}^n e_{iT} e_{jT} \sigma_{ij}^- \end{pmatrix} \end{aligned}$$

O operador traço é responsável por somar todos os elementos da diagonal principal de uma matriz. Supondo que os termos de erro  $\mathbf{E}$  são correlacionados apenas contemporaneamente, apenas os termos da diagonal principal da matriz acima são relevantes, somando-os, chegamos na mesma expressão 3.14:

$$\text{tr}[(\mathbf{Y} - \mathbf{X}\mathbf{B})'\boldsymbol{\Sigma}^{-2}(\mathbf{Y} - \mathbf{X}\mathbf{B})] = \sum_{t=1}^T \sum_{i=1}^n \sum_{j=1}^n e_{it}e_{jt}\sigma_{ij}^-$$

E a f.d.p. pode ser reescrita como:

$$f(\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{nT}{2}}|\sigma^2|^{\frac{T}{2}}} \exp\left\{-\frac{1}{2}\text{tr}[(\mathbf{Y} - \mathbf{X}\mathbf{B})'\boldsymbol{\Sigma}^{-2}(\mathbf{Y} - \mathbf{X}\mathbf{B})]\right\}$$

Sob os mesmos pressupostos feitos para o modelo univariado, a matriz de parâmetros  $\mathbf{B}$  amostral tem distribuição de probabilidade normal multivariada, com ressalva à matriz de covariância que pode ser agora apresentada sob a forma kronecker.

Seja como antes  $\boldsymbol{\Sigma}$  a matriz de covariâncias do termo de erro  $E$  de dimensão  $n \times n$ . Tomando cada equação do VAR individualmente temos que:

$$y_{it} = \beta_i^{(1)}y_{it-1} \sum_j^{n-1} \beta_j^{(1)}y_{jt-1} + \dots + \beta_i^{(p)}y_{it-p} \sum_j^{n-1} \beta_j^{(p)}y_{jt-p} + e_{it} \quad i \neq j$$

A expressão 3.2.1.1 também pode ser escrita como  $y_{it} = \mathbf{x}'_t \mathbf{B}^i + e_{it}$  onde o índice  $i$  agora representa a  $i$ -ésima coluna da matriz  $\mathbf{B}$ . De modo similar, temos  $\mathbf{Y}^{(i)} = \mathbf{X}\mathbf{B}^i + E^{(i)}$ .

Para cada equação do VAR, temos que a variância do conjunto de parâmetros é igual à dada na subseção anterior,  $\sigma_i^-(\mathbf{X}'\mathbf{X})^{-1}$ , que é uma matriz  $np \times np$ . Se os  $n$  termos de erro são correlacionados temporalmente, então suas covariâncias podem ser representadas matricialmente por:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{bmatrix}$$

E a matriz de covariâncias de  $\mathbf{B}$  pode ser representada multiplicando-se cada elemento de  $\boldsymbol{\Sigma}$  pela matriz  $(\mathbf{X}'\mathbf{X})^{-1}$ :

$$\begin{aligned} \boldsymbol{\Sigma} &= \begin{bmatrix} \sigma_{11}(\mathbf{X}'\mathbf{X})^{-1} & \sigma_{12}(\mathbf{X}'\mathbf{X})^{-1} & \dots & \sigma_{1n}(\mathbf{X}'\mathbf{X})^{-1} \\ \sigma_{21}(\mathbf{X}'\mathbf{X})^{-1} & \sigma_{22}(\mathbf{X}'\mathbf{X})^{-1} & \dots & \sigma_{2n}(\mathbf{X}'\mathbf{X})^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1}(\mathbf{X}'\mathbf{X})^{-1} & \sigma_{n2}(\mathbf{X}'\mathbf{X})^{-1} & \dots & \sigma_{nn}(\mathbf{X}'\mathbf{X})^{-1} \end{bmatrix} \\ &= \boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

A distribuição conjunta dos parâmetros tem a seguinte expressão:

$$f(\boldsymbol{\beta}|\mathbf{X}) = \frac{1}{(2\pi)^{\frac{nnT}{2}} |\boldsymbol{\Sigma}|^{\frac{T}{2}}} \exp \left\{ -\frac{1}{2} \left[ (\text{vec}(\mathbf{B}) - \text{vec}(\tilde{\mathbf{B}}))' \boldsymbol{\Sigma}^{-1} \otimes (\mathbf{X}'\mathbf{X}) (\text{vec}(\mathbf{B}) - \text{vec}(\tilde{\mathbf{B}})) \right] \right\}$$

Onde  $\text{vec}()$  é um operador que empilha os dados de uma matriz qualquer coluna a coluna:

$$\text{vec}(\mathbf{B})_{n^2t \times 1} = \begin{bmatrix} \beta_{11}^{(1)} & \beta_{12}^{(1)} & \dots & \beta_{1n}^{(1)} & \beta_{11}^{(2)} & \beta_{12}^{(2)} & \dots & \beta_{1n}^{(2)} \\ \dots & \beta_{11}^{(p)} & \beta_{12}^{(p)} & \dots & \beta_{1n}^{(p)} & \beta_{21}^{(1)} & \beta_{22}^{(1)} & \dots \\ \beta_{2n}^{(p)} & \dots & \beta_{n1}^{(1)} & \beta_{n2}^{(1)} & \dots & \beta_{nn}^{(1)} & \dots & \beta_{n1}^{(p)} & \beta_{n2}^{(p)} & \dots & \beta_{nn}^{(p)} \end{bmatrix}'$$

Então, de forma compacta, os parâmetros  $\mathbf{B} \sim N(\text{vec}(\tilde{\mathbf{B}}), \boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1})$

Tal como ocorrido quando do modelo univariado, os parâmetros do VAR também são escritos na forma da normal métrica. Essa parametrização tem a vantagem de não se utilizar do operador  $\text{vec}$  nem do kronecker:

$$f(\mathbf{B}|\mathbf{X}) = \frac{1}{(2\pi)^{\frac{Tn}{2}} |\boldsymbol{\Sigma}|^{\frac{n}{2}} |\mathbf{X}'\mathbf{X}|^{\frac{-nT}{2}}} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \boldsymbol{\Sigma}^{-1} (\mathbf{B} - \tilde{\mathbf{B}})' (\mathbf{X}'\mathbf{X}) (\mathbf{B} - \tilde{\mathbf{B}}) \right] \right\}$$

A generalização desse tipo de distribuição então pode ser feita de modo que, se  $\mathbf{Y}$  é uma matriz  $T \times N$  cujos valores empilhados tem uma distribuição normal multivariada, cuja matriz de covariância pode ser decomposta em duas matrizes na forma kronecker  $\mathbf{Q} \otimes \mathbf{P}$ , ambas positivas definidas e simétricas de dimensão  $T \times T$  e  $N \times N$  respectivamente, temos que:

$$\text{vec}(\mathbf{Y}) \sim N(\text{vec}(\mathbf{M}), \mathbf{Q} \otimes \mathbf{P})$$

Então, alternativamente,  $\mathbf{Y}$  tem uma distribuição normal métrica, dada pela expressão:

$$\mathbf{Y} \sim MN_{pq}(\mathbf{M}, \mathbf{Q}, \mathbf{P})$$

E a f.d.p. por:

$$f(\mathbf{Y}) = \frac{1}{(2\pi)^{\frac{TN}{2}} |\mathbf{Q}|^{\frac{N}{2}} |\mathbf{P}|^{\frac{T}{2}}} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \mathbf{Q}^{-1} (\mathbf{Y} - \mathbf{M})' \mathbf{P}^{-1} (\mathbf{Y} - \mathbf{M}) \right] \right\}$$

### 3.2.2 Distribuição Wishart

A distribuição Wishart pode ser considerada como uma distribuição  $\chi_n^2$  multivariada, escrita na forma métrica. Retomando conceitos das distribuições univariadas, a distribuição  $\chi_n^2$  é uma forma especial da distribuição  $\Gamma$ .

Considere que  $\Sigma$  e  $\mathbf{S}$  sejam matrizes simétricas positivas definidas nos reais, ambas de dimensão  $n$ . Se  $\mathbf{S}$  for distribuída pela  $\Gamma$  multivariada, segue que sua distribuição tem a forma:

$$f(\mathbf{S}) = \frac{|\Sigma^{-1}|^{\lambda + \frac{n-1}{2}}}{\beta^{\lambda n + \frac{n(n-1)}{2}} \pi^{\frac{n(n-1)}{4}} \prod_{j=1}^p \Gamma(\lambda + \frac{n-j}{2})} \exp \left\{ -\frac{1}{\beta} \text{tr} \Sigma^{-1} \mathbf{S} \right\} |\mathbf{S}|^{\lambda-1}$$

Onde  $\lambda > 0$  e  $\beta > 0$ . Escolhendo  $\alpha = \lambda + \frac{n-1}{2}$ , então a função pode ser reescrita como:

$$f(\mathbf{S}) = \frac{|\Sigma^{-1}|^{\alpha}}{\beta^{n\alpha} \pi^{\frac{n(n-1)}{4}} \prod_{j=1}^p \Gamma(\alpha - \frac{1}{2}(j-1))} \exp \left\{ -\frac{1}{\beta} \text{tr} \Sigma^{-1} \mathbf{S} \right\} |\mathbf{S}|^{\lambda-1}$$

Ainda, definindo a função gama multivariada por Anderson et al. (1958):

$$\Gamma_n(\alpha) = \pi^{\frac{n(n-1)}{4}} \prod_{j=1}^p \Gamma(\alpha - \frac{1}{2}(j-1))$$

A forma geral da distribuição é reescrita como:

$$f(\mathbf{S}) = \frac{|\Sigma|^{-\alpha}}{\beta^{n\alpha} \Gamma_n(\alpha)} \exp \left\{ -\frac{1}{\beta} \text{tr} \Sigma^{-1} \mathbf{S} \right\} |\mathbf{S}|^{\lambda-1}$$

Ou de forma compacta,  $\mathbf{S} \sim \Gamma_n(\alpha, \beta, \Sigma)$ . O caso especial da  $\Gamma$  que é de interesse ocorre quando  $\alpha = \nu/2$  e  $\beta = 2$ , similar à distribuição  $\chi_\nu^2$ , agora a chamamos de Wishart, em homenagem ao trabalho de Wishart (1928), ao formular a distribuição:

$$f(\mathbf{S}) = \frac{|\Sigma|^{-\frac{n}{2}}}{2^{\frac{n\nu}{2}} \Gamma_n(\frac{\nu}{2})} \exp \left\{ -\frac{1}{\beta} \text{tr} \Sigma^{-1} \mathbf{S} \right\} |\mathbf{S}|^{\frac{1}{2}(\nu-n-1)}$$

Generalizando, temos que, se  $\mathbf{Y}_t$  for um vetor i.i.d.  $N(\mathbf{0}, \Sigma)$  de dimensão  $N$  então:

$$\mathbf{S} = \sum_{i=1}^T (\mathbf{Y}_t - \mathbf{M})(\mathbf{Y}_t - \mathbf{M})' \sim W(\Sigma, T)$$

A derivação das distribuições Gama multivariada e Wishart demandam cálculo complexo e serão omitidos nesse trabalho, porém boas referências podem ser encontradas no trabalho original de Wishart ...

### 3.2.2.1 Wishart inversa

De forma genérica, a f.d.p. Wishart inversa é dada pela expressão  $\mathbf{S} \sim iW(\boldsymbol{\Sigma}, T)$  se  $\mathbf{S}^{-1} \sim W(\boldsymbol{\Sigma}^{-1}, T)$ . A derivação dessa função pode ser feita de modo similar à distribuição Gama inversa.

Inicialmente, convém colocar a matriz de parâmetros  $\mathbf{B}$  na sua forma inversa. Da distribuição Wishart, temos que  $\mathbf{A}^{-1} \sim W(\mathbf{B}^{-1}, \nu)$  deve ser escrita como:

$$f(\mathbf{A}^{-1}) = \frac{|\mathbf{B}|^{-\frac{\nu}{2}}}{2^{\frac{n\nu}{2}} \Gamma_n(\frac{\nu}{2})} \exp \left\{ -\frac{1}{2} \text{tr} \mathbf{B}^{-1} \mathbf{A}^{-1} \right\} |\mathbf{A}|^{\frac{1}{2}(-\nu+n+1)}$$

Ao fazer a transformação  $\mathbf{C} = \mathbf{A}^{-1}$ , a inversa da transformação torna-se  $\mathbf{A} = \mathbf{C}^{-1}$  e o jacobiano então  $|\mathbf{C}|^{-n-1}$ :

$$\begin{aligned} f(\mathbf{C}) &= \frac{|\mathbf{B}|^{-\frac{\nu}{2}}}{2^{\frac{n\nu}{2}} \Gamma_n(\frac{\nu}{2})} \exp \left\{ -\frac{1}{2} \text{tr} \mathbf{B}^{-1} \mathbf{C} \right\} |\mathbf{C}|^{\frac{1}{2}(-\nu+n+1)} |\mathbf{C}|^{-n-1} \\ &= \frac{|\mathbf{B}|^{-\frac{\nu}{2}}}{2^{\frac{n\nu}{2}} \Gamma_n(\frac{\nu}{2})} \exp \left\{ -\frac{1}{2} \text{tr} \mathbf{B}^{-1} \mathbf{C} \right\} |\mathbf{C}|^{-\frac{1}{2}(+\nu+n+1)} \end{aligned}$$

A função acima é conhecida como Wishart inversa, também escrita de forma genérica como  $\mathbf{C} \sim iW(\mathbf{B}, \nu)$ . A aplicação mais comum ocorre da distribuição da matriz de covariância amostral  $\mathbf{S}$ , com  $\mathbf{S} \sim iW(\boldsymbol{\Sigma}, \nu)$ :

$$f(\mathbf{S}) = \frac{|\boldsymbol{\Sigma}|^{-\frac{\nu}{2}}}{2^{\frac{n\nu}{2}} \Gamma_n(\frac{\nu}{2})} \exp \left\{ -\frac{1}{2} \text{tr} \mathbf{S}^{-1} \boldsymbol{\Sigma} \right\} |\mathbf{S}|^{-\frac{1}{2}(+\nu+n+1)} \quad (3.15)$$

Para o caso quando a análise bayesiana é feita, convém colocar a matriz de covariância  $\boldsymbol{\Sigma}$  em função de  $\mathbf{S}$ , invertendo a lógica:

$$f(\boldsymbol{\Sigma}) = \frac{|\mathbf{S}|^{-\frac{\nu}{2}}}{2^{\frac{n\nu}{2}} \Gamma_n(\frac{\nu}{2})} \exp \left\{ -\frac{1}{2} \text{tr} \boldsymbol{\Sigma}^{-1} \mathbf{S} \right\} |\boldsymbol{\Sigma}|^{-\frac{1}{2}(+\nu+n+1)} \quad (3.16)$$

### 3.2.3 Distribuição t

A distribuição t multivariada foi desenvolvida inicialmente por Coornish (1954) e sua derivação pode ser vista em seu trabalho, devido à complexidade dos cálculos, será omitida nesse trabalho. Essa distribuição pode ser vista como uma extensão natural da t-student univariada, até mesmo pela sua forma similar, porém algumas diferenças notáveis entre elas surgem ao trabalharmos na forma matricial.

Seja  $\mathbf{X}$  um vetor  $n$ -dimensional, então  $\mathbf{X}$  é distribuída pela t multivariada, com média  $\boldsymbol{\mu}$ , matriz de correlação  $\mathbf{R}$  e  $\nu$  graus de liberdade se sua distribuição conjunta é dada por:

$$f(\mathbf{X}) = \frac{\gamma(\frac{\nu+n}{2})}{(\pi\nu)^{\frac{n}{2}}\gamma(\frac{\nu}{2})|\mathbf{R}|^{\frac{1}{2}}} \left[ 1 + \frac{1}{\nu}(\mathbf{X} - \boldsymbol{\mu})'\mathbf{R}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \right]^{-\frac{\nu+n}{2}}$$

Se  $\boldsymbol{\mu} = \mathbf{0}$ , então a distribuição é chamada de central, caso contrário, não central. Percebe-se que, se  $n = 1$ ,  $\boldsymbol{\mu} = \mathbf{0}$ , e  $\mathbf{R} = 1$ , então a distribuição se reduz à t-student univariada. Outro ponto a ser notado é que, conforme  $\nu \rightarrow \infty$ , então o formato da distribuição tende à normal multivariada padronizada.

Há dois modos de enxergar a representação dessa distribuição, o primeiro ocorre quando um vetor aleatório  $\mathbf{Y}$   $n$ -dimensional tem distribuição  $\mathbf{Y} \sim (\mathbf{0}, \boldsymbol{\Sigma})$  e a razão da covariância amostral pela populacional  $(\nu S^2)/\sigma^2$  tem distribuição  $\chi^2_\nu$ , então  $\mathbf{X}$  tem distribuição t multivariada se:

$$\mathbf{X} = S^{-1}\mathbf{Y} + \boldsymbol{\mu}$$

Por essa representação, temos que  $\mathbf{X}|S \sim N(\boldsymbol{\mu}, (1/s^2)\boldsymbol{\Sigma})$ .

Outra representação, mais comum, ocorre quando  $\mathbf{S}$  é uma matriz simétrica definida positiva tal que a decomposição  $\mathbf{S} = \mathbf{S}^{\frac{1}{2}}\mathbf{S}^{\frac{1}{2}}$  existe e  $\mathbf{S} \sim W_n(\mathbf{R}, \nu + p - 1)$ , então:

$$\mathbf{X} = (\mathbf{V}^{\frac{1}{2}})^{-1}\mathbf{Y} + \boldsymbol{\mu}$$

Dessa relação, obtém-se que  $\mathbf{X}|\mathbf{V} \sim (\boldsymbol{\mu}, \nu\mathbf{S}^{-1})$ .

Essa representação é comum na inferência bayesiana quando, da regressão  $\mathbf{Y}' = \mathbf{X}'\mathbf{B} + \mathbf{E}$ , onde  $\mathbf{Y}$  e  $\mathbf{X}$  são vetores de variáveis endógenas e exógenas com dimensão  $n$  e  $np$  respectivamente,  $\mathbf{E}$  é um vetor  $n$ -dimensional de erros aleatórios, e a matriz de parâmetros  $\mathbf{B}$  tem dimensão  $np \times n$ , a distribuição dos parâmetros  $\mathbf{B}$  pode então ser feita de modo que  $\text{vec}(\mathbf{B}) \sim t_\nu$ .

### 3.2.3.1 Distribuição t matricial

Essa distribuição pode ser considerada em relação à t multivariada de forma similar à relação da normal matricial e a normal multivariada. Foi elaborada por Dickey (1966) e, devido à complexidade dos cálculos, será omitida aqui.

É dito que, sendo  $\mathbf{X}$  uma matriz  $p \times q$ , esta tem distribuição t matricial com parâmetros  $\boldsymbol{\mu}$ ,  $\mathbf{R}$  e  $\mathbf{Q}$  se a f.d.p. tem a seguinte forma:

$$f(\mathbf{X}) = \frac{\Gamma_q(\frac{\nu+p}{2})}{\pi^{\frac{pq}{2}}\Gamma_q(\frac{\nu}{2})|\mathbf{Q}|^{\frac{\nu}{2}}|\mathbf{R}|^{-\frac{q}{2}}|\mathbf{Q} + (\mathbf{X} - \boldsymbol{\mu})'\mathbf{R}^{-1}(\mathbf{X} - \boldsymbol{\mu})|^{-\frac{\nu+p}{2}}}$$

Onde  $\boldsymbol{\mu}$  é uma matriz  $p \times q$  e  $\mathbf{R}$  e  $\mathbf{Q}$  são matrizes ambas positivas definidas de dimensão  $p \times p$  e  $q \times q$  respectivamente. Novamente, tal como para o caso da t

multivariada, se  $\boldsymbol{\mu} = \mathbf{0}$ , então a distribuição é chamada de central. Ainda, se  $\mathbf{R} = \mathbf{I}$ , então é uma mistura da normal padronizada com média  $\mathbf{0}$  e matriz de covariância  $\mathbf{I} \otimes \mathbf{V}$ .

É interessante notar ainda que, se  $p = 1$  ou  $q = 1$ , a distribuição se reduz à  $t$  multivariada. Logo, esta última é um caso especial da distribuição  $t$  matricial.

## 4 VETOR AUTORREGRESSIVO

O vetor autorregressivo, ou VAR, é um dos modelos mais consagrados na literatura em modelagem de séries temporais, e seu uso é bem recorrente sobretudo nas pesquisas em macroeconomia. Seu sucesso foi conquistado pela sua capacidade em relacionar linearmente quaisquer variáveis endógenas que entram no modelo e assim fazer previsões boas  $h$  passos a frente. Através dela é também permitido que se calcule a função impulso-resposta, responsável por prever qual o comportamento das variáveis do modelo ante um choque ocorrido em alguma delas.

Como proposto por Sims (1980), o VAR( $p$ ) é considerado uma generalização da modelagem de séries univariadas AR( $p$ ), onde  $p$  representa o número de defasagens do modelo. Para apresentá-lo inicialmente será dada uma breve explicação sobre modelagens de séries univariadas tipo ARMA( $p,q$ ), da qual é possível obter uma intuição melhor de como funciona esse tipo de modelagem e sua motivação à aplicação em séries multivariadas. De acordo com a proposta do trabalho, não será apresentado a versão mais genérica VARMA( $p,q$ ), uma vez que não será necessária sua utilização. Contudo, para o modelo VAR( $p$ ), é dada uma introdução da versão clássica, e logo após é introduzida a versão Bayesiana para tal tipo de modelagem. Boas referências desse modelo podem ser encontradas nos livros de econometria de séries temporais, como em Hamilton (1994).

### 4.1 Modelo Autorregressivo de Média Móvel

Como descrito, o modelo VAR pode ser entendido como uma generalização de um modelo autorregressivo AR( $p$ ), com  $p$  defasagens. Considere um modelo mais genérico Autorregressivo de Média Móvel, ARMA( $p,q$ ), amplamente utilizado para modelagens de séries temporais econômicas. Esse modelo é uma consequência direta de Wold (1938), que estabelece que quaisquer processos não determinísticos, estacionários em segunda ordem, podem ser representados como uma soma infinita de médias móveis:

$$y_t = e_t + \sum_{i=1}^{\infty} a_i e_{t-i}$$

Onde  $e_t$  é uma inovação no processo, com  $E(e_t) = 0$  para todo  $t = 1, \dots, T$ , ademais,  $e_t$  segue as propriedades de um ruído branco. De outra forma, considere que:

$$y_t(q) = e_t + \sum_{i=1}^q a_i e_{t-i}$$

Então a relação é válida:

$$\|y_t(q) - y_t\|^2 = E(e_t^2 \sum_{i>q} a_i^2) \rightarrow 0, \quad q \rightarrow \infty$$

A estimação de um modelo MA(q) com  $q$  suficientemente grande afim de se ajustar bem ao processo torna a estimação dos parâmetros  $a_i$  custosa em termos estatísticos. Deste modo, é mais conveniente botar termos defasados  $y_{t-p}$  de modo que o processo possa ser explicado por ambos  $y_{t-p}$  e  $e_{t-q}$ .

Então, um processo estacionário de segunda ordem segue um ARMA(p,q), para  $p$  e  $q$  nos inteiros, se:

$$y_t + \sum_{j=1}^p b_j y_{t-j} = c + e_t + \sum_{i=1}^q b_i e_{t-1}$$

#### 4.1.1 Estimação

Suponha que o processo de  $y_t$  seja dado por um ARMA(p,q), então sob suposição de homoscedasticidade do termo de erro  $e_t \sim N(0, \sigma^2)$ , o modelo pode ser completamente definido por:

$$y_t - c = \sum_{j=1}^p b_j (y_{t-j} - c) + e_t - \sum_{i=1}^q a_i e_{t-j}$$

De forma compacta, o vetor de parâmetros pode ser escrito como:

$$\boldsymbol{\phi} = [c \ a_1 \ \dots \ a_q \ b_1 \ \dots \ b_p]'$$

Então o espaço paramétrico é definido como:

$$\boldsymbol{\Phi} \subset \mathbb{R}^{P=Q+1}$$

Condicionada aos valores iniciais  $[y_0, \dots, y_{1-p+q}, e_q, \dots, e_1, \sigma_0^2]$  a função de quase-verossimilhança é dada como:

$$L_n(\boldsymbol{\phi}) = L_n(\boldsymbol{\phi}; \tilde{e}_T) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{e_t^2}{2\sigma_t^2}\right)$$

E o estimador de Quase-Máxima Verossimilhança, ou *Quasi-Maximum Likelihood Estimator* (QMLE)<sup>1</sup> do vetor  $\boldsymbol{\theta}$  é dada pela maximização:

<sup>1</sup> Esse estimador é amplamente utilizado pelas suas propriedades assintóticas. Por muitas vezes é necessário aproximar a f.d.p. dos dados à outra função que tem as mesmas características, tal como o caso quando dados de séries financeiras são modelados. Sabe-se que estes não advêm de uma distribuição normal gaussiana, porém o uso desta apresenta-se como uma boa aproximação para tal.

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} L_n(\theta)$$

E a maximização pode ser feita por métodos numéricos.

## 4.2 O Vetor Autorregressivo

Até agora, os modelos apresentados são univariados, ou seja, o comportamento apresentado pela série pode ser explicado por ela mesma sem que haja a necessidade da observação de outras variáveis. Sabe-se que em economia, grande parte dos fenômenos são explicados por diversas variáveis, o que torna os modelos univariados limitados de forma geral. O Vetor Autorregressivo (VAR) pode ser considerado uma generalização dos modelos AR(p) no sentido que as séries econômicas podem ser explicadas por elas mesmas e por outras. Esse tipo de abordagem econométrica pode ser escrita de forma simples algebricamente como:

$$\mathbf{A}\mathbf{Y}_t = \mathbf{B}_0 + \sum_{m=1}^p \mathbf{B}_m \mathbf{Y}_{t-m} + \sum_{o=1}^r \mathbf{H}_o \mathbf{Z}_{t-o} + \mathbf{e}_t \quad (4.1)$$

Onde:

- a)  $\mathbf{Y}$  é um vetor de variáveis endógenas  $n \times 1$  variando no tempo  $t$ ,  $\mathbf{Z}$  é um vetor de variáveis endógenas  $s \times 1$ , igualmente variante em  $t$  e  $\mathbf{e}$  é um vetor  $n \times n$  de choques aleatórios não correlacionados com o vetor  $\mathbf{Z}$  e  $\mathbf{e}_t \sim i.i.d.(\mathbf{0}, \mathbf{I}_n)$ ;
- b)  $\mathbf{A}$  é uma matriz  $n \times n$  que relaciona as variáveis endógenas  $\mathbf{Y}$  contemporaneamente, os elementos da diagonal principal geralmente assumem o valor unitário, ou seja,  $\alpha_{ij} = 1$  para todo  $i = j$ ;
- c)  $\mathbf{B}_0$  é um vetor coluna  $n \times 1$  de constantes;
- d)  $\mathbf{B}_m$  é um conjunto de matrizes  $n \times n$  que associam de forma linear o vetor de variáveis endógenas  $\mathbf{Y}$  com seus respectivos valores defasados até a ordem  $p$ ;
- e)  $\mathbf{H}_o$  é um conjunto de matrizes  $n \times s$  que associam de forma linear o vetor de variáveis exógenas  $\mathbf{Z}$  defasados até a ordem  $r$  com o vetor de variáveis endógenas  $\mathbf{Y}$ .

A forma apresentada acima é também conhecida como VAR estrutural. Nesse modelo, a matriz  $\mathbf{A}$  é responsável por captar todas as relações contemporâneas entre as variáveis  $\mathbf{Y}$ , restando aos choques afetar individualmente cada uma das variáveis. Há ainda outra forma de reescrever o VAR, chamada de forma reduzida:

$$\mathbf{Y}_t = \mathbf{A}^{-1}\mathbf{B}_0 + \sum_{m=1}^p \mathbf{A}^{-1}\mathbf{B}_m \mathbf{Y}_{t-m} + \sum_{o=1}^r \mathbf{A}^{-1}\mathbf{H}_o \mathbf{Z}_{t-o} + \mathbf{A}^{-1}\mathbf{e}_t \quad (4.2)$$

$$= \phi_0 + \sum_{m=1}^p \phi_m \mathbf{Y}_{t-m} + \sum_{o=1}^r \psi_o \mathbf{Z}_{t-o} + \mathbf{A}^{-1}\mathbf{e}_t \quad (4.3)$$

A forma reduzida é bem popular na literatura por conta da estimação que, sob algumas restrições, é possível retornar aos parâmetros originais.

Por simplicidade, ainda, a partir de então as variáveis exógenas e o termo de constante serão omitidas, sem perda de generalidade. Suponha um modelo bivariado cujo processo estocástico é dado por um VAR(1):

$$\mathbf{A}_{(2 \times 2)} \mathbf{Y}_{t(2 \times 1)} = \mathbf{B}_{(2 \times 2)} \mathbf{Y}_{t-1(2 \times 1)}$$

$$\begin{pmatrix} 1 & \alpha_{12} \\ \alpha_{21} & 1 \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$$

Abrindo o sistema acima, o VAR estrutural pode ser facilmente reescrito como:

$$y_{1t} = -\alpha_{12} y_{2t} + \beta_{11} y_{1t-1} + \beta_{12} y_{2t-1} + e_{1t}$$

$$y_{2t} = -\alpha_{21} y_{1t} + \beta_{21} y_{1t-1} + \beta_{22} y_{2t-1} + e_{2t}$$

Substituindo a segunda equação na primeira obtém-se:

$$y_{1t} = -\alpha_{12}(-\alpha_{21} y_{1t} + \beta_{21} y_{1t-1} + \beta_{22} y_{2t-1} + e_{2t}) + \beta_{11} y_{1t-1} + \beta_{12} y_{2t-1} + e_{1t} \quad (4.4)$$

$$(1 - \alpha_{12} \alpha_{21}) y_{1t} = (\beta_{11} - \alpha_{12} \beta_{21}) y_{1t-1} + (\beta_{12} - \alpha_{12} \beta_{22}) y_{2t-1} + (e_{1t} - \alpha_{12} e_{2t}) \quad (4.5)$$

$$y_{1t} = \frac{\beta_{11} - \alpha_{12} \beta_{21}}{1 - \alpha_{12} \alpha_{21}} y_{1t-1} + \frac{\beta_{12} - \alpha_{12} \beta_{22}}{1 - \alpha_{12} \alpha_{21}} y_{2t-1} + \frac{e_{1t} - \alpha_{12} e_{2t}}{1 - \alpha_{12} \alpha_{21}} \quad (4.6)$$

$$y_{1t} = \beta_{11}^* y_{1t-1} + \beta_{12}^* y_{2t-1} + e_{1t}^* \quad (4.7)$$

A manipulação feita acima pode ser feita para VAR(p) com  $p > 1$  e de dimensão  $n > 2$ . Analisando a equação 4.7, fica claro que os choques estruturais  $\mathbf{e}$  relacionam-se com os regressores através da matriz  $\mathbf{A}$ . É sabido que as técnicas de estimação impõem que os erros sejam independentes dos regressores, ou seja,  $E(\mathbf{Y}_{t-1} \mathbf{e}_t^*) = 0$ . Para que seja feita a correta estimação, geralmente são impostas algumas restrições no VAR a fim de que haja a correta identificação dos parâmetros. Nesse trabalho não serão aprofundadas técnicas de restrição no VAR nesse sentido, será imposto daqui por diante que a matriz

$\mathbf{A}$  é uma matriz identidade de dimensão  $n$ ,  $\mathbf{I}_n$ , o que restringe as variáveis endógenas a dependerem exclusivamente de seus valores defasados. Em outras palavras, restringindo  $\mathbf{A} = \mathbf{I}$ , onde  $\mathbf{I}$  é uma matriz identidade, já não há quaisquer relações contemporâneas entre as variáveis.

#### 4.2.1 Estimação

Para exemplificar o método clássico de estimação, considere ainda quando são impostas algumas suposições adicionais sobre o modelo. A primeira dela é a inexistência de quaisquer relações contemporâneas entre as variáveis, o que reduz a matriz  $\mathbf{A} = \mathbf{I}_n$ , e iguala o VAR estrutural à sua forma reduzida. A segunda suposição diz a respeito do termo de erro aleatório, os quais são distribuídos pela normal gaussiana.

$$\mathbf{y}_t = \sum_{m=1}^p \phi \mathbf{y}_{t-m} + \epsilon_t, \quad \epsilon_t \sim N(\mathbf{0}, \Sigma) \quad (4.8)$$

Para estimar os parâmetros  $\phi$ , é necessária uma amostra de tamanho  $p + T$ . Reescrevendo de forma compacta toda a amostra  $T$  como  $\tilde{\mathbf{y}}_T = (\mathbf{y}_T, \mathbf{y}_{T-1}, \dots, \mathbf{y}_1)$ , e as variáveis e os parâmetros como:

$$\mathbf{Y}_t \equiv \begin{bmatrix} \mathbf{y}_{t-1} \\ \mathbf{y}_{T-2} \\ \vdots \\ \mathbf{y}_{t-p} \end{bmatrix} \quad \mathbf{\Pi} \equiv \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}$$

Podemos então reescrever a distribuição condicional de  $\mathbf{y}_t$  como uma média dada por  $\mathbf{\Pi}'\mathbf{Y}_t$  mais a variância em torno desta, dada pela matriz  $\Gamma$ , por conta do termo de perturbação  $\epsilon_t$ :

$$\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_{-p+1} \sim N(\mathbf{\Pi}'\mathbf{Y}_t, \Omega) \quad (4.9)$$

A densidade condicional de  $\mathbf{y}$  é escrita pela expressão:

$$f_{\mathbf{Y}_t | \tilde{\mathbf{Y}}_{t-1}, \dots, \mathbf{Y}_{-p+1}}(\mathbf{y}_t | \tilde{\mathbf{y}}_{t-1}, \dots, \mathbf{y}_{-p+1}; \boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Omega|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{y}_t - \mathbf{\Pi}'\mathbf{Y}_t)' \Gamma^{-1}(\mathbf{y}_t - \mathbf{\Pi}'\mathbf{Y}_t)\right]$$

Para chegar na função densidade conjunta, basta multiplicar as  $T$  distribuições condicionais de  $\mathbf{y}_t$ :

$$f_{\tilde{\mathbf{Y}}_t|\tilde{\mathbf{Y}}_{t-1},\dots,\mathbf{Y}_{-p+1}}(\tilde{\mathbf{y}}_t|\tilde{\mathbf{y}}_{t-1},\dots,\mathbf{y}_{-p+1};\boldsymbol{\theta}) \quad (4.10)$$

$$= \prod_{t=1}^T \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Omega}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{Y}_t)'\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{Y}_t)\right] \quad (4.11)$$

Deixando os parâmetros a serem estimados variarem em função da amostra, obtemos a função de verossimilhança  $L(\boldsymbol{\theta}; \tilde{\mathbf{y}}_t)$ , que é popularmente escrita em sua forma logarítmica:

$$\begin{aligned} L(\boldsymbol{\theta}; \tilde{\mathbf{y}}_t) &= \sum_{t=1}^T \log f_{\tilde{\mathbf{Y}}_t|\tilde{\mathbf{Y}}_{t-1},\dots,\mathbf{Y}_{-p+1}}(\tilde{\mathbf{y}}_t|\tilde{\mathbf{y}}_{t-1},\dots,\mathbf{y}_{-p+1};\boldsymbol{\theta}) \\ &= -\left(\frac{Tn}{2}\right) \log(2\pi) - \left(\frac{T}{2}\right) \log|\boldsymbol{\Omega}| - \left(\frac{1}{2}\right) \sum_{t=1}^T [(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{Y}_t)'\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{Y}_t)] \end{aligned} \quad (4.12)$$

A estimação do modelo é feita derivando-se a função de log-verossimilhança pelos respectivos parâmetros e igualando-a a zero. Há dois conjuntos de parâmetros a serem maximizados:

$$\frac{\partial L(\boldsymbol{\theta}; \tilde{\mathbf{y}}_t)}{\partial \boldsymbol{\Pi}} = 0 \quad \frac{\partial L(\boldsymbol{\theta}; \tilde{\mathbf{y}}_t)}{\partial \boldsymbol{\Omega}} = 0$$

A primeira derivada pode ser facilmente obtida através da regra da cadeia, derivando  $L(\boldsymbol{\theta}; \tilde{\mathbf{y}}_t)$  em relação a  $\boldsymbol{\Pi}$ , a matriz  $\boldsymbol{\Omega}$  passa a ser considerada uma constante, podendo ser ignorada sem perda de informação:

$$\begin{aligned} \left(\frac{1}{2}\right) \sum_{t=1}^T 2(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{Y}_t)'\mathbf{Y}_t' &= 0 \\ \sum_{t=1}^T [\mathbf{y}_t\mathbf{Y}_t' - \boldsymbol{\Pi}'\mathbf{Y}_t\mathbf{Y}_t'] &= 0 \\ \boldsymbol{\Pi}' \sum_{t=1}^T \mathbf{Y}_t\mathbf{Y}_t' &= \sum_{t=1}^T \mathbf{y}_t\mathbf{Y}_t' \\ \boldsymbol{\Pi}' &= \left[ \sum_{t=1}^T \mathbf{y}_t\mathbf{Y}_t' \right] \left[ \sum_{t=1}^T \mathbf{Y}_t\mathbf{Y}_t' \right]^{-1} \end{aligned} \quad (4.13)$$

Através da expressão acima, fica claro que o estimador de Máxima Verossimilhança para  $\boldsymbol{\Pi}$  iguala-se ao estimador por Mínimos Quadrados. A matriz  $\boldsymbol{\Pi}$  é um arranjo do tipo  $[n \times np]$ , não havendo variáveis exógenas e restringindo o termo de constante a  $\mathbf{0}$ . Cada linha de  $\boldsymbol{\Pi}$  equivale ao vetor de parâmetros  $[\phi_{i1}, \phi_{i2}, \dots, \phi_{ip}]$  relativo à variável  $y_i$ , para todo  $i$  no intervalo  $[1, n]$ .

Derivando-se a matriz  $\mathbf{\Gamma}$  e igualando-se a  $\mathbf{0}$  obtém-se:

$$\begin{aligned}
& - \left( \frac{T}{2\Omega} \right) + \frac{1}{2\Omega^2} \sum_{t=1}^T [(\mathbf{y}_t - \mathbf{\Pi}'\mathbf{Y}_t)'(\mathbf{y}_t - \mathbf{\Pi}'\mathbf{Y}_t)] = 0 \\
& - \left( \frac{T}{2\Omega} \right) + \frac{1}{2\Omega^2} \sum_{t=1}^T \epsilon_t' \epsilon_t = 0 \\
& \left( \frac{T}{2\Omega} \right) = \frac{1}{2\Omega^2} \sum_{t=1}^T \epsilon_t' \epsilon_t \\
& \Omega = \frac{1}{T} \sum_{t=1}^T \epsilon_t' \epsilon_t
\end{aligned} \tag{4.14}$$

### 4.3 VAR Bayesiano

A seguir serão abordados tópicos referentes à ótica Bayesiana do VAR, tal como seu estimador, obtido pela amostragem de Gibbs.

#### 4.3.1 A abordagem Bayesiana

Para introduzir a abordagem Bayesiana ao VAR, primeiramente deve-se ater às diferenças entre as inferências clássica e Bayesiana. Suponha-se o seguinte modelo de regressão linear conhecido:

$$y_t = \mathbf{X}_t \boldsymbol{\beta} + \epsilon_t, \quad \epsilon \sim N(0, \sigma^2) \tag{4.15}$$

Por questões de simplificação, aqui temos uma única variável dependente  $y$  contra  $K$  variáveis explicativas  $X$ . A equação 4.15 pode ser reescrita como  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ , de forma que  $\mathbf{y}$  e  $\boldsymbol{\epsilon}$  são vetores  $T \times 1$  e  $\mathbf{X}$  matriz  $T \times K$ ,  $\boldsymbol{\beta}$  e  $\sigma^2$  são fixos ao longo de  $T$ . A inferência clássica diz que a amostra coletada advém de uma população cujos parâmetros são fixos, porém desconhecidos. Ao coletar diversas amostras dessa população, digamos  $J$  coletas de  $\mathbf{y}$  e  $\mathbf{X}$ , obtém-se  $J$  estimativas para  $\boldsymbol{\beta}$  de forma que  $\mathbf{y}^j = \mathbf{X}^j \hat{\boldsymbol{\beta}}^j + \hat{\boldsymbol{\epsilon}}^j$ . No limite, quando  $j \rightarrow \infty$ , a médias do  $\hat{\boldsymbol{\beta}}^j$  tende aos valores dos verdadeiros parâmetros.

Em economia, ao trabalharmos com séries temporais, incorremos geralmente no fato de que há apenas uma amostra disponível, e é de interesse que esta seja o mais próximo dos valores verdadeiros dos parâmetros. A inferência clássica, então, se dá na mensuração da probabilidade dos valores estimados dos parâmetros  $\boldsymbol{\beta}$  estarem próximos dos parâmetros populacionais, de onde a amostra foi retirada.

Em contraste com a inferência clássica, a Bayesiana trata os parâmetros populacionais como aleatórios, impondo uma distribuição de probabilidade para tais. Essa distribuição deve representar o conhecimento *a priori* sobre os parâmetros populacionais e o quão certo estamos a respeito desse conhecimento. A partir do momento que uma

amostra da população é retirada, a distribuição *priori* é atualizada a fim de incorporar a nova informação. Essa abordagem é feita com base no teorema de Bayes:

$$\begin{aligned}
 f(Y|\theta) &= \frac{h(Y, \theta)}{g(\theta)}, & \text{rearranjando os termos:} \\
 h(Y, \theta) &= f(Y|\theta)g(\theta) \\
 &= p(\theta|Y)m(Y) \\
 p(\theta|Y) &= \frac{f(Y|\theta)g(\theta)}{m(Y)} & (4.16)
 \end{aligned}$$

A interpretação da equação 4.16 é direta, se  $\theta$  é o conjunto de parâmetros populacionais e  $Y$  é a mostra obtida, a probabilidade de obter-se os parâmetros  $\theta$  dado que a amostra  $Y$  foi retirada equivale à multiplicação da probabilidade da amostra  $Y$  ser retirada, dado que os parâmetros são  $\theta$  e a distribuição do conjunto paramétrico  $\theta$ . Aqui a distribuição da amostra dada pela função de probabilidade  $f(Y)$  pode ser ignorada, uma vez que ela não depende dos parâmetros, ela pode ser considerada uma constante e descartada sem perda relevante de informação sobre  $\theta$ :

$$p(\theta|Y) \propto f(Y|\theta)g(\theta) \quad (4.17)$$

Em 4.17, a f.d.p.  $p(\theta|Y)$  é chamada de *a posteriori*, ou seja, a distribuição de  $\theta$  atualizada em relação à  $Y$ , a distribuição  $g(\theta)$  é a distribuição *a priori* dos parâmetros. Na prática, a f.d.p.  $f(Y|\theta)$  é reescrita em forma da função de verossimilhança:

$$p(\theta|Y) \propto L(\theta|Y)g(\theta) \quad (4.18)$$

Alguns pontos merecem ser ressaltados quanto à abordagem Bayesiana. O que pode parecer uma vantagem em adotar-se uma distribuição *a priori* para os parâmetros, pode ser uma desvantagem em termos de custos computacionais. Nem sempre é possível obter uma forma analítica para a f.d.p. *posteriori*, caso positivo, diz-se que a função de verossimilhança e a *priori* são conjugadas, caso negativo, incorremos em métodos de simulação para que a maximização da *posteriori* em relação à  $\theta$  seja feita. O segundo aspecto a ser notado tange à especificação da *priori*, ela é feita subjetivamente pelo pesquisador com base em algum conhecimento prévio. Caso não há muita informação *a priori* sobre a distribuição dos parâmetros, a *priori* é chamada de difusa pois não é informativa, e existem métodos para incorporar essa ausência de informação à estimação.

### 4.3.2 Markov Chain Monte Carlo

De modo geral, quando dispomos de todas as distribuições, ou seja, das *posteriors* condicionais e marginais, a inferência Bayesiana pode ser feita de modo direto. Na prática,

geralmente essas distribuições não são dispostas numa forma analítica tratável. Deste modo, dois procedimentos principais podem ser utilizados, o primeiro dele, conhecido como *importance sampling*, o qual não será abordado neste trabalho, e o segundo são estimadores do tipo *Markov Chain Monte Carlo*, também conhecido como MCMC.

Este último método requer que sejam gerados valores aleatórios a partir da distribuição condicional  $f(\boldsymbol{\theta}^{(j+1)}|\boldsymbol{\theta}^{(j)})$ . Se a distribuição for estacionária, então o vetor inicial de parâmetros  $\boldsymbol{\theta}^{(0)}$  pode ser gerado a partir da *posteriori* e então todos os demais  $\boldsymbol{\theta}^{(j)}$ , para  $j = 1, \dots, J$  poderão ser obtidos com essa mesma distribuição. Como na prática, o uso desse método é vantajoso para o caso quando não dispomos da forma analítica da *posteriori*, o vetor inicial de parâmetros  $\boldsymbol{\theta}^{(0)}$  parte de um ponto arbitrário do espaço paramétrico e nasce então o problema da convergência. Ao passo que os valores  $\boldsymbol{\theta}^{(j)}$  vão sendo gerados, estes convergem à verdadeira distribuição dos parâmetros, mas o número de iterações necessárias a convergência é o real problema.

É possível, então, assumir condições para a convergência de modo que  $p(\boldsymbol{\theta}^{(j+1)} \in \Theta|\boldsymbol{\theta}^{(j)})$ , ou seja, o novo valor gerado em  $j + 1$  encontra-se restrito no espaço paramétrico condicionado ao valor gerado em  $j$ . Se pudermos gerar infinitos valores  $\boldsymbol{\theta}^{(j)}$  tal que  $j \rightarrow \infty$ , a convergência é garantida. Como na prática fazer isto é irreal, a aproximação então é feita de modo a gerar parâmetros com  $J$  suficientemente grande a fim de que se tenha uma boa convergência, sob a pena de estimações imprecisas caso contrário. Quando gerados esses  $J$  valores suficientemente grande, os  $B$  primeiros devem ser descartados, período o qual também é conhecido como *burn-in*.

Não há uma regra clara para a definição de quantos valores dos parâmetros devem ser gerados, intuitivamente, se o pesquisador já tiver uma boa experiência sobre o provável valor assumido pelos parâmetros, este assumirá  $\boldsymbol{\theta}^{(0)}$  de modo que esteja próximo da média da distribuição *posteriori*. De outra forma, a convergência ainda está intrinsecamente relacionada ao quão auto-correlacionados estão esses valores gerados. Caso essa auto-correlação seja alta, a convergência será mais lenta e serão necessários mais iterações a fim de que o *burn-in* seja suficientemente grande a ponto de dar estimações precisas dos verdadeiros parâmetros.

#### 4.3.2.1 A amostragem de Gibbs

Esse tipo de amostragem é uma simulação tipo Monte Carlo feita em Cadeias de Markov (*Markov Chain Monte Carlo*, MCMC), e encontra-se dentro da classe de estimadores discutida acima. Sua motivação, de igual modo, reside no fato de que nem sempre é possível transformar a *posteriori* numa forma analítica tratável do ponto de vista matemático, a fim de que a inferência estatística sobre os parâmetros seja feita. Para elucidar

sua utilidade, suponhamos, por simplificação, o seguinte modelo:

$$y_t = \mathbf{X}_t\boldsymbol{\beta} + \epsilon, \quad \epsilon \sim N(0, \sigma^2) \quad (4.19)$$

Grande parte dos modelos em economia pressupõe que os erros  $\epsilon$  seguem uma distribuição normal. O modelo acima possui dois grupos de parâmetros a serem estimados, digamos  $\theta = \boldsymbol{\beta}, \sigma^2$ . Se nenhum desses parâmetros for dado, a abordagem Bayesiana diz que devemos ter um consentimento antes da amostra a respeito do comportamento de  $\theta$  representado pela distribuição da *priori*, que no presente caso, é dada pela f.d.p. conjunta:

$$p(\boldsymbol{\beta}, \sigma^2) = p(\boldsymbol{\beta}|\sigma^2)p(\sigma^2) \quad (4.20)$$

Após a observação da amostra, a *priori* 4.20 deve ser atualizada de forma que a *posteriori* é dada:

$$p(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) = p(\boldsymbol{\beta}, \sigma^2)L(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) \quad (4.21)$$

$$= p(\boldsymbol{\beta}|\sigma^2, \mathbf{y})p(\sigma^2|\mathbf{y}) \quad (4.22)$$

Para que seja feita a inferência sobre os parâmetros, é necessário ter em mãos as distribuições marginais da *posteriori*:

$$p(\boldsymbol{\beta}|\mathbf{y}) = \int_0^\infty p(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) d\sigma^2 \quad (4.23)$$

Integrar 4.23 nem sempre é direto por não ter uma forma analítica, o algoritmo de Gibbs permite-nos reamostrar, com base nas distribuições condicionais da *posteriori*, e então conseguir via simulação a distribuição das marginais. Para tal, supomos valores iniciais para os parâmetros  $\boldsymbol{\beta}^0$  e  $\{\sigma^2\}^0$ :

- a) com base no valor inicial  $\{\sigma^2\}^0$ , temos a distribuição da *posteriori*  $p(\boldsymbol{\beta}|\{\sigma^2\}^0, \mathbf{y})$  e, através desta, obtemos um valor simulado  $\boldsymbol{\beta}^1$ ;
- b) com  $\boldsymbol{\beta}^1$  em mãos, a distribuição da *posteriori* para  $\{\sigma^2\}$  será  $p(\sigma^2|\boldsymbol{\beta}^1, \mathbf{y})$ . Com essa distribuição em mãos, geramos um valor simulado  $\{\sigma^2\}^1$ ;
- c) com base no valor obtido no item anterior  $\{\sigma^2\}^1$ , temos a distribuição da *posteriori*  $p(\boldsymbol{\beta}|\{\sigma^2\}^1, \mathbf{y})$  e, através desta, novamente obtemos um valor simulado, agora chamado  $\boldsymbol{\beta}^2$ ;
- d) com  $\boldsymbol{\beta}^2$ , a distribuição da *posteriori* para  $\{\sigma^2\}$  fica  $p(\sigma^2|\boldsymbol{\beta}^2, \mathbf{y})$ . Novamente, geramos um valor simulado, agora chamado  $\{\sigma^2\}^2$ ;
- e) esse processo deve-se repetir  $J$  vezes a fim de que haja convergência.

Repetindo o processo descrito acima  $J = K + M$  vezes, devemos descartar os  $K$  primeiros valores de todos os parâmetros e inferimos a respeito dos demais  $M$  valores. Esse descarte deve ser feito no ponto da iteração onde já não há grandes modificações nos parâmetros do modelo, significando que o algoritmo de Gibbs convergiu à distribuição da *posteriori*. Em posse dos  $M$  valores, a inferência é feita com base na distribuição marginal empírica  $(\theta^1, \theta^2, \dots, \theta^M)$ , ou seja, por exemplo, a média de  $\beta_i$ , para todo  $i$ , pode ser feita com  $(1/m) \sum_{m=1}^M \beta_i^m$ .

#### 4.4 Escolha das *Priori*

A escolha das distribuições *priori* torna a abordagem Bayesiana, atrativa em termos de flexibilidade na modelagem, redução do problema da dimensionalidade e incorporação de outros tipos de informação que não a da amostra. Por outro lado, a escolha dessas distribuições exigem cuidado do pesquisador quanto aos custos computacionais em relação à estimação e inferência estatística.

É natural que, quanto da escolha das *priori*, priorize-se distribuições que expliquem o comportamento dos parâmetros e, ao mesmo tempo, sejam de fácil manipulação. O cálculo da *posteriori* condicional nem sempre tem uma forma analítica, ou mesmo das *posteriori* marginais. Quando duas distribuições que, advindas da mesma família, possuem uma forma fechada para as *posteriori* condicionais e marginais, dizemos que são conjugadas. Uma das abordagens mais recorrentes para BVAR é um conjunto de *prioris* conjugadas conhecidas como Normal-Wishart.

##### 4.4.1 Normal-Wishart

Consideremos o modelo VAR em sua forma reduzida já apresentada:

$$\mathbf{y}_t = \sum_{m=1}^p \phi \mathbf{y}_{t-m} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}) \quad (4.24)$$

Há várias formas possíveis de se escrever um VAR, uma delas ocorre ao escrever o vetor  $\mathbf{y}_t$  de forma a empilhar os  $T$  valores assumidos pela primeira variável endógena  $y_1$ , em seguida, os  $T$  valores assumidos pela segunda variável endógena  $y_2$  e assim por diante, até a  $n$ -ésima variável, formando um vetor  $nT \times 1$ , a mesma os respectivos erros aleatórios  $\boldsymbol{\epsilon}_t$ . Uma segunda forma, alternativa, ocorre ao deixarmos em cada coluna uma variável endógena  $y_n$  e nas linhas os valores por elas assumidos em  $t = 1, \dots, T$ . Nesse trabalho vamos explorar a segunda forma.

Exemplificando a forma matricial proposta, de forma simples, e sem perda de generalidade, um VAR(p) sem constante e sem variáveis exógenas é escrito da forma:

$$\begin{bmatrix}
y_{(1)1} & y_{(2)1} & \cdots & y_{(n)1} \\
y_{(1)2} & y_{(2)2} & \cdots & y_{(n)2} \\
\vdots & \vdots & \ddots & \vdots \\
y_{(1)t} & y_{(2)t} & \cdots & y_{(n)t} \\
\vdots & \vdots & \ddots & \vdots \\
y_{(1)T} & y_{(2)T} & \cdots & y_{(n)T}
\end{bmatrix} =
\begin{bmatrix}
y_{(1)0} & \cdots & y_{(1)-p+1} & y_{(2)0} & \cdots & y_{(2)-p+1} & \cdots & y_{(n)0} & \cdots & y_{(n)-p+1} \\
y_{(1)1} & \cdots & y_{(1)-p+2} & y_{(2)1} & \cdots & y_{(2)-p+2} & \cdots & y_{(n)1} & \cdots & y_{(n)-p+2} \\
\vdots & \vdots \\
y_{(1)t-1} & \cdots & y_{(1)t-p+1} & y_{(2)t-1} & \cdots & y_{(2)t-p+1} & \cdots & y_{(n)t-1} & \cdots & y_{(n)t-p+1} \\
\vdots & \vdots \\
y_{(1)T-1} & \cdots & y_{(1)T-p+1} & y_{(2)T-1} & \cdots & y_{(2)T-p+1} & \cdots & y_{(n)T-1} & \cdots & y_{(n)T-p+1}
\end{bmatrix} \times$$

$$\begin{bmatrix}
\beta_{111} & \beta_{211} & \cdots & \beta_{n11} \\
\beta_{121} & \beta_{221} & \cdots & \beta_{n21} \\
\vdots & \vdots & \ddots & \vdots \\
\beta_{1n1} & \beta_{2n1} & \cdots & \beta_{nn1} \\
\beta_{112} & \beta_{212} & \cdots & \beta_{n12} \\
\beta_{122} & \beta_{222} & \cdots & \beta_{n22} \\
\vdots & \vdots & \ddots & \vdots \\
\beta_{1n2} & \beta_{2n2} & \cdots & \beta_{nn2} \\
\vdots & \vdots & \vdots & \vdots \\
\beta_{11p} & \beta_{21p} & \cdots & \beta_{n1p} \\
\beta_{12p} & \beta_{22p} & \cdots & \beta_{n2p} \\
\vdots & \vdots & \ddots & \vdots \\
\beta_{1np} & \beta_{2np} & \cdots & \beta_{nnp}
\end{bmatrix} +
\begin{bmatrix}
\epsilon_{(1)1} & \epsilon_{(2)1} & \cdots & \epsilon_{(n)1} \\
\epsilon_{(1)2} & \epsilon_{(2)2} & \cdots & \epsilon_{(n)2} \\
\vdots & \vdots & \ddots & \vdots \\
\epsilon_{(1)t} & \epsilon_{(2)t} & \cdots & \epsilon_{(n)t} \\
\vdots & \vdots & \vdots & \vdots \\
\epsilon_{(1)T} & \epsilon_{(2)T} & \cdots & \epsilon_{(n)T}
\end{bmatrix}$$

$$\begin{bmatrix}
\mathbf{Y}_1 \\
\mathbf{Y}_2 \\
\vdots \\
\mathbf{Y}_T
\end{bmatrix} =
\begin{bmatrix}
\mathbf{X}_1 \\
\mathbf{X}_2 \\
\vdots \\
\mathbf{X}_T
\end{bmatrix}
\begin{bmatrix}
\mathbf{B}'_1 \\
\mathbf{B}'_2 \\
\vdots \\
\mathbf{B}'_P
\end{bmatrix} +
\begin{bmatrix}
\mathbf{E}_1 \\
\mathbf{E}_2 \\
\vdots \\
\mathbf{E}_T
\end{bmatrix}$$

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}$$

Observada a amostra para  $t = 1, \dots, T$ , como  $\mathbf{Y}$  tem uma do tipo  $T \times n$ , a distribuição conjunta das observações pode ser diretamente escrita na forma da distribuição

normal métrica:

$$P(\mathbf{Y}|\mathbf{B}, \Sigma) = (2\pi)^{-\frac{mT}{2}} |\Sigma|^{\frac{T}{2}} \exp \left\{ -\frac{1}{2} \text{tr}[\Sigma^{-1}(\mathbf{Y} - \mathbf{X}\mathbf{B})'(\mathbf{Y} - \mathbf{X}\mathbf{B})] \right\} \quad (4.25)$$

Lembrando que o estimador de mínimos quadrados para  $\mathbf{B}$  é  $\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ , pode-se adicionar e subtrair o  $\mathbf{Y}$  predito,  $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}}$ , em 4.25. Colocando os parâmetros  $\boldsymbol{\theta}$  em função da amostra, temos que a função de verossimilhança fica na forma:

$$\begin{aligned} L(\mathbf{B}, \Sigma|\mathbf{Y}) &\propto |\Sigma|^{\frac{T}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \left( (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) + (\mathbf{X}\hat{\mathbf{B}} - \mathbf{X}\mathbf{B}) \right)' \left( (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) + (\mathbf{X}\hat{\mathbf{B}} - \mathbf{X}\mathbf{B}) \right) \right] \right\} \\ &= |\Sigma|^{\frac{T}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \left( (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) + (\hat{\mathbf{B}} - \mathbf{B})' \mathbf{X}' \mathbf{X} (\hat{\mathbf{B}} - \mathbf{B}) \right. \right. \right. \\ &\quad \left. \left. \left. + 2(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})' \mathbf{X} (\hat{\mathbf{B}} - \mathbf{B}) \right) \right] \right\} \\ &= |\Sigma|^{\frac{T}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \left( (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) + (\hat{\mathbf{B}} - \mathbf{B})' \mathbf{X}' \mathbf{X} (\hat{\mathbf{B}} - \mathbf{B}) \right) \right] \right\} \\ &= |\Sigma|^{\frac{T}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \left( (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) \right) \right] \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \left( (\hat{\mathbf{B}} - \mathbf{B})' \mathbf{X}' \mathbf{X} (\hat{\mathbf{B}} - \mathbf{B}) \right) \right] \right\} \\ &= |\Sigma|^{\frac{T}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \left( \hat{\mathbf{E}}' \hat{\mathbf{E}} \right) \right] \right\} \times \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \left( (\hat{\mathbf{B}} - \mathbf{B})' \mathbf{X}' \mathbf{X} (\hat{\mathbf{B}} - \mathbf{B}) \right) \right] \right\} \end{aligned}$$

Através da equação acima<sup>2</sup> fica claro que  $\mathbf{B}|\Sigma \sim MN_{np \times n}(\hat{\mathbf{B}}, \Sigma, (\mathbf{Z}'\mathbf{Z})^{-1})$ . A abordagem Bayesiana permite que o conjunto de parâmetros  $\boldsymbol{\theta} = \mathbf{B}, \Sigma$  tenha uma distribuição antes de observarmos a amostra. Há três possibilidades para o presente caso:

- a) caso 1 - ocorre quando o valor da matriz de covariâncias  $\Sigma$  é conhecido e o valor de  $\mathbf{B}$  não. Para tal, atribui-se uma distribuição de probabilidade *priori* apenas para este último;
- b) caso 2 - diferente do Caso 1, este ocorre quando o valor da matriz de covariâncias  $\Sigma$  não é conhecido porém o é para  $\mathbf{B}$ . Quando assim, é atribuída uma distribuição de probabilidade *priori* apenas para o primeiro;
- c) caso 3 - a última possibilidade ocorre quando ambos os parâmetros  $\mathbf{B}$  e  $\Sigma$  são desconhecidos. Dentre as possibilidades, é a preferida quanto à flexibilidade do modelo, uma vez que é permitido que todos os parâmetros variem em conjunto.

De acordo com o escopo deste trabalho, os casos 1 e 3 serão abordados. Para tal, um conjunto de *priori* muito popular em BVAR é a Normal-Wishart. A partir de então, distingue-se os valores dos parâmetros da *priori* pelo travessão embaixo  $\underline{\boldsymbol{\theta}}$ , e os da *posteriori* pelo travessão em cima,  $\bar{\boldsymbol{\theta}}$ .

Visto que o valor da *posteriori* de  $\mathbf{B}$ , é o mesmo para ambos os casos, são demonstrados os resultados apenas para este terceiro.

<sup>2</sup> Lembrar de acrescentar como eu fiz as manipulações algébricas.

#### 4.4.1.1 Distribuição *posteriori* conjunta

Neste este caso, ambos os parâmetros são desconhecidos, logo devem-se ser atribuídas a eles distribuições de probabilidade. Para os parâmetros da *priori* é utilizada a notação  $\underline{\theta}$ , ao passo que para os da *posteriori* é utilizado  $\underline{\theta}$ .

Assim, sob a hipótese de normalidade do termo do erro  $\mathbf{E}$ , sabe-se que, da inferência clássica, o  $\mathbf{B}$  tem distribuição  $N(\mathbf{B}_{MQO}, \Sigma \otimes (\mathbf{X}\mathbf{X}')^{-1})$ , ou alternativamente  $\mathbf{B}|\Sigma \sim MN_{\Sigma, \mathbf{X}\mathbf{X}'}(\mathbf{B}_\mu, \Sigma, \mathbf{X}\mathbf{X}')$ . É natural então que a distribuição *priori* de  $\mathbf{B}$  tenha essa forma:

$$\mathbf{B}|\Sigma \sim MN_{\Sigma, \mathbf{X}\mathbf{X}'}(\underline{\mathbf{B}}, \Sigma, \underline{\Omega})$$

A f.d.p. *priori* de  $\mathbf{B}$  então é dada por:

$$f(\mathbf{B}) = \frac{1}{(2\pi)^{\frac{NN}{2}} |\underline{\Omega}|^{\frac{N}{2}} |\Sigma|^{\frac{N}{2}}} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} (\mathbf{B} - \underline{\mathbf{B}})' \underline{\Omega} (\mathbf{B} - \underline{\mathbf{B}}) \right] \right\} \quad (4.26)$$

$$\propto |\Sigma|^{np+n+1} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} (\mathbf{B} - \underline{\mathbf{B}})' \underline{\Omega} (\mathbf{B} - \underline{\mathbf{B}}) \right] \right\} \quad (4.27)$$

Ainda é necessário atribuir uma distribuição de probabilidade para a matriz de covariâncias dos erros do modelo  $\Sigma$ . Uma opção direta é a Wishart inversa

$$f(\Sigma) = \frac{|\underline{\mathbf{S}}|^{-\frac{\nu}{2}}}{2^{\frac{n\nu}{2}} \Gamma_n(\frac{\nu}{2})} \exp \left\{ -\frac{1}{2} \text{tr} \Sigma^{-1} \underline{\mathbf{S}} \right\} |\Sigma|^{-\frac{1}{2}(\nu+n+1)} \quad (4.28)$$

$$\Sigma \sim iW(\underline{\mathbf{S}}, \nu) \quad (4.29)$$

Para estimar a distribuição *posteriori*, é necessário multiplicar a função de verossimilhança derivada anteriormente pela *priori* condicional de  $\mathbf{B}$  e pela marginal de  $\mathbf{S}$ :

$$\begin{aligned} p(\mathbf{B}, \Sigma | \tilde{\mathbf{Y}}_T) &\propto L(\mathbf{B}, \Sigma | \mathbf{Y}) p(\mathbf{B}, \Sigma) \\ &= L(\mathbf{B}, \Sigma | \mathbf{Y}) p(\mathbf{B}|\Sigma) p(\Sigma) \end{aligned} \quad (4.30)$$

De modo que:

$$\begin{aligned}
f(\mathbf{B}, \Sigma | \tilde{\mathbf{Y}}_T) &\propto \exp |\Sigma|^{-\frac{T}{2}} \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \left( (\mathbf{B} - \hat{\mathbf{B}})' \mathbf{X}' \mathbf{X} (\mathbf{B} - \hat{\mathbf{B}}) \right) \right] \right\} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \mathbf{S} \right] \right\} \\
&\times |\Sigma|^{-(np+n+1)} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} (\mathbf{B} - \underline{\mathbf{B}})' \underline{\Omega}^{-1} (\mathbf{B} - \underline{\mathbf{B}}) \right] \right\} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \underline{\mathbf{S}} \right] \right\} \\
&= |\Sigma|^{-(T+v+np+n+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} (\mathbf{B} - \overline{\mathbf{B}})' \overline{\Omega}^{-1} (\mathbf{B} - \overline{\mathbf{B}}) \right] \right\} \\
&\times \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \overline{\mathbf{S}} \right] \right\} \tag{4.31}
\end{aligned}$$

Abrindo o termo quadrado da última linha, a *posteriori* de  $\mathbf{B}$  deve apresentar a seguinte forma:

$$\begin{aligned}
f(\mathbf{B} | \Sigma, \tilde{\mathbf{Y}}_T) &\propto \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} (\mathbf{B} - \overline{\mathbf{B}})' \overline{\Omega} (\mathbf{B} - \overline{\mathbf{B}}) \right] \right\} \\
&= \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} (\mathbf{B}' - \overline{\mathbf{B}}') \overline{\Omega} (\mathbf{B} - \overline{\mathbf{B}}) \right] \right\} \\
&= \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} (\mathbf{B}' \overline{\Omega}^{-1} \mathbf{B} - \mathbf{B}' \overline{\Omega}^{-1} \overline{\mathbf{B}} - \overline{\mathbf{B}}' \overline{\Omega}^{-1} \mathbf{B} + \overline{\mathbf{B}}' \overline{\Omega}^{-1} \overline{\mathbf{B}}) \right] \right\}
\end{aligned}$$

O próximo passo então consiste em abrir os termos quadrados de  $f(\mathbf{B}, \Sigma | \tilde{\mathbf{Y}}_T)$  de forma que se assemelhe à expressão anterior. Dado que todos os termos são expoentes de mesma base, subtemos o termo, por ora, a fim de trabalhar com as propriedades de traço:

$$\begin{aligned}
&\text{tr} \left[ \Sigma^{-1} \left( (\mathbf{B} - \hat{\mathbf{B}})' \mathbf{X}' \mathbf{X} (\mathbf{B} - \hat{\mathbf{B}}) \right) \right] + \text{tr} \left[ \Sigma^{-1} (\mathbf{B} - \underline{\mathbf{B}})' \underline{\Omega}^{-1} (\mathbf{B} - \underline{\mathbf{B}}) \right] \\
&= \text{tr} \left[ \Sigma^{-1} \left( (\mathbf{B} - \hat{\mathbf{B}})' \mathbf{X}' \mathbf{X} (\mathbf{B} - \hat{\mathbf{B}}) + (\mathbf{B} - \underline{\mathbf{B}})' \underline{\Omega}^{-1} (\mathbf{B} - \underline{\mathbf{B}}) \right) \right]
\end{aligned}$$

Focando nos termos quadrados:

$$\begin{aligned}
&(\mathbf{B} - \hat{\mathbf{B}})' \mathbf{X}' \mathbf{X} (\mathbf{B} - \hat{\mathbf{B}}) + (\mathbf{B} - \underline{\mathbf{B}})' \underline{\Omega}^{-1} (\mathbf{B} - \underline{\mathbf{B}}) \\
&= \mathbf{B}' \mathbf{X}' \mathbf{X} \mathbf{B} - \mathbf{B} \mathbf{X}' \mathbf{X}' \hat{\mathbf{B}} - \hat{\mathbf{B}}' \mathbf{X}' \mathbf{X} \mathbf{B} + \hat{\mathbf{B}}' \mathbf{X}' \mathbf{X} \hat{\mathbf{B}} \\
&\quad + \mathbf{B}' \underline{\Omega}^{-1} \mathbf{B} - \mathbf{B} \underline{\Omega}^{-1} \underline{\mathbf{B}} - \underline{\mathbf{B}} \underline{\Omega}^{-1} \mathbf{B} + \underline{\mathbf{B}}' \underline{\Omega}^{-1} \underline{\mathbf{B}} \\
&= \mathbf{B}' (\mathbf{X}' \mathbf{X} + \underline{\Omega}^{-1}) \mathbf{B} - \mathbf{B}' (\mathbf{X}' \mathbf{X} \hat{\mathbf{B}} + \underline{\Omega}^{-1} \underline{\mathbf{B}}) - (\hat{\mathbf{B}}' \mathbf{X}' \mathbf{X} + \underline{\mathbf{B}}' \underline{\Omega}^{-1}) \mathbf{B} \\
&\quad + \hat{\mathbf{B}}' \mathbf{X}' \mathbf{X} \hat{\mathbf{B}} + \underline{\mathbf{B}}' \underline{\Omega}^{-1} \underline{\mathbf{B}} \\
&= \mathbf{B}' (\mathbf{X}' \mathbf{X} + \underline{\Omega}^{-1}) \mathbf{B} - \mathbf{B}' (\mathbf{X}' \mathbf{X} + \underline{\Omega}^{-1}) (\mathbf{X}' \mathbf{X} + \underline{\Omega}^{-1})^{-1} (\mathbf{X}' \mathbf{X} \hat{\mathbf{B}} + \underline{\Omega}^{-1} \underline{\mathbf{B}}) \\
&\quad - (\mathbf{X}' \mathbf{X} \hat{\mathbf{B}} + \underline{\Omega}^{-1} \underline{\mathbf{B}})' (\mathbf{X}' \mathbf{X} + \underline{\Omega}^{-1})^{-1} (\mathbf{X}' \mathbf{X} + \underline{\Omega}^{-1}) \mathbf{B} + \hat{\mathbf{B}}' \mathbf{X}' \mathbf{X} \hat{\mathbf{B}} + \underline{\mathbf{B}}' \underline{\Omega}^{-1} \underline{\mathbf{B}}
\end{aligned}$$

A próxima etapa compreende deixar os dois últimos termos da expressão acima de forma que possamos completar o quadrado:



$$\bar{\Omega}^{-1} = (\mathbf{X}'\mathbf{X} + \underline{\Omega}^{-1}) \quad (4.32)$$

$$\bar{\mathbf{B}} = (\mathbf{X}'\mathbf{X} + \underline{\Omega}^{-1})^{-1}(\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} + \underline{\Omega}^{-1}\underline{\mathbf{B}}) \quad (4.33)$$

$$= \bar{\Omega}(\mathbf{X}'\hat{\mathbf{Y}} + \underline{\Omega}^{-1}\underline{\mathbf{B}}) \quad (4.34)$$

$$\bar{\mathbf{S}} = \mathbf{S} + \underline{\mathbf{S}} + (\hat{\mathbf{B}} - \underline{\mathbf{B}})' \mathbf{X}' \mathbf{X} (\mathbf{X}' \mathbf{X} + \underline{\Omega}^{-1})^{-1} \underline{\Omega}^{-1} (\hat{\mathbf{B}} - \underline{\mathbf{B}}) \quad (4.35)$$

e então

$$\mathbf{B} | \Sigma, \tilde{\mathbf{Y}}_T \sim MN_{npn}(\bar{\mathbf{B}}, \Sigma, \bar{\Omega}) \quad (4.36)$$

$$\Sigma | \mathbf{Y}_T \sim iW(\bar{\mathbf{S}}, \bar{\nu}), \bar{\nu} = T + \nu \quad (4.37)$$

Uma vez que temos uma forma analítica da distribuição *posteriori* para o caso da conjugação normal-Wishart, a simulação desta pode ser feita pela da seguinte maneira:

- primeiramente deve-se gerar a matriz de covariâncias  $\Sigma^{(j)}$  para  $j = 1, \dots, R$  a partir da *posteriori*  $\Sigma | \mathbf{Y}_T \sim iW_n(\bar{\mathbf{S}}, T - np)$ ;
- tendo em mãos os  $R$  valores gerados para a matriz de covariâncias, deve-se gerar a matriz de parâmetros  $\mathbf{B}^{(j)}$  a partir da *posteriori* condicional  $\mathbf{B} | \Sigma^{(j)}, \mathbf{Y}_T \sim MN_{npn}(\bar{\mathbf{B}}, \Sigma^{(j)}, \bar{\Omega}^{-1})$ ;
- então gerar os valores  $\mathbf{e}_{T+1}^{(j)}, \dots, \mathbf{e}_{T+H}^{(j)}$  da distribuição  $\mathbf{e} \sim N(\mathbf{0}, \Sigma^{(j)})$ ;
- para gerar os valores preditos dos vetores  $\{\tilde{\mathbf{y}}_{T+h}^{(j)'}\}_{h=1}^H$  basta calcular recursivamente:

$$\begin{aligned} \tilde{\mathbf{y}}_{T+1}^{(j)'} &= \sum_{i=1}^p \mathbf{y}_{T+1-i}^{(j)'} \mathbf{B}_i^{(j)} + \mathbf{e}_{T+1}^{(j)'} \\ \tilde{\mathbf{y}}_{T+2}^{(j)'} &= \tilde{\mathbf{y}}_{T+1}^{(j)'} \mathbf{B}_1^{(j)} + \sum_{i=2}^p \mathbf{y}_{T+2-i}^{(j)'} \mathbf{B}_i^{(j)} + \mathbf{e}_{T+2}^{(j)'} \\ \tilde{\mathbf{y}}_{T+3}^{(j)'} &= \sum_{i=1}^2 \tilde{\mathbf{y}}_{T+3-i}^{(j)'} \mathbf{B}_i^{(j)} + \sum_{i=3}^p \mathbf{y}_{T+3-i}^{(j)'} \mathbf{B}_i^{(j)} + \mathbf{e}_{T+3}^{(j)'} \\ &\vdots \\ \tilde{\mathbf{y}}_{T+h}^{(j)'} &= \sum_{i=1}^{h-1} \tilde{\mathbf{y}}_{T+h-i}^{(j)'} \mathbf{B}_i^{(j)} + \sum_{i=h}^p \mathbf{y}_{T+h-i}^{(j)'} \mathbf{B}_i^{(j)} + \mathbf{e}_{T+h}^{(j)'} \end{aligned}$$

$$\text{Onde } \mathbf{B}^{(j)} = [\mathbf{B}_1^{(j)'} \quad \mathbf{B}_2^{(j)'} \quad \dots \quad \mathbf{B}_p^{(j)'}]'$$

Pegando apenas os vetores  $\{\tilde{\mathbf{y}}_{T+1}^{(j)}, \dots, \tilde{\mathbf{y}}_{T+H}^{(j)}\}_{j=1}^R$  tem-se então uma amostra de valores gerados independentes a partir da distribuição conjunta dos valores preditos.

A inferência pode então ser feita pelo intervalo de confiança  $1-\alpha$  para  $[\tilde{\mathbf{y}}_{T+1}, \dots, \tilde{\mathbf{y}}_{T+H}]$ , o qual é obtido pegando os valores  $(\tilde{\mathbf{y}}_{T+h}^{(l)}, \tilde{\mathbf{y}}_{T+h}^{(s)})$ , para todo  $h = 1, \dots, H$ , onde  $l = \lfloor M(\alpha/2) \rfloor$  e  $s = \lfloor M(\alpha - (1/2)) \rfloor$ , e os limitadores  $\lfloor \cdot \rfloor$  indicam a parte inteira do número gerado, e o valor  $\tilde{\mathbf{y}}_{T+h}^{(i)}$  indica a  $i$ -ésima estatística ordenada de  $\tilde{\mathbf{y}}_{T+h}$ .

#### 4.4.1.2 Distribuições marginais

Uma vez que, tendo em mãos a função *posteriori* condicional com os resultados acima, chegar nas marginais torna-se direto, integrando a condicional em relação aos parâmetros. Não há necessidade de calcular a marginal de  $\Sigma$ , uma vez que esta é utilizada diretamente no cálculo da condicional.

Para obter a distribuição marginal de  $\mathbf{B}$  dada a amostra  $\mathbf{Y}$ , é preciso integrar a distribuição *posteriori* conjunta dos parâmetros  $\mathbf{B}, \Sigma | \mathbf{Y} \sim MN_{npn}iW_n(\bar{\mathbf{B}}, \bar{\Omega}, \bar{\mathbf{S}}, \bar{\nu})$  em relação à  $\Sigma$ :

$$\begin{aligned} f(\mathbf{B}, \Sigma | \mathbf{Y}) &= \frac{|\bar{\mathbf{S}}|^{-\frac{\bar{\nu}}{2}} |\bar{\Omega}|^{-\frac{\bar{p}n}{2}}}{\Gamma_n(\frac{\nu}{2})(2\pi)^{\frac{npn}{2}} 2^{\frac{\nu np}{2}}} |\Sigma|^{-\frac{\bar{\nu}+np+n+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} (\mathbf{B} - \bar{\mathbf{B}})' \bar{\Omega} (\mathbf{B} - \bar{\mathbf{B}}) \right] \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \bar{\mathbf{S}} \right] \right\} \\ &= \frac{|\bar{\mathbf{S}}|^{-\frac{\bar{\nu}}{2}} |\bar{\Omega}|^{-\frac{\bar{p}n}{2}}}{\Gamma_n(\frac{\nu}{2})(2\pi)^{\frac{npn}{2}} 2^{\frac{\nu np}{2}}} |\Sigma|^{-\frac{\bar{\nu}+np+n+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \left( (\mathbf{B} - \bar{\mathbf{B}})' \bar{\Omega} (\mathbf{B} - \bar{\mathbf{B}}) + \bar{\mathbf{S}} \right) \right] \right\} \end{aligned}$$

Integrando em relação à  $\Sigma$ :

$$\begin{aligned} f(\mathbf{B} | \mathbf{Y}) &= \int_{\Sigma > 0} f(\mathbf{B}, \Sigma | \mathbf{Y}) d\Sigma \\ &= \frac{|\bar{\mathbf{S}}|^{-\frac{\bar{\nu}}{2}} |\bar{\Omega}|^{-\frac{\bar{p}n}{2}}}{\Gamma_n(\frac{\nu}{2})(2\pi)^{\frac{npn}{2}} 2^{\frac{\nu np}{2}}} \int_{\Sigma > 0} |\Sigma^{-1}|^{\frac{\bar{\nu}+np+n+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \left( (\mathbf{B} - \bar{\mathbf{B}})' \bar{\Omega} (\mathbf{B} - \bar{\mathbf{B}}) + \bar{\mathbf{S}} \right) \right] \right\} d\Sigma \end{aligned} \quad (4.38)$$

Seguindo Anderson et al. (1958), é tido que:

$$\int_{\mathbf{Z} > 0} |\mathbf{Z}|^{\alpha - \frac{n+1}{2}} \exp \{-\text{tr}(\mathbf{Z}\mathbf{\Lambda})\} = |\mathbf{\Lambda}|^{-\alpha} \Gamma_n(\alpha)$$

Então, para deixar a expressão 4.38 da forma conhecida, substitui-se  $\mathbf{\Lambda} = (1/2)((\mathbf{B} - \bar{\mathbf{B}})' \bar{\Omega} (\mathbf{B} - \bar{\mathbf{B}}) + \bar{\mathbf{S}})$ , ademais, é necessário ainda fazer a transformação  $\mathbf{Z} = \Sigma^{-1}$ , cujo jacobiano da transformação é  $|\mathbf{Z}|^{-(n+1)}$ .

Chamando  $k = |\bar{\mathbf{S}}|^{-\frac{\bar{\nu}}{2}} |\bar{\Omega}|^{-\frac{\bar{p}n}{2}} \left[ \Gamma_n(\frac{\nu}{2})(2\pi)^{\frac{npn}{2}} 2^{\frac{\nu np}{2}} \right]^{-1}$ :

$$\begin{aligned}
f(\mathbf{B}|\mathbf{Y}) &= k \int_{\mathbf{Z}>0} |\mathbf{Z}|^{\frac{\bar{\nu}+np}{2} + \frac{n+1}{2}} |\mathbf{Z}|^{-(n+1)} \exp \{-\text{tr}[\mathbf{Z}\mathbf{\Lambda}]\} d\mathbf{Z} \\
&= k \int_{\mathbf{Z}>0} |\mathbf{Z}|^{\frac{\bar{\nu}+np}{2} - \frac{n+1}{2}} \exp \{-\text{tr}[\mathbf{Z}\mathbf{\Lambda}]\} d\mathbf{Z} \\
&= k |\mathbf{\Lambda}|^{\frac{\nu+np}{2}} \Gamma_n \left( \frac{\bar{\nu} + np}{2} \right)
\end{aligned} \tag{4.39}$$

Então chegamos ao seguinte resultado:

$$f(\mathbf{B}|\mathbf{Y}) = \frac{\Gamma_n(\frac{\nu+np}{2}) |\bar{\mathbf{S}}|^{-\frac{\bar{\nu}}{2}} |\bar{\mathbf{\Omega}}|^{-\frac{n\bar{n}}{2}}}{\Gamma_n(\frac{\nu}{2}) (\pi)^{\frac{npn}{2}}} \left| \bar{\mathbf{S}} + (\mathbf{B} - \bar{\mathbf{B}})' \bar{\mathbf{\Omega}} (\mathbf{B} - \bar{\mathbf{B}}) \right|^{-\frac{\bar{\nu}+np}{2}} \tag{4.40}$$

$$= \frac{\Gamma_n(\frac{\nu+np}{2}) |\bar{\mathbf{S}}|^{-\frac{\bar{\nu}}{2}} |\bar{\mathbf{\Omega}}|^{-\frac{n\bar{n}}{2}}}{\Gamma_n(\frac{\nu}{2}) (\pi)^{\frac{npn}{2}}} \left| \bar{\mathbf{I}} + \mathbf{S}^{-1} (\mathbf{B} - \bar{\mathbf{B}})' \bar{\mathbf{\Omega}} (\mathbf{B} - \bar{\mathbf{B}}) \right|^{-\frac{\bar{\nu}+np}{2}} \tag{4.41}$$

As duas expressões acima são recorrentes para a distribuição *t* matricial da *posteriori* marginal de  $\mathbf{B}$ .

A inferência sobre os parâmetros  $\mathbf{B}$  pode ser feita diretamente ao gerar  $J$  matrizes de parâmetro  $\tilde{\mathbf{B}}$  e então tendo-os em mãos,  $\left\{ \tilde{\mathbf{B}}^{(j)} \right\}_{j=1}^J$ , a média é calculada como:

$$E(\hat{\mathbf{B}}) = \frac{1}{J} \sum_{j=1}^J \tilde{\mathbf{B}}^{(j)}$$

Onde os parâmetros são gerados da distribuição *posteriori* marginal  $\tilde{\mathbf{B}} \sim Mt_{npn}(\bar{\mathbf{B}}, \bar{\mathbf{\Omega}}, \bar{\mathbf{S}}, \bar{\nu})$ .

Havendo  $V(\mathbf{B}|\mathbf{Y})$ , é garantido assintoticamente que converge em distribuição:

$$\sqrt{M}(\text{vec}(\hat{\mathbf{B}}) - \text{vec}(\mathbf{B})) \rightarrow N(\mathbf{0}, V(\text{vec}(\mathbf{B})|\mathbf{Y}))$$

O intervalo de confiança  $1 - \alpha$  para  $\tilde{\mathbf{B}}$  pode ser obtido pelos valores  $(\tilde{\mathbf{B}}_{(l)}, \tilde{\mathbf{B}}_{(s)})$ , onde  $l = \lfloor M(\alpha/2) \rfloor$  e  $s = \lfloor M(\alpha - (1/2)) \rfloor$ , onde os limitadores  $\lfloor \cdot \rfloor$  indicam a parte inteira do número gerado, e o valor  $\tilde{\mathbf{B}}_{(i)}$  indica a  $i$ -ésima estatística ordenada de  $\tilde{\mathbf{B}}$ .

## 5 MODELAGEM DA VARIÂNCIA

Um dos fatos estilizados mais abordados das séries financeiras é o fenômeno conhecido como agrupamento da volatilidade. Em outras palavras, esse fenômeno diz que a variância dos retornos apresenta correlação serial no tempo, períodos de grande instabilidade nos retornos tendem a ser seguidos por períodos de grande instabilidade, ao passo que períodos mais estáveis tendem a ser seguidos por períodos de baixa volatilidade.

As pesquisas em séries temporais financeiras tem focado nas últimas décadas a modelar esse fenômeno. Os modelos de volatilidade podem ser divididos em dois grandes grupos, os da família de heterocedasticidade condicional auto-regressiva (ARCH, Autoregressive Conditional Heteroscedasticity) e os da família de Volatilidade Estocástica (SV, Stochastic Volatility).

O desenvolvimento de modelos econométricos que incorporam esse fenômeno tem mostrado êxito no meio acadêmico, tanto quanto aplicação prática destes no mercado. Apesar de sua popularidade, tem enfrentado dificuldades à incorporação de outros fatos estilizados, como, por exemplo, a assimetria da volatilidade (leverage effect). Logo, diversos trabalhos preocuparam-se em modificar os ARCH e SV a fim de ajustar-se melhor ao que é observado na prática. Nesse ponto, esbarrou-se na exigência de modelos cada vez mais flexíveis, capazes de explicar ao máximo os fenômenos e, conseguinte, aumentar seu poder preditivo.

Os modelos ARCH e SV surgiram e desenvolveram-se ao longo das décadas de 80 e 90 através de diversos autores. Os primeiros trabalhos contiveram-se a modelagem de séries univariadas, porém conforme a necessidade, logo generalizações surgiram desses modelos para o caso multivariado. Ainda hoje, pode-se dizer que o caso univariado atingiu um grau de sofisticação maior do que o multivariado. Assim que as extensões surgiram para esse segundo caso, esbarrou-se num problema já conhecido "maldição" da dimensionalidade, em outras palavras, o excessivo número de parâmetros a serem estimados.

Os esforços das pesquisas em modelos multivariados se voltaram a driblar esse problema, deixando a incorporação de outros fatos estilizados com um papel coadjuvante. O problema da dimensionalidade tem sido resolvido com excelência por modelos que reduzem o número de parâmetros sem perder sua capacidade explicativa. Para a família ARCH, esse processo avançou significativamente, ao passo que para a família SV, não. O desenvolvimento de trabalhos relativos a essa segunda família de modelos, mais flexível que a primeira no que tange à incorporação de um componente estocástico para a variância, encontrou então dificuldades quanto aos métodos clássicos de estimação. Diversos autores então propuseram métodos mais sofisticados para os SV, como o MCMC, enquanto que para os ARCH, estimadores populares se mostraram bem eficazes.

## 5.1 Modelos univariados

A seguir são discutidos sucintamente alguns modelos de volatilidade univariados e suas propriedades.

### 5.1.1 GARCH

Os modelos da família ARCH são também conhecidos como de volatilidade determinística. O trabalho seminal na área pode ser atribuído a Engle (1982) que propôs estimar parâmetros para a persistência da volatilidade no tempo. Por vezes esse modelo exige o uso de estimações demasiadas e então surge a popular generalização do modelo por Bollerslev (1986), o de Heterocedasticidade Condicionada Auto Regressiva Generalizada (GARCH, Generalized Auto Regressive Conditional Heterocedasticity). Esses trabalhos iniciais preocuparam-se, sobretudo, na modelagem do agrupamento de volatilidade, mas conseguinte, surgiram uma gama de trabalhos propondo incorporar os demais fatos estilizados. Dentre eles, o GARCH Exponencial (EGARCH, Exponential GARCH) de Nelson (1991), o GARCH Limiar (TARCH, Threshold GARCH) de Rabemananjara e Zakoian (1993) e Glosten, Jagannathan e Runkle (1993) e uma versão onde o processo gerador da volatilidade muda temporalmente, de Bauwens, Preminger e Rombouts (2006), (RS-GARCH, Regime Switching GARCH).

Apesar de serem bem populares, Nelson (1991) indica algumas desvantagens no uso destes, para garantir que a variância condicional seja positiva, deve-se impor restrições no espaço paramétrico que, por vezes, são violadas. Outra desvantagem desses modelos é que eles não levam em consideração que existe um componente aleatório no processo de variância condicional.

#### 5.1.1.1 O modelo

É dito que o erro  $e_t$  em  $t = 1, \dots, T$  segue um processo GARCH(p,q) se os dois primeiros momentos condicionais existem e, sendo  $\Psi_{t-1}$  o conjunto de informações disponíveis até  $t - 1$ , então:

- a)  $E(e_t | \Psi_{t-1}) = 0$
- b) Existem constantes  $\omega$ ,  $a_i$  para  $i = 1, \dots, q$  e  $b_j$  para  $j = 1, \dots, p$  tais que:

$$\text{Var}(e_t | \Psi_{t-1}) = \sigma_t^2 = \omega + \sum_{i=1}^q a_i e_{t-i}^2 + \sum_{j=1}^p b_j \sigma_{t-j}^2$$

Onde a variância de  $e_t$  é dada pelos termos de inovação defasados  $q$  vezes,  $e_{t-i}^2$  e pela própria variância defasada  $p$  vezes, dada por  $h_{t-j}$ . Se  $p = 0$ , então o modelo é reduzido à forma ARCH(q).

Considere um modelo GARCH(1,1). A condição de estacionariedade do processo é dada se:

$$-\infty \leq E[\log(a\eta_t^2 + b)] < 0$$

Então

$$h_t = \left\{ 1 + \sum_{i=1}^{\infty} a(\eta_{t-1}) \dots a(\eta_{t-i}) \right\} \omega$$

converge quase-certamente e  $e_t = \sqrt{h_t}\eta_t$  é solução única e estritamente estacionária do modelo. Se  $E[\log(a\eta_t^2 + b)] \geq 0$  e  $\omega > 0$ , então não há solução estritamente estacionária. A prova da condição de estacionariedade foge do escopo do trabalho. A condição de estacionariedade para o modelo genérico GARCH(p,q), tal como as respectivas provas podem ser encontradas em Francq e Zakoian (2011).

#### 5.1.1.2 Estimação

A estimação do GARCH geralmente é feita através da Quase-Máxima Verossimilhança, ou *Quasi-Maximum Likelihood Estimator* (QMLE). Seja novamente o modelo GARCH dado por

$$\begin{aligned} e_t &= \sqrt{h_t}\eta_t \\ h_t &= \omega + \sum_{i=1}^q a_i e_{t-i}^2 + \sum_{j=1}^p b_j h_{t-j} \end{aligned}$$

Onde  $\eta_t$  é uma sequência de variáveis aleatórias com variância 1. Sejam também  $\omega \geq 0$ ,  $a_i \geq 0$  para todo  $i = 1, \dots, q$  e  $b_j \geq 0$  para todo  $j = 1, \dots, p$ . Então de forma compacta, o vetor de parâmetros é escrito como:

$$\boldsymbol{\theta} = [\omega \ a_1 \ \dots \ a_q \ b_1 \ \dots \ b_p]'$$

O espaço paramétrico é definido como:

$$\Theta \subset (0, +\infty) \times [0, +\infty)^{p+q}$$

Para escrever a verossimilhança do modelo, então deve-se especificar uma distribuição para as variáveis aleatórias i.i.d.  $\eta_t$ . A função de quase-máxima verossimilhança coincide com a verossimilhança quando, condicionada a alguns valores iniciais,  $\eta_t$  é gaussiano. Os valores iniciais são dados por:

$$e_0^2 = \dots = e_{1-q}^2 = \sigma_0^2 = \dots = \sigma_{1-p}^2 = \frac{\omega}{1 - \sum_{i=1}^q a_i - \sum_{j=1}^p b_j}$$

E então a função de quase-verossimilhança é dada por:

$$L_n(\boldsymbol{\theta}) = L_n(\boldsymbol{\theta}; \tilde{e}_T) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{e_t^2}{2\sigma_t^2}\right)$$

Onde  $\sigma_t^2$  são recursivamente definidas em  $t \geq 1$  por:

$$\sigma_t^2 = \omega + \sum_{i=1}^q a_i e_{t-i}^2 + \sum_{j=1}^p b_j \sigma_{t-j}^2$$

A QMLE do vetor  $\boldsymbol{\theta}$  é dada pela maximização:

$$\hat{\boldsymbol{\theta}}_n = \arg \boldsymbol{\theta} \in \Theta L_n(\boldsymbol{\theta})$$

### 5.1.2 Volatilidade estocástica

Taylor (1986) introduz o SV como alternativa direta ao ARCH, com a vantagem de que, agora, a variância dos retornos passa a ser um processo randômico não observado. O SV original, de forma popular, modela a variância em logaritmo natural, o que aproxima bem o modelo de uma forma temporal contínua. O SV pressupõe que há um componente determinístico auto-regressivo, que capta a persistência da volatilidade no tempo, mais um componente aleatório. Após a introdução desse novo tipo de modelagem, tal como no caso ARCH, diferentes parametrizações surgiram, e.g. Jacquier, Polson e Rossi (2002), Yu, Yang e Zhang (2002) que propõe um modelo SV não-linear e Koopman e Uspensky (2002) formulam uma variação do SV na média.

O SV é um modelo mais flexível do que os de volatilidade determinística ARCH, através do componente aleatório, pode-se incorporar ao modelo os demais fatos estilizados como feito nos trabalhos de Nelson (1988), Barndorff-Nielsen (1997), Gallant, Hsieh e Tauchen (1997), Mahieu e Schotman (1998), Sandmann e Koopman (1998), Ruiz (1994), Harvey, Ruiz e Shephard (1994), Chib, Nardari e Shephard (2002), Jacquier, Polson e Rossi (2004) e Harvey e Shephard (1996). Ghysels et al. (1995) discute detalhadamente as propriedades do SV. As propriedades dos modelos SV podem ser encontradas nos trabalhos de Taylor (1994), Shephard (1996), Ghysels et al. (1995), Capobianco (1996) e Barndorff-Nielsen e Shephard (2001). Harvey e Shephard (1996) e So, Li e Lam (2002) ainda incorporam o *leverage effect* ao SV.

### 5.1.3 O modelo

O modelo SV se ajusta bem aos dados e à moderna teoria em finanças. Esse método tem maior flexibilidade que o GARCH em incorporar os fatos estilizados. Outra grande vantagem em seu uso ocorre da sua capacidade em ser colocado na forma estado-espço. Não há apenas um tipo de parametrização deste modelo, porém a mais popular é dada como:

$$e_t = \exp\left(\frac{h_t}{2}\right) \epsilon_t \quad (5.1)$$

$$h_t = \gamma + \phi h_{t-1} + \eta_t \quad (5.2)$$

Onde  $\epsilon_t \sim N(0, 1)$  e  $\eta_t \sim N(0, \sigma_\eta^2)$ . Como se pode perceber, a SV trabalha com o logaritmo das volatilidades, o que na prática não há necessidade de restringir que a variância seja sempre positiva. O parâmetro  $h_t$  então segue um processo aleatório e  $\phi$  é o parâmetro responsável pela persistência da variância no tempo.

Sob a condição de que  $|\phi| < 1$ , então  $h_t$  é um processo estritamente estacionário cuja média e variância são dadas como:

$$E(h_t) = \mu_h = \frac{\gamma}{1 - \phi} \quad \text{Var}(h_t) = \sigma_h^2 = \frac{\sigma_\eta^2}{1 - \phi^2} \quad (5.3)$$

Onde os erros  $\epsilon_t$  e  $\eta_t$  são gaussianos e independentes.

Para ver mais claramente como se dá a dinâmica deste modelo, considere elevar ao quadrado os termos e colocar na forma logarítmica:

$$\ln e_t^2 = h_t + \ln \epsilon_t^2$$

Se  $\epsilon_t \sim N(0, 1)$ , então  $\ln \epsilon_t^2$  tem distribuição  $\log \chi^2$  cuja média é  $\psi(1) - \ln 2 \simeq -1,2704$ , onde  $\psi(\cdot)$  é a função digama, e variância  $\pi^2/2 \simeq 4,9348$ . Definindo  $\zeta_t = \ln \epsilon_t^2 + 1,2704$ , então  $\zeta_t$  é i.i.d. com média 0 e variância  $\pi^2/2$ :

$$\ln e_t^2 = -1,2704 + h_t \zeta_t$$

### 5.1.4 Estimação

Nas próximas seções serão apresentados alguns métodos principais de estimação dos modelos SV.

#### 5.1.4.1 Métodos tradicionais

As limitações desse método encontram-se na estimação, não tão trivial quanto às do GARCH. A distribuição de  $\epsilon_t | \Psi_{t-1}$ , onde  $\Psi_{t-1}$  é o conjunto de informações até  $t - 1$ , não possui forma analítica, então a função de verossimilhança não é direta. Um dos métodos de se obter a verossimilhança é integrando a f.d.p. conjunta em relação à log-volatilidade, seja  $\boldsymbol{\theta} = [\gamma, \phi, \sigma_\eta^2]$  o conjunto de parâmetros, então:

$$\mathcal{L}(\mathbf{e}; \boldsymbol{\theta}) = \int p(\mathbf{e} | \mathbf{h}, \boldsymbol{\theta}) p(\mathbf{h} | \boldsymbol{\theta}) d\mathbf{h}$$

Esse método envolve métodos numéricos para a estimação do conjunto de parâmetros  $\boldsymbol{\theta}$ . Alternativamente, métodos do tipo integração via Monte Carlo são mais fáceis de implementar porém estas envolvem erros computacionais extras, não sendo tão atrativas. Dentre outros métodos, temos o *Generalized Method-of-Moments* proposto então por Taylor (1986), o qual não envolve a especificação da função de verossimilhança. Harvey, Ruiz e Shephard (1994) propõe a estimação via *Quasi-Maximum Likelihood*, (QML). A ideia básica por detrás deste método envolve considerar que a variável aleatória  $\zeta_t$  é *i.i.d.*  $N(0, \pi^2/2)$  e então o filtro de Kalman é aplicado às equações:

$$w_t = -1,2704 + h_t + \zeta_t \quad (5.4)$$

$$h_t = \gamma + \phi h_{t-1} + \eta_t \quad (5.5)$$

Onde  $\zeta_t = \ln \epsilon_t^2 - E(\ln \epsilon_t^2)$  e  $\text{Var}(\zeta_t) = \sigma_\zeta^2 = \pi^2/2$ . O modelo acima é claramente um modelo na forma estado-espço, porém fere as hipóteses de normalidade quanto à variável  $\zeta_t$ . A vantagem desta parametrização ocorre quanto à correlação contemporânea originalmente existente entre  $\epsilon_t$  e  $\eta_t$ . Como estão elevadas ao quadrado, essa correlação é perdida. O estimador de QML então trata os erros  $\zeta_t$  como se fossem distribuídos *i.i.d.*  $N(0, \pi^2/2)$ , o que é uma aproximação razoável.

#### 5.1.4.2 A abordagem Bayesiana

A abordagem Bayesiana é uma alternativa atrativa à estimação dos parâmetros da SV. Como anteriormente, essa abordagem, diferente da clássica, trata os verdadeiros parâmetros populacionais  $\boldsymbol{\theta}$  como aleatórios, sob uma distribuição *priori*, antes de observar a amostra  $\mathbf{Y}$ , é possível construir a distribuição *posteriori* como:

$$p(\boldsymbol{\theta} | \mathbf{Y}) = \frac{p(\mathbf{Y} | \boldsymbol{\theta}) p(\boldsymbol{\theta})}{\int_{\Theta} p(\mathbf{Y} | \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}} \propto p(\mathbf{Y} | \boldsymbol{\theta}) p(\boldsymbol{\theta})$$

Então, usando da distribuição *posteriori* de  $\boldsymbol{\theta}$ , é possível fazer a inferência estatística nos parâmetros, sob a suposição de que os dados foram gerados pelo modelo em escolhido.

A vantagem dessa abordagem é a não necessidade de que a *posteriori* tenha uma forma analítica, sendo que esta pode ser calculada via métodos MCMC, o que torna o método atrativo para o modelo. Há duas vertentes principais para essa abordagem, a primeira delas é a estimação em movimentos únicos, também conhecida como *Single-Move* MCMC, e a outra, mais explorada neste trabalho por suas propriedades é a estimação em movimentos múltiplos, também conhecida como *Multi-Move* MCMC.

#### 5.1.4.3 *Single-Move* MCMC

Esse método foi inicialmente proposto em Shephard (1993) e Jacquier, Polson e Rossi (2002). Essa estimação segue uma estrutura hierárquica de distribuições condicionais. A primeira delas é  $p(\mathbf{Y}|\mathbf{h})$ , a segunda coloca a volatilidade condicionada aos parâmetros da SV  $p(\mathbf{h}|\boldsymbol{\theta})$  e a terceira distribuição é a marginal dos parâmetros da SV ( $\boldsymbol{\theta}$ ). Desta forma:

$$p(\mathbf{h}, \boldsymbol{\theta}|\mathbf{Y}) \propto p(\mathbf{Y}|\mathbf{h})p(\mathbf{h}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

De outra forma, é dito que a distribuição *posteriori* de  $\boldsymbol{\theta}$  é dada pela expressão acima. Como é sabido, a *posteriori* não possui forma analítica, e o método MCMC não pode ser feito diretamente, sendo necessário reamostragens via Gibbs:

#### 5.1.4.4 *Algoritmo para a estimação de SV com Single-Move*

Para os valores  $J = B + R$ :

- a) inicialmente deve-se gerar  $\mathbf{h}^{i+1}$  a partir da distribuição  $\mathbf{h}^{j+1} \sim p(\mathbf{h}|\mathbf{y}\boldsymbol{\theta}^{(j)})$ ;
- b) então gerar os valores dos parâmetros  $\boldsymbol{\theta}^{(j+1)}$  a partir da distribuição  $\boldsymbol{\theta}^{(j+1)} \sim p(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{h}^{(j+1)})$ ;
- c) e então voltar para o primeiro item e repetir o algoritmo  $J$  vezes e excluir os  $B$  valores iniciais obtidos por questão de convergência.

Uma vez que o segundo passo do algoritmo é direto, a desvantagem deste método encontra-se no primeiro passo, o qual não pode ser feito diretamente. O *Single-Move* decompõe a f.d.p.  $p(\mathbf{h}|\mathbf{y}\boldsymbol{\theta}^{(j)})$  nas condicionais  $p(h_t|\mathbf{h}_{-t}^{(j)}, \mathbf{Y}, \boldsymbol{\theta}^{(j)})$ , onde  $\mathbf{h}_{-t}$  é o conjunto de todas as log-volatilidades em  $T$ , exceto em  $t$ . Usando das propriedades Markovianas, desta forma pode-se reduzir a condicional a:

$$p(h_t|\mathbf{h}_{-t}^{(j)}, \mathbf{Y}, \boldsymbol{\theta}^{(j)}) = p(h_t|h_{t-1}, h_{t+1}, y_t, \boldsymbol{\theta}) \propto p(y_t|h_t)p(h_{t+1}|h_t, \boldsymbol{\theta})p(h_t|h_{t-1}, \boldsymbol{\theta}) \quad (5.6)$$

### 5.1.4.5 Multi-Move MCMC

Esse método foi inicialmente sugerido em Kim, Shephard e Chib (1998). O procedimento é feito com base na forma estado-espço da SV como

$$\begin{aligned}\ln e_t^2 &= w_t = h_t + z_t \\ h_t &= \mu_h + \phi(h_{t-1} - \mu_h) + \eta_t\end{aligned}$$

A grande vantagem desta abordagem diz quanto à aproximação  $z_t = \ln e_t^2 \sim N(-1, 2704, \pi^2/2)$  para o caso *Single-Move*, a qual os autores consideram sub-ótima, sobretudo em pequenas amostras. Então, Kim, Shephard e Chib (1998) propõe uma aproximação da distribuição  $\ln \chi_1^2$  com base na mistura de 7 distribuições normais, de modo que:

$$p_z(z_t) \simeq \sum_{j=1}^7 q_j p_{z|\delta=j}(z_t | \delta_t = j)$$

Onde  $z_t | s_t = j \sim N(m_j - 1, 2704, v_j^2)$ , onde  $\delta$  é uma variável discreta que assume os valores no intervalo  $[1, 7]$  no conjunto  $\mathbb{N}$ , para  $t = 1, \dots, T$ . A expressão acima pode ser considerada então uma média ponderada de 7 distribuições normais, dado o fato que  $p(\delta_t = j) = q_j$  e  $\sum_{j=1}^7 q_j = 1$ .

Através desta abordagem, então, a aproximação das distribuições  $\ln \chi^2$  e gaussianas pode ser feita com um nível de acurácia maior que o *Single-Move*, e o filtro de Kalman pode ser aplicado de forma direta. Essa forma permite que todo o vetor  $\tilde{h}_T = \mathbf{h}_T$  seja estimado através de  $p(\mathbf{h} | \tilde{\mathbf{w}}, \tilde{\boldsymbol{\delta}})$  de uma só vez. O algoritmo de Gibbs assume então a forma genérica:

### 5.1.4.6 Algoritmo de Gibbs para estimação da SV com Multi-Move

Inicialmente deve-se escolher os valores iniciais  $\boldsymbol{\theta}^{(0)}$  e  $\mathbf{h}^{(0)}$ . Para os valores  $J = B + R$ :

- a) gerar os valores  $\boldsymbol{\delta}^{(j+1)} \sim p(\boldsymbol{\delta} | \mathbf{w}, \mathbf{h}^{(j)})$ ;
- b) então gerar os valores de  $\mathbf{h}_T$  com base na distribuição  $\mathbf{h}^{(j+1)} \sim p(\mathbf{h} | \mathbf{w}, \boldsymbol{\delta}^{(j+1)}, \boldsymbol{\theta}^{(j)})$ ;
- c) o próximo passo será gerar o vetor  $\boldsymbol{\theta}$  a partir da distribuição  $\boldsymbol{\theta}^{(j+1)} \sim p(\boldsymbol{\theta} | Z\mathbf{h}^{(j+1)})$ ;
- d) e então voltar para o primeiro item e repetir o algoritmo  $J$  vezes e excluir os  $B$  valores iniciais obtidos por questão de convergência.

Em Kim, Shephard e Chib (1998) encontramos que a correlação entre as recursivas gerações de parâmetros é reduzida drasticamente. Uma desvantagem deste método é que, apesar de ser mais eficiente do que o *Single-Move*, ele ainda é uma aproximação.

## 5.2 Modelos multivariados

Os modelos multivariados surgem com a proposta de modelar simultaneamente volatilidades de diversas séries financeiras distintas, a importância se dá na captura do último fato estilizado mencionado, quando a volatilidade tende a caminhar conjuntamente entre diversos retornos de ativos, de modo similar, podemos dizer que eles tem "algo" em comum, ou então que apresentam fatores comuns não observados. Para Platanioti, McCoy e Stephens (2005), os modelos multivariados podem ser melhor compreendidos que os univariados, na medida que eles contém mais informação: a relação entre as mudanças nos retornos dos ativos. Pitt e Shephard (1999) motivam o estudo de modelagem de variância para o caso multivariado quanto a sua interpretação econômica em Teoria de Precificação por Arbitragem (APT, Arbitrage Pricing Theory).

### 5.2.1 MGARCH

As versões multivariadas dos modelos GARCH (MGARCH) surgiram no começo do desenvolvimento desse tipo de abordagem, com as univariadas, dentre todas as especificações, são relevantes os trabalhos de Bollerslev, Engle e Wooldridge (1988) que faz uso do operador vech para modelar simultaneamente as volatilidades e uma generalização desse modelo conhecida como BEKK, de Engle e Kroner (1995). Esse tipo de modelagem frequentemente sofre com a quantidade alta de parâmetros a serem estimados, fenômeno também conhecido como problema de dimensionalidade. Os esforços no desenvolvimento destes deu-se então na redução do número de parâmetros, como exemplo, temos o modelo fatorial que, motivado pela teoria econômica como o APT de Ross (1976), o qual esforça-se em achar fatores comuns às diversas séries de retornos, como o factor-GARCH introduzido por Engle, Ng e Rothschild (1990) e o modelo ortogonal GARCH de Alexander (2001) e sua versão generalizada, de Weide (2002), que faz uso de autovalores e autovetores para redução do problema.

Ainda dentro das versões multivariadas da família GARCH, temos dois principais que trabalham com a correlação dos retornos, o modelo de Correlação Condicional Constante (CCC, Constant Conditional Correlation) de Bollerslev (1990) e o de Correlação Condicional Dinâmica (DCC, Dynamic Conditional Correlation) de Engle (2002). Jeantheau (1998) propôs uma extensão ao CCC que relaxa algumas das hipóteses do modelo, também conhecido como ECCC. De modo similar, há variações do DCC, como o proposto por Tse e Tsui (2002) e o DCC flexível quadrático de Billio, Caporin e Gobbo (2003) que impõe uma estrutura BEKK para as correlações (QFDCC, Quadratic Flexible Dynamic Conditional Correlation) e Silvennoinen e Teräsvirta (2009b) que lançam uma versão suavizada do modelo (STCC-GARCH, Smooth Transition Conditional Correlation GARCH). Cappiello, Engle e Sheppard (2006) generalizam o DCC a incorporar efeitos assimétricos no modelo (AG-DCC, Asymmetric Generalized DCC). Dentro outras

variações dessa ampla classe, encontramos trabalhos de Pelletier (2006) e Silvennoinen e Teräsvirta (2009a).

Nesta seção será introduzida as duas versões mais populares do MGARCH, o VECH e o BEKK-GARCH. Pode ser mostrado que, sob certas restrições, o BEKK-GARCH é um modelo mais genérico que aninha o VECH.

### 5.2.1.1 BEKK-GARCH

Dentre os modelos de volatilidade que buscam incorporar os *clusters* da volatilidade dos retornos, o mais famoso é o GARCH (Generalized Autoregressive Conditional Heteroskedasticity) de Bollerslev (1986). Desde então muitas variações deste modelo surgiram objetivando incorporar os demais fatos estilizados. O GARCH pode ser estendido ao caso multivariado (MGARCH), como o modelo BEKK de Engle e Kroner (1995). Esse modelo entende que há correlação temporal entre todas as covariâncias.

Seja o vetor de retornos dado por  $\mathbf{y}_t = [y_{(1)t} \ y_{(2)t} \ \dots \ y_{(n)t}]' \in \mathbb{R}^n$ . Aplicando um modelo estocástico, por exemplo o VAR, à série de retornos  $\mathbf{y}_t$  para  $t = 1, \dots, T$ , obtemos a média condicionada ao conjunto de informações passadas  $\Psi_{t-1}$ , dada pelo vetor  $\hat{\mathbf{y}}_t$ . A diferença entre o estimado e o observado na realidade é dada pelo termo de erro,  $\mathbf{e}_t$ . O modelo então assume que:

$$\mathbf{e}_t = \mathbf{H}_t^{-\frac{1}{2}} \mathbf{z}_t$$

Onde  $\mathbf{z}_t$  é uma sequência i.i.d. que obedece às seguintes características:

$$E(\mathbf{z}_t) = \mathbf{0}$$

$$E(\mathbf{z}_t \mathbf{z}_t') = \mathbf{I}_n$$

Em outras palavras, o vetor de aleatoriedade  $\mathbf{z}_t$  tem média  $\mathbf{0}$  e são não correlacionados, ademais, possuem variância unitária. Para o vetor de erros do modelo, é restrito que este obedeça as seguintes características:

$$E_{t-1}(\mathbf{e}_t) = E(\mathbf{e}_t | \Psi_{t-1}) = \mathbf{0} \tag{5.7}$$

$$E_{t-1}(\mathbf{e}_t \mathbf{e}_t') = \text{Cov}(\mathbf{e}_t | \Psi_{t-1}) = \mathbf{H}_t \tag{5.8}$$

$$E(\mathbf{e}_t \mathbf{e}_t') = \Sigma \tag{5.9}$$

O conjunto de informações disponíveis  $\Psi_{t-1}$  é geralmente tido como o conjunto de vetores  $\mathbf{y}$  disponíveis até  $t - 1$ . Através desse conjunto de informações a matriz  $\mathbf{H}_t$ , restrita como positiva definida, pode ser mensurada. De outra forma, os elementos da

diagonal de  $\mathbf{H}_t$  são as variâncias individuais de cada erro do modelo, e restringem-se à positividade estrita pela definição de variância. O modelo BEKK-GARCH, como proposto em Engle e Kroner (1995) segue a seguinte parametrização:

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}' + \sum_{k=1}^K \sum_{i=1}^q \mathbf{A}_{ik} \mathbf{e}_{t-i} \mathbf{e}'_{t-1} \mathbf{A}'_{ik} + \sum_{k=1}^K \sum_{j=1}^p \mathbf{B}_{ik} \mathbf{H}_{t-j} \mathbf{B}'_{ik}$$

O termo de intercepto, ou volatilidade incondicional, como dado, é uma decomposição onde  $\mathbf{C}$  é uma matriz triangular inferior, restrita de forma que os elementos da diagonal principal  $C_{k,l}$  para todo  $k = l$ , são positivos, garantindo que o modelo será corretamente identificado, ou seja, não há outras combinações que darão a mesma representação, então  $K = 1$ :

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}' + \sum_{i=1}^q \mathbf{A}_i \mathbf{e}_{t-i} \mathbf{e}'_{t-1} \mathbf{A}'_i + \sum_{j=1}^p \mathbf{B}_j \mathbf{H}_{t-j} \mathbf{B}'_j$$

A estacionariedade da matriz de covariâncias é dada se os autovalores de  $\sum_{i=1}^q \mathbf{A}_i \otimes \mathbf{A}_j + \sum_{j=1}^p \mathbf{B}_j \otimes \mathbf{B}_j$  estiverem dentro do círculo unitário.

O BEKK, tal como outros modelos multivariados da mesma natureza, sofrem do problema conhecido como maldição da dimensionalidade, onde o número de parâmetros  $R$  aumenta rapidamente à medida que são acrescentadas mais variáveis  $n$ ,  $R = (p+q)n^2 + \frac{n(n+1)}{2}$ . Suponha um modelo com  $n$  retornos, cuja volatilidade dos erros segue uma estrutura BEKK-GARCH( $p=1, q=1$ ) a forma matricial é melhor compreendida na forma aberta:

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}' + \mathbf{A}\mathbf{e}_{t-1} \mathbf{e}'_{t-1} \mathbf{A}' + \mathbf{B}\mathbf{H}_{t-1} \mathbf{B}' \quad (5.10)$$

$$(5.11)$$

$$\begin{aligned}
& \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ h_{21} & h_{22} & \dots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \dots & h_{nn} \end{bmatrix}_t \\
&= \mathbf{CC}' + \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{bmatrix} \begin{bmatrix} e_1^2 & e_1 e_2 & \dots & e_1 e_n \\ e_2 e_1 & e_2^2 & \dots & e_2 e_n \\ \vdots & \vdots & \ddots & \vdots \\ e_n e_1 & e_n e_2 & \dots & e_n^2 \end{bmatrix}_{t-1} \begin{bmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{bmatrix} \\
&+ \begin{bmatrix} B_{11} & B_{12} & \dots & B_{1n} \\ B_{21} & B_{22} & \dots & B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ B_{n1} & B_{n2} & \dots & B_{nn} \end{bmatrix} \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ h_{21} & h_{22} & \dots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \dots & h_{nn} \end{bmatrix}_{t-1} \begin{bmatrix} B_{11} & B_{21} & \dots & B_{n1} \\ B_{12} & B_{22} & \dots & B_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ B_{1n} & B_{2n} & \dots & B_{nn} \end{bmatrix} \\
&= \mathbf{CC}' \\
&+ \begin{bmatrix} \sum_{i=1}^n \sum_j^n A_{1j} A_{1i} e_{i,t-1} e_{j,t-1} & \sum_{i=1}^n \sum_j^n A_{1j} A_{2i} e_{i,t-1} e_{j,t-1} & \dots & \sum_{i=1}^n \sum_j^n A_{1j} A_{ni} e_{i,t-1} e_{j,t-1} \\ \sum_{i=1}^n \sum_j^n A_{2j} A_{1i} e_{i,t-1} e_{j,t-1} & \sum_{i=1}^n \sum_j^n A_{2j} A_{2i} e_{i,t-1} e_{j,t-1} & \dots & \sum_{i=1}^n \sum_j^n A_{2j} A_{ni} e_{i,t-1} e_{j,t-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n \sum_j^n A_{nj} A_{1i} e_{i,t-1} e_{j,t-1} & \sum_{i=1}^n \sum_j^n A_{nj} A_{2i} e_{i,t-1} e_{j,t-1} & \dots & \sum_{i=1}^n \sum_j^n A_{nj} A_{ni} e_{i,t-1} e_{j,t-1} \end{bmatrix} \\
&+ \begin{bmatrix} \sum_{i=1}^n \sum_j^n B_{1j} B_{1i} h_{ij,t-1} & \sum_{i=1}^n \sum_j^n B_{1j} B_{2i} h_{ij,t-1} & \dots & \sum_{i=1}^n \sum_j^n B_{1j} B_{ni} h_{ij,t-1} \\ \sum_{i=1}^n \sum_j^n B_{2j} B_{1i} h_{ij,t-1} & \sum_{i=1}^n \sum_j^n B_{2j} B_{2i} h_{ij,t-1} & \dots & \sum_{i=1}^n \sum_j^n B_{2j} B_{ni} h_{ij,t-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n \sum_j^n B_{nj} B_{1i} h_{ij,t-1} & \sum_{i=1}^n \sum_j^n B_{nj} B_{2i} h_{ij,t-1} & \dots & \sum_{i=1}^n \sum_j^n B_{nj} B_{ni} h_{ij,t-1} \end{bmatrix}
\end{aligned}$$

De outra forma, suponha então que separamos individualmente qualquer covariância entre duas séries da matriz de covariâncias  $\mathbf{H}_t$ , denotada por  $h_{IJ}$ , para  $I, J = 1, \dots, n$  então segue que esta é dada pela expressão:

$$\begin{aligned}
h_{IJ,t} &= K_{IJ} + A_{IJ}^2 e_{I,t-1} e_{J,t-1} + \sum_{i=1, i \neq I}^n \sum_{j=1, j \neq J}^n A_{Ij} A_{Ji} e_{i,t-1} e_{j,t-1} + B_{IJ}^2 h_{IJ,t-1} \\
&+ \sum_{i=1, i \neq I}^n \sum_{j=1, j \neq J}^n B_{Ij} B_{Ji} h_{ij,t-1}
\end{aligned}$$

Onde  $K_{IJ}$  é uma constante, gerada pelo termo  $\mathbf{CC}'$ . Essa parametrização é genérica o suficiente para captar quaisquer relações existentes entre as variâncias e covariâncias, por exemplo, a covariância dada por  $h_{IJ}$  em  $t$  é dada pelo seu próprio termo de inovação  $A_{IJ}^2 e_{I,t-1} e_{J,t-1}$ , mais todos os outros, mais a própria covariância defasada um período,  $B_{IJ}^2 h_{IJ,t-1}$  mais todas as outras covariâncias defasadas um período. Geralmente espera-se que os próprios termos defasados tenham um coeficiente mais significativo, o que leva o

BEKK-GARCH a sofrer restrições diretamente, devido a alta quantidade de parâmetros a ser estimada, típica de modelos multivariados com muitas variáveis endógenas. Levando em consideração essa restrição, as matrizes  $\mathbf{A}$  e  $\mathbf{B}$  passam a ser diagonais:

$$\begin{aligned}
& \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ h_{21} & h_{22} & \dots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \dots & h_{nn} \end{bmatrix}_t = \mathbf{C}\mathbf{C}' \\
& + \begin{bmatrix} A_{11} & 0 & \dots & 0 \\ 0 & A_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_{nn} \end{bmatrix} \begin{bmatrix} e_1^2 & e_1e_2 & \dots & e_1e_n \\ e_2e_1 & e_2^2 & \dots & e_2e_n \\ \vdots & \vdots & \ddots & \vdots \\ e_n e_1 & e_n e_2 & \dots & e_n^2 \end{bmatrix}_{t-1} \begin{bmatrix} A_{11} & 0 & \dots & 0 \\ 0 & A_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_{nn} \end{bmatrix} \\
& + \begin{bmatrix} B_{11} & 0 & \dots & 0 \\ 0 & B_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & B_{nn} \end{bmatrix} \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ h_{21} & h_{22} & \dots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \dots & h_{nn} \end{bmatrix}_{t-1} \begin{bmatrix} B_{11} & 0 & \dots & 0 \\ 0 & B_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & B_{nn} \end{bmatrix} \\
& = \mathbf{C}\mathbf{C}' + \begin{bmatrix} A_{11}^2 e_{1,t-1}^2 & A_{11}A_{22}e_{1,t-1}e_{2,t-1} & \dots & A_{11}A_{nn}e_{1,t-1}e_{n,t-1} \\ A_{11}A_{22}e_{1,t-1}e_{2,t-1} & A_{22}^2 e_{2,t-1}^2 & \dots & A_{22}A_{nn}e_{2,t-1}e_{n,t-1} \\ \vdots & \vdots & \ddots & \vdots \\ A_{nn}A_{11}e_{1,t-1}e_{n,t-1} & A_{nn}A_{22}e_{2,t-1}e_{n,t-1} & \dots & A_{nn}^2 e_{n,t-1}^2 \end{bmatrix} \\
& + \begin{bmatrix} B_{11}^2 h_{1,t-1} & B_{11}B_{22}h_{12,t-1} & \dots & B_{11}B_{nn}h_{1n,t-1} \\ B_{11}B_{22}h_{21,t-1} & B_{22}^2 h_{22,t-1} & \dots & B_{22}B_{nn}h_{2n,t-1} \\ \vdots & \vdots & \ddots & \vdots \\ B_{nn}B_{11}h_{n1,t-1} & B_{nn}B_{22}h_{n2,t-1} & \dots & B_{nn}^2 h_{n2,t-1} \end{bmatrix}
\end{aligned}$$

Outra forma de restringir o BEKK-GARCH é considerar que  $\mathbf{A}$  e  $\mathbf{B}$  não são diagonais, porém são escalares que multiplicam as matrizes  $\mathbf{e}_{t-1}\mathbf{e}'_{t-1}$  e  $\mathbf{H}_{t-1}$ . Desta forma então temos que:

$$\begin{aligned}
& \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ h_{21} & h_{22} & \dots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \dots & h_{nn} \end{bmatrix}_t = \mathbf{C}\mathbf{C}' + A \begin{bmatrix} e_1^2 & e_1e_2 & \dots & e_1e_n \\ e_2e_1 & e_2^2 & \dots & e_2e_n \\ \vdots & \vdots & \ddots & \vdots \\ e_n e_1 & e_n e_2 & \dots & e_n^2 \end{bmatrix}_{t-1} \\
& + B \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ h_{21} & h_{22} & \dots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \dots & h_{nn} \end{bmatrix}_{t-1}
\end{aligned}$$

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}' + A(\mathbf{e}_{t-1}\mathbf{e}_{t-1}') + B\mathbf{H}_{t-1}$$

Essa forma reduz drasticamente o número de parâmetros a ser estimado, sobretudo se considerarmos que  $\mathbf{C} = \mathbf{0}_{n \times n}$ . Essa forma é muito parecida com modelos do tipo *Exponential Weighted Moving Average*, ou EWMA, bastando restringir adicionalmente que a soma dos parâmetros seja  $A + B = 1$ , então:

$$\mathbf{H}_t = (1 - B)(\mathbf{e}_{t-1}\mathbf{e}_{t-1}') + B\mathbf{H}_{t-1}$$

Esse último modelo é bastante popular na literatura e amplamente aceito quando  $B = 0,94$ , para o caso quando são trabalhados dados diários, proposto por RiskMetrics.

#### 5.2.1.2 VECH-GARCH

O *vector-half* GARCH, ou VECH-GARCH, por vezes chamado apenas de VEGARCH, é o nome de uma parametrização do MGARCH considerada uma das mais diretas em relação ao GARCH univariado. Nesse modelo, todas as variâncias e covariâncias dependem de seus próprios valores defasados e das demais covariâncias defasadas, e dos seus próprios termos de inovação defasados, tal como dos demais igualmente defasados.

O nome faz referência ao operador matricial  $\text{vech}()$ . Tal como o tradicional operador  $\text{vec}()$ , responsável por vetorizar matrizes empilhando seus termos coluna a coluna, esse operador trabalha da mesma forma, porém diferentemente do outro, este só empilha os valores que estão no triângulo inferior da matriz, o que se ajusta bem ao caso, quando a matriz de covariâncias é simétrica e não há necessidade de se empilhar todos os valores, pois estes se repetem no triângulo superior.

Sob as mesmas suposições do BEKK-GARCH, considere que  $\boldsymbol{\eta}_t$  seja uma sequência i.i.d. do tipo ruído branco. Considere novamente  $\mathbf{e}_t$  como erros aleatórios em  $\mathbb{R}^n$ , então a decomposição pode ser feita:

$$\mathbf{e}_t = \mathbf{H}_t^{\frac{1}{2}}\boldsymbol{\eta}_t$$

Então a matriz  $\mathbf{H}_t$  é a matriz de covariâncias, e segue um processo determinístico dado pela seguinte parametrização:

$$\begin{aligned}
\begin{bmatrix} h_{11} \\ h_{21} \\ \vdots \\ h_{m1} \\ h_{22} \\ h_{32} \\ \vdots \\ h_{mm} \end{bmatrix} &= \begin{bmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{m1} \\ c_{22} \\ c_{32} \\ \vdots \\ c_{mm} \end{bmatrix} + \sum_{i=1}^p \begin{bmatrix} a_{11}^{(i)} & a_{12}^{(i)} & \dots & a_{1m}^{(i)} \\ a_{21}^{(i)} & a_{22}^{(i)} & \dots & a_{2m}^{(i)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}^{(i)} & a_{m2}^{(i)} & \dots & a_{mm}^{(i)} \end{bmatrix} \begin{bmatrix} e_1^2 \\ e_2 e_1 \\ \vdots \\ e_m e_1 \\ e_2^2 \\ e_3 e_2 \\ \vdots \\ e_m^2 \end{bmatrix}_{t-i} \\
&+ \sum_{j=1}^q \begin{bmatrix} b_{11}^{(j)} & b_{12}^{(j)} & \dots & b_{1m}^{(j)} \\ b_{21}^{(j)} & b_{22}^{(j)} & \dots & b_{2m}^{(j)} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1}^{(j)} & b_{m2}^{(j)} & \dots & b_{mm}^{(j)} \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{21} \\ \vdots \\ h_{m1} \\ h_{22} \\ h_{32} \\ \vdots \\ h_{mm} \end{bmatrix}_{t-j} \\
\text{vech}(\mathbf{H}_t) &= \mathbf{c} + \sum_{i=1}^p \mathbf{A}^{(i)} \text{vech}(\mathbf{e}_{t-i} \mathbf{e}'_{t-i}) + \sum_{j=1}^q \mathbf{B}^{(j)} \text{vech}(\mathbf{H}_{t-j}) \quad (5.12)
\end{aligned}$$

Onde as matrizes  $\mathbf{A}^{(i)}$  e  $\mathbf{B}^{(j)}$  são quadradas de dimensão  $m = n(n+1)/2$ . Deste modo, o VECH-GARCH se assemelha ao modelo BEKK-GARCH e a covariância entre dois erros, chamados  $I$  e  $J$ , para todo  $I \geq J$  em  $I, J = 1, \dots, n$  quaisquer é dada pegando-se cada equação individualmente:

$$h_{IJ,t} = c_{IJ} + \sum_{i=1}^p \sum_{k=1}^m a_{lk} e_{k,t-i} e_{k',t-i} + \sum_{j=1}^q \sum_{k=1}^m b_{lk} h_{k,t-j}$$

Onde  $l$  é a linha das matrizes  $\mathbf{A}^{(i)}$  e  $\mathbf{B}^{(j)}$  referentes à  $h_{IJ}$  no  $\text{vech}(\mathbf{H}_t)$  e  $k$  é o novo coeficiente obtido da vetorização da matriz de covariâncias e da matriz de inovações.

A desvantagem deste método encontra-se em restringir o modelo à estacionariedade. O VECH-GARCH não possui uma solução na forma explícita que garanta a estacionariedade da matriz de covariâncias. A abordagem deste tópico é extensiva e boas referências podem ser encontradas em Francq e Zakoian (2011).

Para fins didáticos, suponha agora uma representação do tipo VECH-GARCH ( $p=1, q=1$ ), então haverá um termo do primeiro somatório do lado direito da expressão que representará a defasagem do erro de inovação referentes aos ativos  $I$  e  $J$  e haverá no segundo somatório um termo referente à defasagem da própria covariância referente à  $I$  e

$J$ . Desses termos, tal como na parametrização BEKK-GARCH, são esperados valores de coeficientes mais altos que nos demais. Então da mesma forma, restringindo as matrizes  $\mathbf{A}^{(1)}$  e  $\mathbf{B}^{(1)}$  a serem diagonais, a dimensão do problema pode ser reduzida:

$$\begin{bmatrix} h_{11} \\ h_{21} \\ \vdots \\ h_{m1} \\ h_{22} \\ h_{32} \\ \vdots \\ h_{mm} \end{bmatrix} = \begin{bmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{m1} \\ c_{22} \\ c_{32} \\ \vdots \\ c_{mm} \end{bmatrix} + \begin{bmatrix} a_1 1^{(i)} & 0 & \dots & 0 \\ 0 & a_2 2^{(i)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{mm}^{(i)} \end{bmatrix} \begin{bmatrix} e_1^2 \\ e_2 e_1 \\ \vdots \\ e_m e_1 \\ e_2^2 \\ e_3 e_2 \\ \vdots \\ e_m^2 \end{bmatrix}_{t-1} \quad (5.13)$$

$$+ \begin{bmatrix} b_1 1^{(j)} & 0 & \dots & 0 \\ 0 & b_2 2^{(j)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & b_{mm}^{(j)} \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{21} \\ \vdots \\ h_{m1} \\ h_{22} \\ h_{32} \\ \vdots \\ h_{mm} \end{bmatrix}_{t-1} \quad (5.14)$$

$$\text{vech}(\mathbf{H}_t) = \mathbf{c} + \text{diag}(\mathbf{A}^{(1)})\text{vech}(\mathbf{e}_{t-1}\mathbf{e}'_{t-1}) + \text{diag}(\mathbf{B}^{(1)})\text{vech}(\mathbf{H}_{t-j}) \quad (5.15)$$

De modo que qualquer covariância entre  $I$  e  $J$  pode ser reduzida à expressão:

$$h_{IJ,t} = c_{IJ} + a_{kk}e_I e_J + b_{kk}h_{IJ,t-1}$$

Onde agora  $k$  representa o termo da diagonal das matrizes  $\mathbf{A}^{(1)}$  e  $\mathbf{B}^{(1)}$  referentes aos termo  $IJ$  na parametrização  $\text{vech}$ . Esse modelo é equivalente ao BEKK-GARCH diagonal e as restrições feitas à essas matrizes torna o modelo mais parcimonioso estatisticamente e, geralmente, pouca informação é perdida ao colocar os coeficientes fora da diagonal principal como 0.

### 5.2.2 MSV

A literatura para os modelos SV é ainda muito mais voltada ao caso univariado, de modo que modelos SV multivariados ainda são pouco percorridos, ao contrário do que ocorreu com os modelos MGARCH que, rapidamente ganharam popularidade nas pesquisas em finanças. Parte da dificuldade em se trabalhar com os modelos MSV fica por conta da sua própria estrutura logarítmica. O desenvolvimento da versão multivariada

pode ser atribuída principalmente a Harvey, Ruiz e Shephard (1994), Pitt e Shephard (1999) e Aguilar e West (2000) e por fim modelos baseados na distribuição de Wishart.

O primeiro caso citado refere-se ao Modelo de Correlação Condicional Constante (CCCM, Constant Conditional Correlation Model) que é uma versão SV do modelo Correlação Condicional de Bollerslev (1990) (CCC, Constant Conditional Correlation). É um modelo que pode ter seus parâmetros estimados via QML como proposto por Harvey, Ruiz e Shephard (1994), Danielsson (1994) estende esse modelo para incorporar *leverage effects*, Smith e Pitts (2006) usam um modelo bivariado similar e Ray e Tsay (2000) usam esse modelo para estudar componentes comuns de memória longa na volatilidade do mercado de capitais. Esse modelo também permite incorporar fatores comuns aos ativos, alguns trabalhos consideraram o caso para um fator, como em Shephard (1996), Liesenfeld e Richard (2003) e Jacquier, Polson e Rossi (1995), que propõe estimar os parâmetros via MCMC.

A segunda versão dos MSV equivale ao modelo fatorial de Engle, Ng e Rothschild (1990), cujo cerne deste é reduzir o problema da dimensão. Diferentes propostas que surgiram ao longo dos anos marcaram a evolução desse tipo de modelo, eles diferem essencialmente do modelo fatorial de Harvey, Ruiz e Shephard (1994) no que tange à incorporação dos fatores, como o desenvolvido inicialmente em Jacquier, Polson e Rossi (1995). Outra abordagem feita deve-se aos trabalhos de Pitt e Shephard (1999) e Aguilar e West (2000), os quais permitem que o termo de erro idiossincrático de cada retorno seja modelado por processos SV independentes. Uma desvantagem dos modelos fatoriais SV é que são passíveis de restrições no espaço paramétrico por conta da não identificabilidade correta dos parâmetros. Chib, Nardari e Shephard (2006) consegue incorporar distribuição marginal t-student aos erros do modelo de forma a captar as caldas mais pesadas, tanto como levar em consideração irregularidades na série de retornos, uma versão discretizada do processo de Lévy (Platanioti, McCoy e Stephens, 2005). Aguilar et al. (1998) e Jungbacker e Koopman (2006) utilizaram apenas um fator para decompor a variância.

Por ultimo, temos as versões MSV baseadas na distribuição de Wishart e sua conjugação com a Beta. O trabalho de Uhlig (1997) apresenta impedimentos à total aceitação em pesquisas de modelagem de séries financeiras, pois não é imposto restrições quanto a estacionariedade das séries. O modelo já apresenta uma versão mais geral desenvolvida por Philipov e Glickman (2006), que sugere estimação dos parâmetros via MCMC. Como alternativa, Gouriéroux, Jasiak e Sufana (2009) introduz um MSV baseado num processo Wishart auto-regressivo de ordem um (WAR(1)). Ademais, encontram-se trabalhos em MSV baseados na variação da correlação no tempo, dentre eles, Yu e Meyer (2006), Tsay (2005) e Jungbacker e Koopman (jungbacker2006monte) utilizaram essa metodologia.

## 6 MODELOS ESTADO-ESPAÇO

Apesar de terem sido concebidos inicialmente para a engenharia, esses modelos ganharam notoriedade nas pesquisas em econometria de séries temporais. Os modelos tipo estado-espço são sistemas que decompõem uma variável ou um vetor observáveis, em componentes não observáveis, também chamados de variáveis de estado, mais um termo de erro aleatório. Esses sistemas permitem, então, que as variáveis de estado tenha sua dinâmica dada de acordo com o comportamento apresentado pelos dados observáveis. A popularidade desses modelos se dá, sobretudo, pela sua flexibilidade em relação à estrutura do processo estocástico que os dados apresentam. Deste modo, diversos tipos de modelagem de dados variantes no tempo podem ser colocados na forma estado-espço, como ARMA, GARCH, tendência, sazonalidade e ciclos, e regressões cujos parâmetros variam temporalmente. Para mais referências sobre o assunto, consultar Kim e Nelson (1999) e Cuthbertson, Hall e Taylor (1992).

### 6.1 O filtro de Kalman

Quando trabalhamos com modelos econométricos temporais, buscamos identificar o processo gerador de uma determinada série  $y_t$  com base em alguns pressupostos. Esses pressupostos nos permitem estimar  $\hat{y}_t$  e predizê-lo  $h$  passos a frente. Para que métodos tradicionais de estimação sejam válidos, é necessário que o processo gerador permaneça constante no tempo, em outras palavras, os parâmetro do modelo são tidos como fixos temporalmente. Em muitas aplicações na econometria, é interessante relaxarmos essa hipótese. Considere o seguinte modelo:

$$y_t = \mathbf{X}'_t \boldsymbol{\beta}_t + \epsilon_t \quad (6.1)$$

A equação (6.1) representa uma regressão linear múltipla e um exemplo de modelo dinâmico em que os parâmetros  $\boldsymbol{\beta}$  já não são fixos, eles variam conforme novas informações chagam em  $t$ . Métodos convencionais de estimação já não funcionarão, então o uso do filtro de Kalman se mostra aplicável para esse tipo. O filtro de Kalman é um método feito em cima da forma estados-espço de modo recursivo, sempre atualizando a variável de estado de acordo com uma ponderação entre o que foi medido e o que foi predito. Para tornar mais claro como o filtro funciona, considere outro modelo similar, porém mais simples:

$$y_t = \pi_t + \epsilon_t \quad (6.2)$$

A equação (6.2) também é conhecida como modelo de componentes não observáveis, pois nele, a série observável  $y_t$  é decomposta num componente permanente  $\pi_t$  mais

um ruído branco, por vezes chamado de componente transitório  $\epsilon_t$ . O filtro de Kalman então provê um meio de estimar  $y_{t|t-1}$  baseando-se em informações até  $t - 1$  e então, atualizar o componente permanente conforme a chegada de novas informações. Essa atualização ocorre de acordo com o componente permanente predito mais a proporção  $k_t$  do erro de predição conforme a chegada de novas informações. Algebricamente podemos escrever como:

$$y_t = \pi_t + (1 - k_t)s_t \quad (6.3)$$

$$\pi_t = \pi_{t-1} + k_t s_t \quad (6.4)$$

Assume-se que  $s_t$  seja um ruído branco não correlacionado temporalmente com  $\pi_t$  e  $k_t$  seja um componente de atualização, também chamado de ganho de Kalman. Substituindo (6.3) em (6.4) temos a seguinte relação:

$$\begin{aligned} y_t &= \pi_{t-1} + s_t && \text{multiplicando por } k_t \text{ e isolando } s_t, \\ k_t s_t &= k_t(y_t + \pi_{t-1}) && \text{substituindo } k_t s_t \text{ em (6.4),} \\ \pi_t &= \pi_{t-1} + k_t(y_t - \pi_{t-1}) \end{aligned} \quad (6.5)$$

De acordo com (6.5) fica claro verificar que trata-se de um modelo adaptativo, onde o componente não observado permanente é ajustado de acordo com a chegada de  $y_t$  na proporção do ganho de kalman  $k_t$ .

De modo geral, como visto, os modelos estado-espço trabalham com dois tipos principais de equações, a equação de mensuração, e a de transição. Visando simplificar como filtro funciona, considere o seguinte modelo popular, com  $x$  e  $\beta$  vetores de entrada do modelo e parâmetros, ambos de dimensão  $n$  e  $x$ , e seja  $\epsilon$  um ruído branco.

$$\begin{aligned} y_t &= \mathbf{x}'_t \boldsymbol{\beta}_t + \epsilon_t && \text{equação mensuração} && (6.6) \\ \boldsymbol{\epsilon}_t &\sim N(0, \mathbf{V}_t) \\ E(\boldsymbol{\beta}_t \boldsymbol{\epsilon}_t) &= 0 \end{aligned}$$

$$\begin{aligned} \boldsymbol{\beta}_t &= \mathbf{T}_t \boldsymbol{\beta}_{t-1} + \mathbf{R}_t \boldsymbol{\eta}_t && \text{equação de transição} && (6.7) \\ \boldsymbol{\eta}_t &\sim N(0, \mathbf{Q}_t) \\ E(\boldsymbol{\beta}_{t-1} \boldsymbol{\eta}_t) &= 0 \end{aligned}$$

Ainda é necessário assumir que os erros são não correlacionados,  $E(\boldsymbol{\epsilon}_t \boldsymbol{\eta}_t) = 0$ . As equações (6.6) e (6.7) de mensuração e de transição juntas formam o modelo estado-espço, as matrizes  $\mathbf{T}_t$  e  $\mathbf{R}_t$  são conhecidas em  $t$ . Nesse sistema, escrito como estado-espço, os  $\beta_t$

são estocásticos e  $\mathbf{X}_t\beta_t$  é o componente permanente do modelo. Para que o filtro funcione, geralmente deve-se ter em mãos o valor inicial estimado  $\hat{\beta}_0$  do verdadeiro vetor  $\beta_0$  e sua matriz covariância estimada  $\hat{\mathbf{P}}_0$ , de acordo com a informação que o agente possui em  $t = 0$ , este vetor e matriz também são conhecidos como *priori* do modelo. Com a *priori*, é possível prever qual será o valor de  $\beta_t$  quando  $t = 1$ . Algebricamente, descrevemos a predição como:

$$\hat{\beta}_{1|0} = \mathbf{T}_0\hat{\beta}_0 \quad (6.8)$$

$$\text{Cov}(\hat{\beta}_{1|0}) = E(\hat{\beta}_{1|0} - \beta_1)(\hat{\beta}_{1|0} - \beta_1) = \mathbf{P}_{1|0} \quad (6.9)$$

Se considerarmos que a *priori* é dada pelo verdadeiro parâmetro inicial mais um termo de erro que representa a incerteza deste:

$$\hat{\beta}_0 = \beta_0 + \psi_0 \quad (6.10)$$

Podemos então isolar  $\beta_0$  em (6.10) e substituí-lo em (6.7) e fazer as devidas manipulações algébricas como segue:

$$\begin{aligned} \beta_0 &= \hat{\beta}_0 - \psi_0 && \text{de (6.7) temos:} \\ \beta_1 &= \mathbf{T}_1(\hat{\beta}_0 - \psi_0) + \mathbf{R}_1\eta_1 \\ \beta_1 &= \underbrace{\mathbf{T}_1\hat{\beta}_0}_{\hat{\beta}_{1|0}} - \mathbf{T}_1\psi_0 + \mathbf{R}_1\eta_1 \\ (\beta_1 - \hat{\beta}_{1|0}) &= -\mathbf{T}_1\psi_0 + \mathbf{R}_1\eta_1 \end{aligned} \quad (6.11)$$

Pela equação (6.11) fica claro que o erro de predição de  $\beta_{1|0}$  pode ser decomposto em dois componentes escritos à direita, o primeiro termo deve-se à incerteza em relação à *priori*, já o segundo termo corresponde à incerteza na equação de transição (6.7). Com esse resultado em mãos, fica fácil modificar (6.9) para obter  $\mathbf{P}_{1|0}$ :

$$\begin{aligned} \mathbf{P}_{1|0} &= E[(-\mathbf{T}_1\psi_0 + \mathbf{R}_1\eta_1)(-\mathbf{T}_1\psi_0 + \mathbf{R}_1\eta_1)'] \\ &= (\mathbf{T}_1\mathbf{P}_0\mathbf{T}_1' + \mathbf{R}_1\mathbf{Q}_1\mathbf{R}_1') \end{aligned} \quad (6.12)$$

Com a chegada de nova informação  $\mathbf{Y}_t$  em  $t = 1$ , o valor do parâmetro  $\beta$  deverá ser atualizado. O modelo então assume a seguinte forma algébrica, onde  $\omega_1$  é o erro na predição do parâmetro  $\beta_{1|0}$ :

$$\begin{aligned} \mathbf{Y}_1 &= \boldsymbol{\beta}_1 \mathbf{X} + \boldsymbol{\epsilon}_1 & \boldsymbol{\epsilon}_1 &\sim N(0, \mathbf{V}_1) \\ \boldsymbol{\beta}_{1|0} &= \boldsymbol{\beta}_1 + \boldsymbol{\omega}_1 & \boldsymbol{\omega}_1 &\sim N(0, \mathbf{P}_{1|0}) \end{aligned}$$

O sistema acima pode ser colocado na forma matricial:

$$\underbrace{\begin{pmatrix} \mathbf{Y}_1 \\ \boldsymbol{\beta}_{1|0} \end{pmatrix}}_{\tilde{\mathbf{Y}}_1} = \underbrace{\begin{pmatrix} \mathbf{X} \\ \mathbf{I} \end{pmatrix}}_{\tilde{\mathbf{X}}} \boldsymbol{\beta}_1 + \underbrace{\begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\omega}_1 \end{pmatrix}}_{\tilde{\boldsymbol{\epsilon}}_1} \quad (6.13)$$

$$\tilde{\mathbf{Y}}_1 = \tilde{\mathbf{X}} \boldsymbol{\beta}_1 + \tilde{\boldsymbol{\epsilon}}_1 \quad (6.14)$$

Se assumirmos que:

$$\text{Cov}(\tilde{\boldsymbol{\epsilon}}_1) = \begin{pmatrix} \mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{1|0} \end{pmatrix} = \tilde{\mathbf{V}}_1 \quad (6.15)$$

Feitas as transformações, pode-se então rodar o estimador de mínimos quadrados generalizados (MQG) afim de obter  $\boldsymbol{\beta}_1$ :

$$\boldsymbol{\beta}_1 = (\tilde{\mathbf{X}}' \tilde{\mathbf{V}}_1^{-1} \tilde{\mathbf{X}})^{-1} (\tilde{\mathbf{X}}' \tilde{\mathbf{V}}_1^{-1} \tilde{\mathbf{Y}}_1) \quad (6.16)$$

$$\text{Cov}(\boldsymbol{\beta}_1) = \begin{pmatrix} \mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{1|0} \end{pmatrix} \left[ (\mathbf{X}' \quad \mathbf{I}) \begin{pmatrix} \mathbf{X} \\ \mathbf{I} \end{pmatrix} \right]^{-1} = \mathbf{P}_1 \quad (6.17)$$

Ao abrirmos a equação (6.17) com base em (6.15) e (6.13), obtemos o resultado que segue:

$$\mathbf{P}_1^{-1} = \tilde{\mathbf{V}}_1^{-1} \tilde{\mathbf{X}}' \tilde{\mathbf{X}} \quad (6.18)$$

$$= \mathbf{X}' \mathbf{V}_1^{-1} \mathbf{X} \quad (6.19)$$

$$\left[ (\mathbf{X}' \quad \mathbf{I}) \begin{pmatrix} \mathbf{V}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{1|0}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{X}' \\ \mathbf{I} \end{pmatrix} \right] \quad (6.20)$$

$$= (\mathbf{X}' \mathbf{V}_1^{-1} \mathbf{X} + \mathbf{P}_{1|0}^{-1}) \quad (6.21)$$

Com o resultado (6.21) em mãos, podemos derivar a equação de atualização como segue:

$$\begin{aligned}\beta_1 &= \mathbf{P}_1 \left[ \begin{pmatrix} \mathbf{X}' & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{V}_1^{-1} & 0 \\ 0 & \mathbf{P}_{1|0}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{Y}_1 \\ \beta_{1|0} \end{pmatrix} \right] \\ &= (\mathbf{P}_1 \mathbf{X}' \mathbf{V}_1^{-1}) \mathbf{Y}_1 + \mathbf{P}_1 \mathbf{P}_{1|0}^{-1} \beta_{1|0}\end{aligned}\quad (6.22)$$

Substituindo  $\mathbf{P}_{1|0}^{-1}$  de (6.21) na equação (6.22) e realizando as seguintes manipulações, chegamos ao seguinte resultado:

$$\beta_1 = (\mathbf{P}_1 \mathbf{X}' \mathbf{V}_1^{-1}) \mathbf{Y}_1 + \mathbf{P}_1 (\mathbf{P}_1^{-1} - \mathbf{X}' \mathbf{V}_1^{-1} \mathbf{X}) \beta_{1|0} \quad (6.23)$$

$$= (\mathbf{P}_1 \mathbf{X}' \mathbf{V}_1^{-1}) \mathbf{Y}_1 + (\mathbf{I} - \mathbf{P}_1 \mathbf{X}' \mathbf{V}_1^{-1} \mathbf{X}) \beta_{1|0} \quad (6.24)$$

$$= \beta_{1|0} + \underbrace{(\mathbf{P}_1 \mathbf{X}' \mathbf{V}_1^{-1})}_{\mathbf{K}_1} (\mathbf{Y}_1 - \mathbf{X} \beta_{1|0})$$

$$= \beta_{1|0} + \mathbf{K}_1 (\mathbf{Y}_1 - \mathbf{X} \beta_{1|0}) \quad \text{equação de atualização} \quad (6.25)$$

De (6.22), através do segundo termo do lado direito, podemos reformular a matriz de covariância de  $\beta_1$  em função do ganho de Kalman  $\mathbf{K}_1$ :

$$\begin{aligned}\mathbf{P}_1 \mathbf{P}_{1|0}^{-1} &= \mathbf{P}_1 (\mathbf{P}_1^{-1} - \mathbf{X}' \mathbf{V}_1^{-1} \mathbf{X}) \\ &= (\mathbf{I} - \mathbf{K}_1 \mathbf{X})\end{aligned}$$

$$\mathbf{P}_1 = (\mathbf{I} - \mathbf{K}_1 \mathbf{X}) \mathbf{P}_{1|0} \quad (6.26)$$

$$(6.27)$$

Outra forma de escrevermos o ganho ótimo de Kalman segue diretamente da definição acima de  $\mathbf{P}_1$ , ao substituir a matriz na definição dada anteriormente em (6.25), segue que:

$$\begin{aligned}\mathbf{K}_1 &= \mathbf{P}_1 \mathbf{X}' \mathbf{V}_1^{-1} \\ &= (\mathbf{I} - \mathbf{K}_1 \mathbf{X}) \mathbf{P}_{1|0} \mathbf{X}' \mathbf{V}_1^{-1} \\ &= \mathbf{P}_{1|0} \mathbf{X}' \mathbf{V}_1^{-1} - \mathbf{K}_1 \mathbf{X} \mathbf{P}_{1|0} \mathbf{X}' \mathbf{V}_1^{-1}\end{aligned}$$

$$\mathbf{K}_1 (\mathbf{I} + \mathbf{X} \mathbf{P}_{1|0} \mathbf{X}' \mathbf{V}_1^{-1}) = \mathbf{P}_{1|0} \mathbf{X}' \mathbf{V}_1^{-1}$$

Multiplicando ambos os lados por  $\mathbf{V}_1$ :

$$\mathbf{K}_1 (\mathbf{V}_1 + \mathbf{X} \mathbf{P}_{1|0} \mathbf{X}') = \mathbf{P}_{1|0} \mathbf{X}'$$

$$\mathbf{K}_1 = \mathbf{P}_{1|0} \mathbf{X}' \underbrace{(\mathbf{V}_1 + \mathbf{X} \mathbf{P}_{1|0} \mathbf{X}')^{-1}}_F$$

$$= \mathbf{P}_{1|0} \mathbf{X}' \mathbf{F}_1^{-1} \quad (6.28)$$

### 6.1.1 Estimação dos parâmetros

Os modelos em estado-espço de modo geral trabalham com a modelagem de séries que envolvem variáveis não observáveis. O filtro de Kalman então cumpre um papel dentro desses modelos como um estimador dos componentes não observáveis. É um método recursivo de estimação onde os cálculos são refeitos a cada chegada de novas informações, funcionando de forma a atualizar o componente não observável.

Dentre as inúmeras possibilidades do uso do filtro em economia, atemo-nos ao presente caso quando é necessário que os parâmetros de um modelo de regressão linear, como o VAR, por exemplo, alteram-se temporalmente, geralmente por uma questão de aprendizagem e expectativas racionais. Tendo esse fenômeno em vista, considere o seguinte modelo em questão:

$$\mathbf{y}_t = \mathbf{x}_t \mathbf{B}_t + \mathbf{e}_t \quad \mathbf{e}_t \sim i.i.d.N(\mathbf{0}, \mathbf{R}) \quad (6.29)$$

$$\mathbf{B}_t = \tilde{\boldsymbol{\mu}} + \mathbf{T} \mathbf{B}_{t-1} + \mathbf{v}_t \quad \mathbf{v}_t \sim i.i.d.N(\mathbf{0}, \mathbf{Q}) \quad (6.30)$$

Onde  $\mathbf{y}_t$  é um vetor  $n$ -dimensional de variáveis endógenas,  $\mathbf{X}_t$  é um vetor  $k$ -dimensional, com  $k = np$  para o caso de um VAR(p), e os termos de erro  $\mathbf{e}_t$  e  $\mathbf{v}_t$  são independentes. Os parâmetros  $\tilde{\boldsymbol{\mu}}$  e  $\mathbf{T}$  são conhecidos como hiperparâmetros, por ora, consideramos como dados. O algoritmo do filtro de Kalman pode ser facilmente estendido para o caso quando esses parâmetros não são conhecidos, caso em que há necessidade de serem estimados.

O filtro então funciona recursivamente a cada chegada de nova informação, esse processo basicamente pode ser dividido em dois passos principais:

- a) **predição:** antes do tempo  $t$  é necessário prever otimamente o valor que  $\mathbf{y}_t$  assumirá. Essa predição é calculada com base em toda informação disponível até  $t - 1$ . De modo compacto o valor predito é escrito como  $\mathbf{y}_{t|t-1}$  e ele é feito com base na matriz de parâmetros predita para  $t$ , escrita compactamente como  $\mathbf{B}_{t|t-1}$ ;
- b) **atualização:** essa etapa ocorre da chegada de uma nova informação, chamamos de  $\mathbf{y}_t$ . O filtro então calcula o erro de predição como  $\boldsymbol{\eta}_{t|t-1} = \mathbf{y}_t - \mathbf{y}_{t|t-1}$ . Com o erro de predição, os parâmetros podem ser atualizados, obtendo  $\mathbf{B}_{t|t}$  e então voltar à etapa de predição, agora prevendo  $\mathbf{y}_{t+1}$ , com base na informação disponível até  $t$ , chamado de  $\mathbf{y}_{t+1|t}$ .

O algoritmo pode ser sumarizado da seguinte forma:

- a) inicialmente são estipulados valores para  $\mathbf{B}_{0|0}$  e  $\mathbf{P}_{0|0}$  com base em conhecimento prévio do pesquisador ou mesmo na estimação dos parâmetros com base numa pré-amostra;

- b) o segundo passo então consiste em calcular os valores dos parâmetros preditos, ela é feita através da equação de transição:

$$\mathbf{B}_{1|0} = \tilde{\boldsymbol{\mu}} + \mathbf{T}\mathbf{B}_{0|0} \quad \mathbf{P}_{1|0} = \mathbf{F}\mathbf{P}_{0|0}\mathbf{F}' + \mathbf{Q}$$

- c) com a chegada de uma nova informação em  $t = 1$ , definida pelo vetor  $\mathbf{y}_1$ , então o erro de predição e sua variância devem ser calculados da seguinte forma:

$$\begin{aligned} \boldsymbol{\eta}_{1|0} &= \mathbf{y}_1 - \mathbf{y}_{1|0} = \mathbf{y}_1 - \mathbf{x}_1\mathbf{B}_{1|0} \\ \mathbf{f}_{1|0} &= \mathbf{x}_1\mathbf{P}_{1|0}\mathbf{x}'_1 + \mathbf{R} \end{aligned}$$

- d) com os resultados do item anterior, o parâmetro  $\mathbf{B}_{1|0}$ , que é uma predição, deve ser corrigido com base no erro de predição  $\boldsymbol{\eta}_{1|0}$ :

$$\begin{aligned} \mathbf{B}_{1|1} &= \mathbf{B}_{1|0} + \mathbf{P}_{1|0}\mathbf{x}'_1\mathbf{f}_{1|0}^{-1}\boldsymbol{\eta}_{1|0} = \mathbf{B}_{1|0} + \mathbf{K}_1\boldsymbol{\eta}_{1|0} \\ \mathbf{P}_{1|1} &= \mathbf{P}_{1|0} - \mathbf{P}_{1|0}\mathbf{x}'_1\mathbf{f}_{1|0}^{-1}\mathbf{x}_1\mathbf{P}_{1|0} \end{aligned}$$

- e) com  $\mathbf{B}_{1|1}$  e  $\mathbf{P}_{1|1}$ , o item 1 deve ser retomado e assim ser feito recursivamente até toda a amostra em  $t = T$ .

O ganho de Kalman  $\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{x}'_t\mathbf{f}_{t|t-1}^{-1}$  é o responsável por determinar o peso dado à nova informação que chegou, dada então pelo erro de predição  $\boldsymbol{\eta}_{t|t-1}$ .

Nesse modelo, o erro de predição pode ser decomposto em duas partes. A primeira deve-se ao erro de inferência ao tentar prever o vetor  $\mathbf{y}_t$  com base na predição  $\mathbf{B}_{t|t-1}$ , ou seja, a diferença  $\mathbf{B}_t - \mathbf{B}_{t|t-1}$ . A segunda parte do erro de predição ocorre por conta da incerteza intrínseca ao modelo, dada pela aleatoriedade  $\mathbf{e}_t$ . Ambas as incertezas são expressas diretamente na equação da variância do erro de predição  $\mathbf{f}_{t|t-1}$ .

Outro ponto a ser notado é que o ganho de Kalman é uma função inversa da matriz de covariâncias  $\mathbf{R}$ , ou seja, quanto maior for a variância dos erros  $\mathbf{e}_t$ , menor peso se dará à chegada da nova informação. O resultado é intuitivo dado que grande parte da incerteza quanto à predição está associada à especificação do modelo dada pela equação de mensuração e não pela incerteza na predição de  $\mathbf{B}_t$ . De outra forma, conforme as incertezas associadas à previsão em  $\mathbf{B}_{t|t-1}$  aumentam, o filtro dará maior peso à chegada da nova informação, uma vez que a predição não tem sido boa.

## 6.2 Suavizador

Ao contrário do filtro de Kalman, por vezes é necessário estimar os estados condicionais por toda a amostra  $\tilde{\mathbf{y}}_T$ . Esse método é muito utilizado na abordagem Bayesiana

quanto ao cômputo da distribuição *posteriori* da marginal dos estados  $\mathbf{B}_t$ , para o presente caso BVAR. Equivalentemente, todos os estados são atualizados com base em toda a amostra em  $T$ :

$$p(\mathbf{B}_t|\tilde{\mathbf{y}}_T)$$

A diferença entre o filtro e o suavizador, mais popularmente conhecido na literatura do inglês *smoother*, basicamente refere-se aos dados considerados para a estimação dos estados  $\mathbf{B}_t$ , ao passo que o filtro se utiliza dos dados passados e do atual  $\tilde{\mathbf{Y}}_{t-1}, \mathbf{y}_t$ , o *smoother* se utiliza dos dados futuros. Depois de obter a *posteriori* dos estado através do filtro,  $\mathbf{B}_T$ , cujo processo é progressivo em  $t = 1, \dots, T$ , então o suavizador pode obter todos os estados condicionados a toda a amostra de forma retroativa, em  $t = T - 1, \dots, 1$ . Para outras referências sobre o tópico, consultar Särkkä (2013).

De modo compacto, as equações retroativas do suavizador Bayesiano são gerados pela seguinte f.d.p.:

$$\begin{aligned} p(\mathbf{B}_{t+1}|\tilde{\mathbf{Y}}_T) &= \int p(\mathbf{B}_{t+1}|\mathbf{B}_t)p(\mathbf{B}_t|\tilde{\mathbf{Y}}_T)d\mathbf{B}_t \\ p(\mathbf{B}_t|\tilde{\mathbf{Y}}_T) &= p(\mathbf{B}_t|\tilde{\mathbf{Y}}_T) \int \frac{p(\mathbf{B}_{t+1}|\mathbf{Y}_t)p(\mathbf{B}_{t+1}|\tilde{\mathbf{Y}}_T)}{p(\mathbf{B}_{T+1}|\tilde{\mathbf{Y}}_T)} \end{aligned}$$

Onde  $p(\mathbf{B}_t|\tilde{\mathbf{Y}}_T)$  é dado pelo filtro de Kalman no estado  $\mathbf{B}_t$ .

A derivação de  $\mathbf{B}_{t|t, \mathbf{B}_{t+1}} = E(\mathbf{B}_t|\mathbf{B}_{t|t}, \mathbf{B}_{t+1})$  equivale a atualizar a estimação de  $\mathbf{B}_t$  combinando  $\mathbf{B}_{t|t}$  com as informações disponíveis no vetor gerado  $\mathbf{B}_{t+1}$ . Desta forma,  $\mathbf{B}_{t+1}$  pode ser considerado um vetor adicional de dados observados que contém informação além de  $\mathbf{B}_{t|t}$ . Tome a seguinte equação de estados:

$$\mathbf{B}_{t+1} = \boldsymbol{\mu} + \mathbf{T}\mathbf{B}_t + \mathbf{v}_{t+1}$$

Então, o erro de predição de  $\mathbf{B}_{t+1}$  condicionado à amostra  $\tilde{\mathbf{y}}_t$  é dado pela expressão  $\boldsymbol{\eta}_{t+1|t} = \mathbf{B}_{t+1} - \boldsymbol{\mu} - \mathbf{T}\mathbf{B}_{t|t}$ , o que expressa a nova informação contida em  $\mathbf{B}_{t+1}$  que não está contida em  $\mathbf{B}_{t|t}$ . A variância do erro de predição é dado por  $\mathbf{f}_{t+1|t}^* = \mathbf{T}\mathbf{P}_{t|t}\mathbf{T}' + \mathbf{Q}$ . Adicionalmente, considere que  $\mathbf{H}^* = \mathbf{T}$  seja a matriz que relaciona o vetor observado  $\mathbf{B}_{t+1}$  com o não observado  $\mathbf{B}_t$ , então as equações de atualização podem ser escrita da forma:

$$\begin{aligned}
\mathbf{B}_{t|t, \mathbf{B}_{t+1}} &= E(\mathbf{B}_t | \tilde{\mathbf{y}}_t, \mathbf{B}_{t+1}) \\
&= E(\mathbf{B}_t | \mathbf{B}_{t|t, \mathbf{B}_{t+1}}) \\
&= \mathbf{B}_{t|t} + \mathbf{P}_{t|t} \mathbf{H}^{*'} \mathbf{f}_{t+1|t}^{*-1} \boldsymbol{\eta}_{t+1|t}^* \tag{6.31}
\end{aligned}$$

$$= \mathbf{B}_{t|t} + \mathbf{P}_{t|t} \mathbf{T}' (\mathbf{T} \mathbf{P}_{t|t} \mathbf{T}' + \mathbf{Q})^{-1} \boldsymbol{\eta}_{t+1|t}^* (\mathbf{B}_{t+1} - \boldsymbol{\mu} - \mathbf{T} \mathbf{B}_{t|t}) \tag{6.32}$$

$$\begin{aligned}
\mathbf{P}_{t|t, \mathbf{B}_{t+1}} &= \text{Cov}(\mathbf{B}_t | \tilde{\mathbf{y}}_t, \mathbf{B}_{t+1}) \\
&= \text{Cov}(\mathbf{B}_t | \mathbf{B}_{t|t, \mathbf{B}_{t+1}}) \\
&= \mathbf{P}_{t|t} + \mathbf{P}_{t|t} \mathbf{H}^{*'} \mathbf{f}_{t+1|t}^{*-1} \mathbf{H}^* \mathbf{P}_{t|t} \tag{6.33}
\end{aligned}$$

$$= \mathbf{P}_{t|t} + \mathbf{P}_{t|t} \mathbf{T}' (\mathbf{T} \mathbf{P}_{t|t} \mathbf{T}' + \mathbf{Q})^{-1} \mathbf{T} \mathbf{P}_{t|t} \tag{6.34}$$

$$\tag{6.35}$$

De modo geral, a derivação do suavizador é a mesma do que a dada pelo filtro de Kalman, porém os dados utilizados são os futuros e não os passados, mas o comportamento é o mesmo e as respectivas demonstrações são as mesmas.

### 6.3 Abordagem Bayesianiana

Novamente, considere o seguinte modelo linear das seções anteriores:

$$\mathbf{y}_t = \mathbf{x}_t \mathbf{B}_t + \mathbf{e}_t \quad \mathbf{e}_t \sim i.i.d.N(\mathbf{0}, \mathbf{R}) \tag{6.36}$$

$$\mathbf{B}_t = \tilde{\boldsymbol{\mu}} + \mathbf{T} \mathbf{B}_{t-1} + \mathbf{v}_t \quad \mathbf{v}_t \sim i.i.d.N(\mathbf{0}, \mathbf{Q}) \tag{6.37}$$

Onde  $\mathbf{y}_t$  é um vetor  $n$ -dimensional de variáveis endógenas,  $\mathbf{X}_t$  é um vetor  $k$ -dimensional, e os termos de erro  $\mathbf{e}_t$  e  $\mathbf{v}_t$  são vetores independentes entre si.

Da abordagem clássica, a estimação e conseqüente inferência sobre o vetor  $\tilde{\mathbf{B}}_T = [\mathbf{B}_1 \ \mathbf{B}_2 \ \dots \ \mathbf{B}_T]'$  é feita com base nos hiperparâmetros do modelo. Em contraste, a abordagem Bayesianiana considera que todos os hiperparâmetros, inclusive as variáveis de estado  $\tilde{\mathbf{B}}_T$ , são aleatórios. Outro ponto onde ambas abordagens diverge refere-se ao fato que a inferência sobre  $\tilde{\mathbf{B}}_T$  é feita conjuntamente com os hiperparâmetros, e não condicionalmente.

A ideia básica por trás da abordagem Bayesianiana é que ambas variáveis de estado e hiperparâmetros são tratados tal como dados omitidos, e devem ser gerados via simulação por métodos MCMC e só então a inferência pode ser feita.

Não existe uma maneira única de gerar o vetor  $\tilde{\mathbf{B}}_T$ . Neste trabalho é abordado o método sugerido em Carter e Kohn (1994) o qual é capaz de gerar todo vetor  $\tilde{\mathbf{B}}_T$  a partir da distribuição com base em toda a amostra  $p(\tilde{\mathbf{B}}_T | \tilde{\mathbf{y}}_T)$ . Esse método consiste então em obter a distribuição  $p(\tilde{\mathbf{B}}_T | \tilde{\mathbf{y}}_T)$  construída pela seguinte cadeia:

$$\begin{aligned}
p(\tilde{\mathbf{B}}_T|\tilde{\mathbf{y}}_T) &= p(\mathbf{B}_T|\tilde{\mathbf{y}}_T)p(\tilde{\mathbf{B}}_{T-1}|\mathbf{B}_T, \tilde{\mathbf{y}}_T) \\
&= p(\mathbf{B}_T|\tilde{\mathbf{y}}_T)p(\mathbf{B}_{T-1}|\mathbf{B}_T, \tilde{\mathbf{y}}_T)p(\tilde{\mathbf{B}}_{T-2}|\mathbf{B}_{T-1}, \mathbf{B}_T, \tilde{\mathbf{y}}_T) \\
&= p(\mathbf{B}_T|\tilde{\mathbf{y}}_T)p(\mathbf{B}_{T-1}|\mathbf{B}_T, \tilde{\mathbf{y}}_T)p(\mathbf{B}_{T-2}|\mathbf{B}_{T-1}, \mathbf{B}_T, \tilde{\mathbf{y}}_T) \\
&\quad \times p(\tilde{\mathbf{B}}_{T-3}|\mathbf{B}_{T-2}, \mathbf{B}_{T-1}, \mathbf{B}_T, \tilde{\mathbf{y}}_T)
\end{aligned}$$

Levando-se em consideração que  $\mathbf{B}_t$ , condicionado a  $\mathbf{B}_{T-1}$  e  $\tilde{\mathbf{y}}_t$  não tem mais informação sobre  $\mathbf{B}_{T-2}$  além da contida em  $\mathbf{B}_{T-1}$ , então os termos condicionantes podem ser omitidos de forma a simplificar:

$$\begin{aligned}
p(\tilde{\mathbf{B}}_T|\tilde{\mathbf{y}}_T) &= p(\mathbf{B}_T|\tilde{\mathbf{y}}_T)p(\mathbf{B}_{T-1}|\mathbf{B}_T, \tilde{\mathbf{y}}_T)p(\mathbf{B}_{T-2}|\mathbf{B}_{T-1}, \tilde{\mathbf{y}}_T)p(\tilde{\mathbf{B}}_{T-3}|\mathbf{B}_{T-2}, \tilde{\mathbf{y}}_T) \\
&\quad \dots p(\mathbf{B}_1|\mathbf{B}_2, \tilde{\mathbf{y}}_T)
\end{aligned}$$

Ainda, se considerarmos que  $\tilde{\mathbf{y}}_T$  não contém informação alguma para gerar  $\mathbf{B}_{t-i}$  além da disponível em  $\mathbf{B}_{t-i+1}$  e  $\tilde{\mathbf{y}}_{t-i}$ , então a expressão pode novamente ser simplificada:

$$\begin{aligned}
p(\tilde{\mathbf{B}}_T|\tilde{\mathbf{y}}_T) &= p(\mathbf{B}_T|\tilde{\mathbf{y}}_T)p(\mathbf{B}_{T-1}|\mathbf{B}_T, \tilde{\mathbf{y}}_{T-1})p(\mathbf{B}_{T-2}|\mathbf{B}_{T-1}, \tilde{\mathbf{y}}_{T-1})p(\tilde{\mathbf{B}}_{T-3}|\mathbf{B}_{T-2}, \tilde{\mathbf{y}}_{T-3}) \\
&\quad \dots p(\mathbf{B}_1|\mathbf{B}_2, \tilde{\mathbf{y}}_1) \\
&= p(\mathbf{B}_T|\tilde{\mathbf{y}}_T) \prod_{t=1}^{T-1} p(\mathbf{B}_t|\mathbf{B}_{t+1}, \tilde{\mathbf{y}}_t)
\end{aligned}$$

Em termos mais práticos, deve-se gerar então todo o vetor  $\tilde{\mathbf{B}}_T$  a partir de  $\mathbf{B}_T$ , o qual pode ser obtido pelo tradicional método do filtro de Kalman, e então gerar  $\mathbf{B}_t$  de modo retroativo a partir da distribuição condicionada ao vetor gerado anteriormente  $p(\mathbf{B}_t|\mathbf{B}_{t+1}, \tilde{\mathbf{y}}_t)$  para  $t = T-1, T-1, \dots, 1$ .

### 6.3.1 TVP-VAR

A constância nos parâmetros do VAR é comumente criticada quando o modelo é aplicado em um longo intervalo temporal. É razoável que se permita um modelo flexível onde os agentes trabalham com expectativas racionais, de forma que aprendam com as informações ao passo que essas vão sendo disponibilizadas, o que na essência, o método do filtro de Kalman faz nos modelos com parâmetros variantes, colocando-os na forma estado-espaco. Essa seção então se empenha em explicar como podem ser construída a inferência Bayesiana para o caso quando os parâmetros do VAR variam temporalmente, ou *Time-Varying Parameter VAR*, ou mesmo TVP-VAR. A base para a inferência Bayesiana

em VAR pode ser atribuída aos trabalhos de Cogley e Sargent (), Cogley e Sargent (2005) e Primiceri (2005).

Suponha agora o seguinte modelo VAR:

$$\mathbf{y}'_t = \mathbf{x}_t \mathbf{B}_t + \mathbf{e}'_t$$

Contendo o arranjo matricial na forma recorrente até então, onde  $\mathbf{y}_t$  e  $\mathbf{e}'_t$  são vetores  $n$ -dimensionais e  $\mathbf{x}'_t = [\mathbf{y}'_{t-1} \quad \mathbf{y}'_{t-2} \quad \dots \quad \mathbf{y}'_{t-p}]$  é um vetor  $k$ -dimensional, onde  $k = np$  e  $\mathbf{B}_t$  é uma matriz de dimensão  $k \times n$ . Vetorizando a matriz de coeficientes, o modelo é facilmente reescrito da forma:

$$\begin{bmatrix} \mathbf{y}_{1t} \\ \mathbf{y}_{2t} \\ \vdots \\ \mathbf{y}_{nt} \end{bmatrix} = \begin{bmatrix} \mathbf{x}'_t & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}'_t & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{x}'_t \end{bmatrix} \times \begin{bmatrix} \beta_{11}^{(1)} \\ \beta_{12}^{(1)} \\ \vdots \\ \beta_{1n}^{(1)} \\ \beta_{11}^{(2)} \\ \vdots \\ \beta_{1n}^{(p)} \\ \vdots \\ \beta_{n1}^{(1)} \\ \beta_{n1}^{(1)} \\ \vdots \\ \beta_{nn}^{(p)} \end{bmatrix}_t + \begin{bmatrix} \mathbf{e}_{1t} \\ \mathbf{e}_{2t} \\ \vdots \\ \mathbf{e}_{nt} \end{bmatrix}$$

$$\mathbf{y}_t = (\mathbf{I}_n \otimes \mathbf{x}'_t) \text{vec}(\mathbf{B})_t + \mathbf{e}_t$$

$$\mathbf{y}_t = \mathbf{W}_t \boldsymbol{\gamma}_t + \mathbf{e}_t$$

A ideia central para que os parâmetros variem no tempo é a mesma para o caso genérico abordado nas seções anteriores, deve-se atribuir um processo estocástico que permita que os parâmetros, agora  $\boldsymbol{\gamma}$  variem em  $t$ . Um processo comumente utilizado é o passeio aleatório que permite que os coeficientes hajam da seguinte forma:

$$\boldsymbol{\gamma}_{t+1} = \boldsymbol{\gamma}_t + \mathbf{v}_t$$

De outra forma, sob a suposição de normalidade dos erros do modelo,  $\mathbf{e}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma})$  e  $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{Q})$ , e independência entre  $\mathbf{e}_t$  e  $\mathbf{v}_t$ , a equação 6.3.1 diz que  $\boldsymbol{\gamma}_{t+1} | \boldsymbol{\gamma}, \mathbf{Q} \sim N(\boldsymbol{\gamma}_t, \mathbf{Q})$ . Analisando por outro sentido, essa distribuição serve de *priori* para a distribuição de  $\boldsymbol{\gamma}_{t+1}$  e então a *priori* para todos os estados  $t = 1, \dots, T$  dos parâmetros deve ser escrita como o produtório de todas as distribuições normais multiplicado pela *priori* do primeiro estado, quando  $t = 0$ ,  $p(\boldsymbol{\gamma}_0)$ , usualmente tida como normal:

$$\boldsymbol{\gamma}_0 \sim N(\boldsymbol{S}_{0|0}, \boldsymbol{P}_{0|0})$$

Sob outra ótica ainda, a equação de estado serve como base para a inferência Bayesiana se entendermos que, se  $\gamma_{t+1}|\gamma_t \sim N(\gamma_t, \boldsymbol{Q})$ , temos uma informação da distribuição *priori* para  $\gamma_{t+1}$ , sob a condição de que os  $\boldsymbol{\gamma}$  são independentes e identicamente distribuídos.

Tal como no VAR com parâmetros fixos, descrevemos *prioris* também para as distribuições dos parâmetros referente às variâncias. A distribuição de Wishart inversa, conjugada da normal, é uma boa opção:

$$\begin{aligned}\boldsymbol{\Sigma} &\sim iW(\underline{\boldsymbol{\Sigma}}, \underline{\nu}) \\ \boldsymbol{Q} &\sim iW(\underline{\boldsymbol{Q}}, \underline{\nu}_Q)\end{aligned}$$

A variação nos parâmetros então se mostra como uma complicação adicional para fazer previsões ao adicionar uma nova variabilidade na distribuição de predição através do termo de erro  $\boldsymbol{v}_t$ . Os valores  $\boldsymbol{\gamma}_{T+h}$  devem então ser simulados via métodos MCMC.

Após a observação da amostra  $\tilde{\boldsymbol{y}}_T$ , a *posteriori* para os parâmetros de variância são dados diretamente da Wishart inversa, tal como derivada anteriormente, a saber:

$$\begin{aligned}\boldsymbol{\Sigma}|\tilde{\boldsymbol{y}}_T, \tilde{\boldsymbol{B}}_T &\sim iW(\bar{\boldsymbol{\Sigma}}, \bar{\nu}) \\ \bar{\boldsymbol{\Sigma}} &= \underline{\boldsymbol{\Sigma}} + \sum_{t=1}^T (\boldsymbol{y}_t - \boldsymbol{w}_t \boldsymbol{\gamma}_t)(\boldsymbol{y}_t - \boldsymbol{w}_t \boldsymbol{\gamma}_t)', \quad \bar{\nu} = \underline{\nu} + T \\ \boldsymbol{Q}|\boldsymbol{\gamma}_T &\sim iW(\bar{\boldsymbol{Q}}, \bar{\nu}_Q) \\ \bar{\boldsymbol{Q}} &= \underline{\boldsymbol{Q}} + \sum_{t=1}^T (\boldsymbol{\gamma}_{t+1} - \boldsymbol{\gamma}_t)(\boldsymbol{\gamma}_{t+1} - \boldsymbol{\gamma}_t)' \quad \bar{\nu}_Q = \underline{\nu}_Q + T\end{aligned}$$

O método para calcular os parâmetros variantes do VAR, tal como fazer previsões sobre as variáveis endógenas é dada pelo seguinte algoritmo com base na amostragem de Gibbs:

#### *Amostragem de Gibbs para gerar TVP-VAR*

Para o modelo estado-espaco apresentado, sob as *prioris* Normal-Wishart apresentadas, o algoritmo necessita de valores iniciais para  $\boldsymbol{\Sigma}^{(0)}$  e  $\boldsymbol{Q}^{(0)}$ . Para  $J = B + R$ , são descartados os primeiros  $B$  valores gerados por questão de convergência, então:

- a) deve-se gerar através da posteriori  $\boldsymbol{\gamma}_T|\boldsymbol{y}_T, \boldsymbol{\Sigma}^{(j-1)}, \boldsymbol{Q}^{(j-1)} \sim N(\boldsymbol{S}_{T|T}, \boldsymbol{P}_{T|T})$  obtido do filtro de Kalman. Então para  $t = T - 1, T - 2, \dots, 1$  devem ser gerados

- $\gamma_t^{(j)}$  a partir da condicional  $\Gamma_t | \tilde{\mathbf{y}}_t, \Sigma^{(T-1)}, \mathbf{Q}^{(j-1)}, \gamma_{t+1}^{(j)}$  através do método de suavização;
- b) gerados então os valores para  $\gamma_T^{(j)}$ , deve-se gerar  $\mathbf{Q}^{(j)}$  a partir da condicional  $\mathbf{Q} | \gamma_T$ ;
- c) o próximo passo será gerar  $\Sigma^{(j)}$  através da condicional  $\Sigma | \tilde{\mathbf{y}}_T, \tilde{\gamma}_T$ ;
- d) deve-se então repetir o procedimento  $j = B + R$  vezes, afim de que sejam descartados os  $B$  valores iniciais obtidos dos parâmetros por questão de convergência;
- e) quando  $j > B$  então os erros aleatórios da equação de transição podem ser gerados  $h$  passos à frente pela distribuição  $\mathbf{e}_t \sim N(\mathbf{0}, \Sigma^{(j)})$ , para  $h = 1, \dots, H$ . Para os parâmetros  $\gamma_{T+h}^{(j)}$ , eles podem ser gerados pela equação de estado. Então recursivamente os  $\mathbf{y}_{T+h}^{(j)'}$  são calculados recursivamente pela expressão:

$$\begin{aligned}\tilde{\mathbf{y}}_{T+1}^{(j)'} &= \sum_{i=1}^p \mathbf{y}_{T+1-i}^{(j)'} \mathbf{B}_i^{(j)} + \mathbf{e}_{T+1}^{(j)'} \\ \tilde{\mathbf{y}}_{T+2}^{(j)'} &= \tilde{\mathbf{y}}_{T+1}^{(j)'} \mathbf{B}_1^{(j)} + \sum_{i=2}^p \mathbf{y}_{T+2-i}^{(j)'} \mathbf{B}_i^{(j)} + \mathbf{e}_{T+2}^{(j)'} \\ \tilde{\mathbf{y}}_{T+3}^{(j)'} &= \sum_{i=1}^2 \tilde{\mathbf{y}}_{T+3-i}^{(j)'} \mathbf{B}_i^{(j)} + \sum_{i=3}^p \mathbf{y}_{T+3-i}^{(j)'} \mathbf{B}_i^{(j)} + \mathbf{e}_{T+3}^{(j)'} \\ &\vdots \\ \tilde{\mathbf{y}}_{T+h}^{(j)'} &= \sum_{i=1}^{h-1} \tilde{\mathbf{y}}_{T+h-i}^{(j)'} \mathbf{B}_i^{(j)} + \sum_{i=h}^p \mathbf{y}_{T+h-i}^{(j)'} \mathbf{B}_i^{(j)} + \mathbf{e}_{T+h}^{(j)'}\end{aligned}$$

$$\text{Onde } \mathbf{B}^{(j)} = [\mathbf{B}_1^{(j)'} \quad \mathbf{B}_2^{(j)'} \quad \dots \quad \mathbf{B}_p^{(j)'}]'$$

Pegando apenas os vetores  $\{\tilde{\mathbf{y}}_{T+1}^{(j)}, \dots, \tilde{\mathbf{y}}_{T+H}^{(j)}\}_{j=B+1}^R$  tem-se então uma amostra de valores gerados independentes a partir da distribuição conjunta dos valores preditos.

A ideia básica por detrás do algoritmo é fazer o uso da estrutura linear da forma estado-espço. Dadas as matrizes de covariâncias  $\Sigma$  e  $\mathbf{Q}$ , o filtro de Kalman pode ser rodado na sua forma convencional, a progressiva, para  $t = 1, \dots, T$ , o que produzirá uma sequência de  $T$  valores para as distribuições condicionais  $\gamma_t | \tilde{\mathbf{y}}_t, \Sigma, \mathbf{Q} \sim N(\mathbf{S}_{t|t}, \mathbf{P}_{t|t})$ . Isso dará, então, uma *posteriori* condicional completa para o vetor  $\gamma_T | \tilde{\mathbf{y}}_T, \Sigma, \mathbf{Q} \sim N(\mathbf{S}_{T|T}, \mathbf{P}_{T|T})$ . Essa probabilidade pode então ser decomposta de forma que obtenhamos a *posteriori* condicional completa para os demais parâmetros, de acordo com a cadeia:

$$p(\tilde{\gamma}_T | \tilde{\mathbf{y}}_T, \Sigma, \mathbf{Q}) = p(\gamma_T | \tilde{\mathbf{y}}_T, \Sigma, \mathbf{Q}) \prod_{t=1}^{T-1} p(\gamma_t | \tilde{\mathbf{y}}_t, \gamma_{t+1}, \Sigma, \mathbf{Q})$$

Onde  $\gamma_t | \tilde{\mathbf{y}}_t, \gamma_{t+1}, \Sigma, \mathbf{Q} \sim N(\mathbf{S}_{t|T}, \mathbf{P}_{t|T})$ .

Alguns pontos relevantes devem ser destacados quanto esse modelo. O primeiro deles refere-se à estrutura passeio aleatório imposta ao comportamento dos parâmetros  $\gamma_t$ . Esses parâmetros são relativamente sensíveis à escolha de  $\mathbf{Q}$ . Escolher bem os elementos de  $\mathbf{Q}$ , priorizando valores baixos, implica que  $\gamma_t$  apresentará um comportamento temporal mais estável, suavizado. É importante atribuir baixa variância à  $\mathbf{v}_t$ , uma vez que o processo gerador tem raiz unitária e é não-estacionário. Em outras palavras, é interessante que os parâmetros não tenham um comportamento explosivo, uma vez que as variáveis endógenas tem um comportamento estacionário.

É comum se utilizar de um período pré-amstral, ou *training sample*, a fim de que os parâmetros iniciais sejam estimados e o modelo seja calibrado. Outra alternativa é usar uma *priori* do tipo Minnesota, a ser explicada com detalhamento no próximo capítulo.

### 6.3.2 TVP-VAR com volatilidade estocástica

A constância da matriz de covariâncias é contestável em muitos casos, especialmente quando o intervalo de tempo analisado é grande. Por exemplo, como no presente caso em finanças, é comum atribuir a variação da variância condicionalmente no tempo por conta do fenômeno do agrupamento de volatilidade. Para o caso quando a variância é dada, ou seja, não há necessidade de atribuir qualquer distribuição de probabilidade para ela, um processo determinístico como o tradicional MGARCH é uma via. Quando tal, a estimação dos parâmetros  $\gamma_t$  é feita diretamente.

Para o caso mais geral Baesiano, comumente é atribuída uma distribuição Wishart inversa para a matriz de covariâncias, um processo estocástico que condicione essa distribuição temporalmente é a Volatilidade Estocástica, que pode ser aplicado ao TVP-VAR Bayesiano, em outros termos, *Time-Varying Parameters with Stochastic Volatility* VAR, ou TVPSV-VAR.

Karlsson et al. (2012) argumenta que, dada a dificuldade de se distinguir empiricamente entre dois modelos, um com parâmetros constantes e variância não constante, e outro com parâmetros variantes e variâncias fixas, um modelo mais flexível que permita ambos é mais desejável. Os trabalhos iniciais atribuídos a essa abordagem podem ser referidos à Cogley e Sargent (2005) e Primiceri (2005).

Para introduzir as variâncias variantes no tempo do modelo TVPSV-VAR, é necessário que a matriz de covariâncias  $\Sigma$  seja decomposta de forma que:

$$\Sigma_t = \mathbf{L}_t^{-1} \mathbf{D}_t \mathbf{L}_t'^{-1}$$

Onde  $\mathbf{D}_t$  é uma matriz diagonal cujos elementos seguem um processo de volatilidade estocástica:

$$\mathbf{D}_t = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{bmatrix}_t \quad \mathbf{L}_t = \begin{bmatrix} 1 & 0 & \dots & 0 \\ l_{21} & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \dots \\ l_{n1} & l_{n2} & \dots & 1 \end{bmatrix}$$

E então os elementos de  $\mathbf{D}_t$  seguem um a um o processo:

$$d_{(i)t} = \exp\left(\frac{h_{it}}{2}\right) \quad (6.38)$$

$$h_{(i)t+1} = \mu_i + \phi_i(h_{it} - \mu_i) + \eta_{it} \quad i = 1, \dots, n \quad (6.39)$$

Onde  $\boldsymbol{\eta}_t = (\eta_{(1)t}, \eta_{(2)t}, \dots, \eta_{(n)t}) \sim i.i.d.N(\mathbf{0}, \mathbf{V}_\eta)$ . A ideia por trás dessa decomposição então é colocar a matriz de correlações  $\mathbf{L}_t$  de forma a variar temporalmente através de um passeio aleatório:

$$l_{(i)t+1} = l_{(i)t} + \zeta_{(i)t} \quad i = 1, \dots, n$$

Vetorizando os  $n(n-1)/2$  elementos não nulos fora da diagonal principal de  $\mathbf{L}_t$  de modo sempre a pegar da maior linha à menor e, chamando esse novo vetor de  $\mathbf{l}_t$ , a equação 6.3.2 pode ser reescrita como  $\mathbf{l}_{t+1} = \mathbf{L}_t + \boldsymbol{\zeta}_t$ , onde  $\boldsymbol{\zeta}_t \sim i.i.d.N(\mathbf{0}, \mathbf{V}_\zeta)$ , e  $\mathbf{V}_\zeta$  é uma matriz bloco-diagonal, com cada bloco correspondendo a uma linha em  $\mathbf{L}_t$ .

Para completar a inferência Bayesiana, é necessário que sejam atribuídas *prioris* sobre os parâmetros  $\mu_i$  e  $\phi_i$  da volatilidade estocástica e para as matrizes de covariâncias  $\mathbf{V}_\eta$  e  $\mathbf{V}_\zeta$ . Para o vetor das log-volatilidades incondicionais  $\boldsymbol{\mu} = [\mu_1 \ \mu_2 \ \dots \ \mu_n]'$  é comum utilizar uma *priori* não informativa  $\boldsymbol{\mu} \sim N(\mathbf{0}, \underline{\Sigma}_\mu)$ , cuja matriz de covariâncias  $\underline{\Sigma}_\mu$  é diagonal. Para o vetor de parâmetros  $\boldsymbol{\phi} = [\phi_1 \ \phi_2 \ \dots \ \phi_n]$  é comum utilizar uma *priori* normal restrita a região estacionária  $\boldsymbol{\phi} \sim N(\underline{\boldsymbol{\phi}}, \underline{\Sigma}_\phi)I(|\boldsymbol{\phi}| < 1)$  onde a matriz de covariâncias  $\underline{\Sigma}_\phi$  é diagonal. Alternativamente em alguns trabalhos, como Primiceri (2005), é plausível atribuir um passeio aleatório para os parâmetros  $\boldsymbol{\phi}$  desde que restrinja-se  $\boldsymbol{\mu} = \mathbf{0}$ , para o modelo não ficar "explosivo". Sob a primeira forma, tal como apresentada quando  $\phi_i$  restringe-se à região estacionária, então  $h_{(i)0}|\mu_i, \phi_i, \sigma_{\eta_i}^2 \sim N(\mu_i, \sigma_{\eta_i}^2/(1 - \phi_i^2))$ , onde  $\sigma_{\eta_i}^2$  é o  $i$ -ésimo elemento da diagonal principal de  $\mathbf{V}_\eta$ , é uma escolha apropriada. Para o segundo caso quando  $\phi_i$  segue um passeio aleatório, uma *priori* não informativa pode ser utilizada.

As distribuições de  $\mathbf{V}_\eta$  e  $\mathbf{V}_\zeta$  são do tipo Wishart inversa e, para este último caso, cada bloco matriz segue uma Wishart inversa independente dos demais blocos. De modo compacto:

$$\mathbf{V}_\eta \sim iW(\underline{\mathbf{S}}_\eta, \underline{\nu}_\eta) \quad (6.40)$$

$$\mathbf{V}_{\zeta,i} \sim iW(\underline{\mathbf{S}}_{\zeta,i}, \underline{\nu}_{\zeta,i}) \quad i = 1, \dots, n \quad (6.41)$$

Por fim, para os elementos em  $\mathbf{L}_0$  é adicionada a condição de que são distribuídos por uma normal não-informativa.

Um ponto importante a ser notado é que a inferência sobre os parâmetros do modelo da variância,  $\mathbf{L}_t$ ,  $\mathbf{V}_\zeta$ ,  $\mathbf{V}_\eta$ ,  $\mathbf{D}_t$ ,  $\boldsymbol{\mu}$  e  $\boldsymbol{\phi}$ , não dependem se os parâmetros do modelo VAR, dados em  $\mathbf{B}$ , variam no tempo ou não. Eles apenas estão condicionados uns aos outros. Então a construção das distribuições *posterioris* desses parâmetros pode ser feita com base em amostragens de Gibbs independentes. Vide que a relação inversa já não será válida, os parâmetros do modelo  $\mathbf{B}$  dependem, quanto à inferência, da variação da matriz de covariâncias  $\boldsymbol{\Sigma}_t$ .

Considere agora o seguinte modelo:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{W}_t \boldsymbol{\gamma}_t + \mathbf{e}_t \\ \boldsymbol{\gamma}_{t+1} &= \mathbf{d}_t + \mathbf{T}_t \boldsymbol{\gamma}_t + \mathbf{u}_t \end{aligned}$$

Se incorreremos ao caso quando os parâmetros do modelo são fixos temporalmente, temos então que este é um caso especial do TVP-VAR quando a matriz de covariâncias da equação de estado é nula,  $\mathbf{Q} = \mathbf{0}$ , com a adição das restrições que  $\mathbf{d}_t = \mathbf{0}$  e  $\mathbf{T}_t = \mathbf{I}$ . Quando tal, a *priori* é dada pelos parâmetros  $\mathbf{B}_{0|0}$  e  $\mathbf{P}_{0|0}$  de modo usual. Então, sob a hipótese de normalidade dos erros  $\mathbf{e}_t$ , a posteriori obtida é a tradicional com base na conjugação Normal-Wishart, cujos resultados já foram derivados.

Voltando para a variabilidade temporal da matriz  $\boldsymbol{\Sigma}_t$ , considere o vetor de erros  $\mathbf{e}_t = \mathbf{y}_t - \mathbf{W}_t \boldsymbol{\gamma}_t$ . Multiplicando-o pela matriz de correlações obtemos que  $\mathbf{L}_t \mathbf{e}_t = \mathbf{u}_t$ , cuja variância é  $\text{Var}(\mathbf{u}_t) = \mathbf{D}_t$ . Então pelo sistema, são obtidas  $n - 1$  equações independentes:

$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ l_{21} & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \dots \\ l_{n1} & l_{n2} & \dots & 1 \end{bmatrix} \times \begin{bmatrix} e_{(1)t} \\ e_{(2)t} \\ \vdots \\ e_{(n)t} \end{bmatrix} = \begin{bmatrix} u_{(1)t} \\ u_{(2)t} \\ \vdots \\ u_{(n)t} \end{bmatrix}$$

$$\begin{bmatrix} e_{(1)t} \\ e_{(2)t} + l_{(21)t}e_{(1)t} \\ e_{(3)t} + l_{(31)t}e_{(1)t} + l_{(32)t}e_{(2)t} \\ \vdots \\ e_{(n)t} + \sum_{j=1}^{i-1} l_{(ij)t}e_{(j)t} \end{bmatrix} = \begin{bmatrix} u_{(1)t} \\ u_{(2)t} \\ \vdots \\ u_{(n)t} \end{bmatrix}$$

Isolando o termo  $e_{(i)t}$  de cada equação, então elas podem ser escritas individualmente como:

$$e_{(i)t} = - \sum_{j=1}^{i-1} l_{(ij)t} e_{(j)t} + u_{(i)t}$$

Então o filtro de Kalman pode ser rodado naturalmente sobre cada equação independente para  $i = 1, \dots, (n-1)$ . Para visualizar melhor, o somatório do lado direito pode ser escrito na forma matricial:

$$e_{(i)t} = \mathbf{z}'_{i,t} \mathbf{l}_{i,t} + u_{(i)t}$$

Onde  $\mathbf{z}'_{i,t} = [e_{1,t}, \dots, e_{i-1,t}]$ ,  $\mathbf{l}_{i,t} = [l_{i1,t}, \dots, l_{i(i-1),t}]'$ ,  $u_{(i)t} = \exp(h_{i,t}/2)$ . A equação de estado tem a forma:

$$\mathbf{l}_{i,t+1} = \mathbf{I} \mathbf{l}_{i,t} + \boldsymbol{\zeta}_{i,t}$$

Onde  $\mathbf{I}$  é uma matriz identidade. A variação em  $\boldsymbol{\zeta}_i$  é dada pelos blocos da matriz bloco-diagonal  $\mathbf{V}_\zeta$ . Tal como nos casos anteriores, sob a hipótese de normalidade dos termos de erro  $\boldsymbol{\zeta}_t$ , então a posteriori para o bloco  $\mathbf{V}_{\zeta,i}$  é dado pela inversa da Wishart:

$$\mathbf{V}_{\zeta,i} | \tilde{\mathbf{l}}_{i,T} \sim iW(\bar{\mathbf{S}}_{\zeta,i}, \bar{\nu}_{\zeta,i}), \quad i = 2, \dots, n \quad (6.42)$$

$$\bar{\mathbf{S}}_{\zeta,i} = \underline{\mathbf{S}}_{\zeta,i} + \sum_{t=1}^T (\mathbf{l}_{i,t} - \mathbf{l}_{i,t-1})(\mathbf{l}_{i,t} - \mathbf{l}_{i,t-1})' \quad (6.43)$$

$$\bar{\nu}_{\zeta,i} = \underline{\nu}_{\zeta,i} + T \quad (6.44)$$

Com todas as informações corretamente especificadas como explicado acima, o filtro de Kalman torna-se direto e completo para os termos de correlação  $l_{ij,t}$ , já o mesmo não ocorre para a modelagem da variância  $h_{i,t}$  pois essa não é linear e o filtro já não pode ser aplicado diretamente na sua forma convencional. Para tal, é necessário que sejam feitas algumas manipulações. Tome novamente a equação que descreve o processo da variância:

$$u_{i,t} = \exp\left(\frac{h_{i,t}}{2}\right) v_{i,t}, \quad v_{i,t} \sim N(0, 1)$$

Visando linearizar a equação 6.3.2, deve-se elevá-la ao quadrado e então colocá-la em logaritmo:

$$u_{i,t}^2 = \left( \exp \left( \frac{2h_{i,t}}{2} \right) \right) v_{i,t}^2$$

$$\ln(u_{i,t}^2) = h_{i,t} + \ln(v_{i,t}^2) \quad (6.45)$$

$$u_{i,t}^* = h_{i,t} + v_{i,t}^* \quad (6.46)$$

Sabendo que, se  $v_{i,t} \sim N(0, 1)$ , então  $v_{i,t}^2 \sim \chi_1$  e  $\ln v_{i,t}^2 \sim \log \chi_1$ . Tendo um modelo linear em mãos, falta que os termos de erro sejam normalmente distribuídos para que o filtro de Kalman possa ser aplicado. Sabendo que, se  $v_{i,t} \sim N(0, 1)$ , então  $v_{i,t}^2 \sim \chi_1$  e  $\ln v_{i,t}^2 \sim \log \chi_1$ , o que não é gaussiano. Kim, Shephard e Chib (1998) aproxima a distribuição de  $\log \chi_1$  de uma mistura de distribuições normais, de forma que a f.d.p. pode ser dada como:

$$p(v_{i,t}^*) \approx \sum_{j=1}^7 q_j N(m_j - 1.2704, \tau_j^2)$$

Onde os coeficientes em  $j$  são dados pela tabela 1:

Tabela 1 – Coeficientes da mistura de distribuições normais para a aproximação  $\ln \chi_1^2$

j	$q_j$	$m_j$	$\tau_j^2$
1	0,00730	-10,12999	5,79596
2	0,10556	-3,97281	2,61369
3	0,00002	-8,56686	5,17950
4	0,04395	2,77786	0,16735
5	0,34001	0,61942	0,64009
6	0,24566	1,79518	0,34023
7	0,25750	-1,08819	1,26261

Fonte: Kim, Shephard e Chib (1998), p. 371

Outra forma de abordar a questão é entender que essa aproximação é nada mais do que uma média ponderada de 7 distribuições normais. Introduzindo uma variável discreta  $\delta_{i,t}$  que assume os valores no intervalo dos inteiros  $[1, 7]$ , entende-se que então:

$$v_{i,t}^* | \delta_{i,t} = j \sim N(m_j - 1.2704, \tau_j^2) \quad p(\delta_{i,t} = j) = q_j$$

O problema então finalmente pode ser reduzido à linearidade e a normalidade condicionados ao parâmetro  $\delta_{i,t}$ , para todo  $t = 1, \dots, T$ , e então o filtro poderá ser rodado. Para colocá-lo na forma estado-espaco, é necessário subtrair do modelo a média condicionada ( $m_j - 1.2704$ ) a fim de que os novos erros,  $v_{i,t}^{**}$  tenham valor esperado 0,  $E(v_{i,t}^{**}) = 0$ .

A equação de transição tem a forma:

$$\begin{bmatrix} u_{(1)t}^* \\ u_{(2)t}^* \\ \vdots \\ u_{(n)t}^* \end{bmatrix} = \begin{bmatrix} (m_j - 1.2704)_t & 1 & 0 & \dots & 0 \\ (m_j - 1.2704)_t & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (m_j - 1.2704)_t & 0 & 0 & \dots & 1 \end{bmatrix} \times \begin{bmatrix} 1 \\ h_{(1)t} \\ h_{(2)t} \\ \vdots \\ h_{(n)t} \end{bmatrix} + \begin{bmatrix} v_{(1)t}^{**} \\ v_{(2)t}^{**} \\ \vdots \\ v_{(n)t}^{**} \end{bmatrix} \mathbf{u}_t^* \quad (6.47)$$

$$= \mathbf{Z}_t \begin{pmatrix} 1 \\ \mathbf{h}_t \end{pmatrix} + \mathbf{v}_t^{**} \quad (6.48)$$

A variância de  $\mathbf{v}_t^{**}$  é dada por uma matriz diagonal cujas variâncias são dadas por  $\text{Var}(v_{i,t}^* | \delta_{i,t} = j) = \tau_j^2$

Necessitamos também da equação de estado, a qual pode ser colocada na forma:

$$\begin{bmatrix} h_{(1)t+1} \\ h_{(2)t+1} \\ \vdots \\ h_{(n)t+1} \end{bmatrix} = \begin{bmatrix} \mu_1(1 - \phi_1) \\ \mu_2(1 - \phi_2) \\ \vdots \\ \mu_n(1 - \phi_n) \end{bmatrix} + \begin{bmatrix} \phi_1 & 0 & \dots & 0 \\ 0 & \phi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \phi_n \end{bmatrix} \times \begin{bmatrix} h_{(1)t} \\ h_{(2)t} \\ \vdots \\ h_{(n)t} \end{bmatrix} + \begin{bmatrix} \eta_{(1)t} \\ \eta_{(2)t} \\ \vdots \\ \eta_{(n)t} \end{bmatrix}$$

$$\mathbf{h}_{t+1} = \mathbf{d} + \mathbf{T}\mathbf{h}_t = \boldsymbol{\eta}_t \quad (6.49)$$

Então para completar o filtro, é necessário atribuir distribuições de probabilidade para os estados  $\delta$ , naturalmente, a escolha é a normal gaussiana, então a *posteriori* é escrita como:

$$p(\delta_{i,t} = j | u_{i,j}^*, h_{i,t}) \propto q_j \frac{1}{\tau_j} \exp\left(-\frac{(u_{i,t}^* - m_j - 1,2704 - h_{i,t})^2}{2\tau_j^2}\right)$$

E também é necessário atribuir uma distribuição à matriz  $\mathbf{V}_\eta$ . Novamente a inversa de Wishart costuma ser uma escolha padrão. De forma compacta a relação é escrita como  $\mathbf{V}_\eta | \tilde{\mathbf{h}}_T, \boldsymbol{\mu}, \boldsymbol{\phi}$ . A *posteriori* tem sua forma analítica como:

$$\bar{\mathbf{S}}_\eta + \sum_{t=1}^T (\mathbf{h}_t - \mathbf{d} - \mathbf{T}\mathbf{h}_{t-1})(\mathbf{h}_t - \mathbf{d} - \mathbf{T}\mathbf{h}_{t-1})' \quad (6.50)$$

$$\bar{\nu}_\eta = \nu_\eta + T \quad (6.51)$$

A inferência sobre os parâmetros  $\boldsymbol{\mu}$  e  $\boldsymbol{\phi}$  pode ser feita isolando os parâmetros de interesse da parte variante no tempo. Para  $\mu$  então colocamos o modelo da forma:

$$\begin{bmatrix} h_{(1)t} - \phi_1 h_{(1)t-1} \\ h_{(2)t} - \phi_2 h_{(2)t-1} \\ \vdots \\ h_{(n)t} - \phi_n h_{(n)t-1} \end{bmatrix} = \begin{bmatrix} (1 - \phi_1) & 0 & \dots & 0 \\ 0 & (1 - \phi_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (1 - \phi_n) \end{bmatrix} \times \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} + \begin{bmatrix} \eta_{(1)t} \\ \eta_{(2)t} \\ \vdots \\ \eta_{(n)t} \end{bmatrix}$$

$$\mathbf{h}_t^* = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\eta}_t$$

E a *posteriori* de  $\boldsymbol{\mu}$  então toma a forma padrão:

$$\boldsymbol{\mu} | \tilde{\mathbf{h}}_T, \boldsymbol{\phi}, \mathbf{V}_\eta \sim N(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}}_\mu) \quad (6.52)$$

$$\boldsymbol{\Sigma}_\mu = (\boldsymbol{\Sigma}_\mu^{-1} + T \mathbf{X}' \mathbf{V}_\eta^{-1} \mathbf{X}_t)^{-1} \quad (6.53)$$

$$\bar{\boldsymbol{\mu}} = \bar{\boldsymbol{\Sigma}}_\mu \left( \boldsymbol{\Sigma}_\mu^{-1} \underline{\boldsymbol{\mu}} + \mathbf{X}' \mathbf{V}_\eta^{-1} \sum_{t=1}^T \mathbf{h}_t^* \right) \quad (6.54)$$

Para a inferência sobre os parâmetros  $\boldsymbol{\phi}$ , a mesma abordagem deve ser feita, de forma que, desta vez,  $\boldsymbol{\phi}$  deve ser isolado:

$$\begin{bmatrix} h_{(1)t} - \mu_1 \\ h_{(2)t} - \mu_2 \\ \vdots \\ h_{(n)t} - \mu_n \end{bmatrix} = \begin{bmatrix} (h_{(1)t-1} - \mu_1) & 0 & \dots & 0 \\ 0 & (h_{(2)t-1} - \mu_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (h_{(n)t-1} - \mu_n) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_n \end{bmatrix} + \begin{bmatrix} \eta_{(1)t} \\ \eta_{(2)t} \\ \vdots \\ \eta_{(n)t} \end{bmatrix}$$

$$\mathbf{h}_t^{**} = \mathbf{X}_t^* \boldsymbol{\phi} + \boldsymbol{\eta}_t$$

E a *posteriori* de  $\boldsymbol{\phi}$  toma a forma similar aos outros parâmetros já derivados:

$$\boldsymbol{\phi} | \tilde{\mathbf{h}}_T, \boldsymbol{\mu}, \mathbf{V}_\eta \sim N(\bar{\boldsymbol{\phi}}, \bar{\boldsymbol{\Sigma}}_\phi) I(|\phi_i| < 1) \quad (6.55)$$

$$\boldsymbol{\Sigma}_\phi = (\boldsymbol{\Sigma}_{phi}^{-1} + T \sum_{t=1}^T \mathbf{X}^{*'} \mathbf{V}_\eta^{-1} \mathbf{X}_t^*)^{-1} \quad (6.56)$$

$$\bar{\boldsymbol{\phi}} = \bar{\boldsymbol{\Sigma}}_\phi \left( \boldsymbol{\Sigma}_\phi^{-1} \underline{\boldsymbol{\phi}} + \sum_{t=1}^T \mathbf{X}^{*'} \mathbf{V}_\eta^{-1} \mathbf{h}_t^{***} \right) \quad (6.57)$$

Sumarizando, tendo todas essas informações disponíveis descritas acima, o algoritmo de Gibbs para a estimação dos parâmetros é dada pelas seguintes etapas:

#### *Amostragem de Gibbs para gerar TVPSV-VAR*

Para construir o TVPSV-VAR, é necessário selecionar:

- a) os valores iniciais para todas as matrizes de covariâncias dos erros do modelo em  $t = 1, \dots, T$ , então  $\tilde{\Sigma}_T^{(0)}$ ;
- b) os valores iniciais para todas as matrizes  $\mathbf{V}_{\zeta,i}$ , para  $i = 2, \dots, n$ ;
- c) os valores iniciais para os parâmetros da volatilidade estocástica  $\boldsymbol{\mu}^{(0)}$ ,  $\boldsymbol{\phi}^{(0)}$  e  $\mathbf{V}_\eta^{(0)}$ ;
- d) os valores iniciais da variável discreta  $\boldsymbol{\delta}_T^{(0)}$ .

Então, para  $j = B+R$  valores dos parâmetros gerados, quando os  $B$  valores iniciais são descartados por convergência, deve-se

- a) inicialmente os parâmetros da regressão  $\tilde{\gamma}_T$  são gerados pelo filtro de Kalman tradicional com base na distribuição *posteriori*  $\boldsymbol{\gamma}_t | \tilde{\mathbf{Y}}_t, \tilde{\Sigma}_t^{(j-1)} \sim N(\bar{\boldsymbol{\gamma}}_t, \bar{\Sigma}_\gamma)$  e então rodado o suavizador no modo convencional;
- b) para  $i = 1, \dots, n$  o filtro de Kalman deve ser rodado para:

$$\begin{aligned} \mathbf{l}_{i,t+1} &= \mathbf{I}\mathbf{l}_{i,t} + \boldsymbol{\zeta}_{i,t} \\ \mathbf{l}_t + 1 &= \mathbf{l}_t + \boldsymbol{\zeta}_t \end{aligned}$$

e então gerar  $\tilde{\mathbf{l}}_T^{(j)}$  a partir da distribuição *posteriori*  $\mathbf{l}_{i,T} | \mathbf{Y}_T, \mathbf{V}_{\zeta,i}^{(j-1)} \sim N(\mathbf{s}_{T|T}, \mathbf{P}_{T|T})$ , cujos parâmetros são obtidos do filtro de Kalman. Então obter  $\mathbf{l}_{i,t}^{(j)} | \tilde{\mathbf{y}}_t, \mathbf{V}_{\zeta,i}^{(j-1)}, \mathbf{l}_{i,t+1}^{(j)}$  do suavizador;

- c) então para  $i = 1, \dots, n$  deve-se gerar as matrizes  $\mathbf{V}_{\zeta,i}^{(j)}$  a partir da *posteriori* condicional  $\mathbf{V}_{i,\zeta} | \tilde{\mathbf{l}}_T^{(j)} \sim iW(\bar{\mathbf{S}}_{\zeta,i}, \bar{\nu}_{\zeta,1})$ ;
- d) o próximo passo é gerar as log-volatilidades  $\mathbf{h}_T^{(j)}$  a partir da distribuição *posteriori* condicional  $\mathbf{h}_T | \tilde{\mathbf{y}}_T, \boldsymbol{\mu}^{(j-1)}, \boldsymbol{\phi}^{(j-1)}, \mathbf{V}_{eta}^{j-1}, \tilde{\boldsymbol{\delta}}_T^{(j-1)} \sim N(\mathbf{s}_{T|T}, \mathbf{V}_{T|T})$ , cujos parâmetros são obtidos do filtro de Kalman. Então para  $t = T-1, \dots, 1$  os  $\mathbf{h}_t^{(j)}$  podem ser obtidos do suavizador, cuja *posteriori* condicional é dada pela distribuição de  $\mathbf{h}_T | \tilde{\mathbf{y}}_t, \boldsymbol{\mu}^{(j-1)}, \boldsymbol{\phi}^{(j-1)}$ ;
- e) para  $t = 1, \dots, T$ , os valores  $\delta_{i,t}$  devem ser gerados para todo  $i = 1, \dots, n$  de acordo com a *posteriori* condicional  $\delta_{i,t} = j | u_{i,j}^*, h_{i,t}$ ;
- f) deve-se então gerar a matriz  $\mathbf{V}_\eta^{(j)}$  da *posteriori* condicional  $\mathbf{V}_\eta | \tilde{\mathbf{h}}_T^{(j)}, \boldsymbol{\mu}^{(j-1)}, \boldsymbol{\phi}^{(j-1)} \sim iW(\bar{\mathbf{S}}_\eta, \bar{\nu}_\eta)$ ;
- g) gerar o vetor  $\boldsymbol{\phi}^{(j)}$  da *posteriori* condicional  $\boldsymbol{\phi} | \tilde{\mathbf{h}}_T^{(j)}, \boldsymbol{\mu}^{(j-1)}, \mathbf{V}^{(j)} \sim N(\bar{\boldsymbol{\phi}}, \bar{\Sigma}_\phi) I(|\phi_i| < 1)$ ;
- h) gerar o vetor  $\boldsymbol{\mu}^{(j)}$  da *posteriori* condicional  $\boldsymbol{\mu} | \tilde{\mathbf{h}}_T^{(j)}, \boldsymbol{\phi}^{(j)}, \mathbf{V}^{(j)} \sim N(\bar{\boldsymbol{\mu}}, \bar{\Sigma}_\mu)$ ;
- i) deve-se então repetir o procedimento  $j = B + R$  vezes, afim de que sejam descartados os  $B$  valores iniciais obtidos dos parâmetros por questão de convergência;

j) quando  $j > B$  então os erros aleatórios da equação de transição podem ser gerados  $h$  passos à frente pela distribuição  $\mathbf{e}_t \sim N(\mathbf{0}, \Sigma^{(j)})$ , para  $h = 1, \dots, H$ . Para os parâmetros  $\gamma_{T+h}^{(j)}$ , eles podem ser gerados pela equação de estado. Então recursivamente os  $\mathbf{y}_{T+h}^{(j)}$  são calculados recursivamente pela expressão:

$$\begin{aligned}\tilde{\mathbf{y}}_{T+1}^{(j)'} &= \sum_{i=1}^p \mathbf{y}_{T+1-i}^{(j)'} \mathbf{B}_i^{(j)} + \mathbf{e}_{T+1}^{(j)'} \\ \tilde{\mathbf{y}}_{T+2}^{(j)'} &= \tilde{\mathbf{y}}_{T+1}^{(j)'} \mathbf{B}_1^{(j)} + \sum_{i=2}^p \mathbf{y}_{T+2-i}^{(j)'} \mathbf{B}_i^{(j)} + \mathbf{e}_{T+2}^{(j)'} \\ \tilde{\mathbf{y}}_{T+3}^{(j)'} &= \sum_{i=1}^2 \tilde{\mathbf{y}}_{T+3-i}^{(j)'} \mathbf{B}_i^{(j)} + \sum_{i=3}^p \mathbf{y}_{T+3-i}^{(j)'} \mathbf{B}_i^{(j)} + \mathbf{e}_{T+3}^{(j)'} \\ &\vdots \\ \tilde{\mathbf{y}}_{T+h}^{(j)'} &= \sum_{i=1}^{h-1} \tilde{\mathbf{y}}_{T+h-i}^{(j)'} \mathbf{B}_i^{(j)} + \sum_{i=h}^p \mathbf{y}_{T+h-i}^{(j)'} \mathbf{B}_i^{(j)} + \mathbf{e}_{T+h}^{(j)'}\end{aligned}$$

onde  $\mathbf{B}^{(j)} = [\mathbf{B}_1^{(j)'} \quad \mathbf{B}_2^{(j)'} \quad \dots \quad \mathbf{B}_p^{(j)'}]'$ .

## 7 TÓPICOS ADICIONAIS

Este capítulo trata de dois tópicos adicionais necessários à complementação dos conceitos expostos anteriormente. O primeiro deles aborda a questão de qual distribuição *priori* escolher para a análise Bayesiana ser completa. Nos capítulos passados foram abordadas as propriedades estatísticas e montagem dos modelos sobre essa ótica, e ficou motivado o uso desse tipo de distribuição, necessária para o *start* da inferência, ou seja, essa distribuição pode ser interpretada como um dado adicional do modelo antes da apresentação da amostra. Ela está intrinsecamente ligada à crença do pesquisador sobre o comportamento dos dados pré-estimação. Através dela, a inferência Bayesiana partirá rumo à distribuição de probabilidade populacional dos parâmetros, feita com base na amostragem. Esse ponto de partida então é trivial na análise, uma vez que é responsável pela boa convergência das estimações. Ela, então, pode ser feita sobre diversas formas, dentre as quais, vale destacar duas.

A primeira trata-se de um período pré-amostral, o qual o pesquisador seleciona um conjunto de dados disponíveis anteriormente ao período da análise, o qual geralmente é feito uma estimação, geralmente por mínimos quadrados, para atribuir valores iniciais aos parâmetros. Esse tipo de abordagem é muito conhecida na literatura como *training-sample*. A segunda abordagem, explorada neste trabalho, trata-se de uma distribuição *priori* dos parâmetros exógena à observação de quaisquer dados, definida pelo pesquisador. É um método pronto, proposto pelo FED de Minneapolis amplamente utilizado na literatura pela praticidade, sobretudo quando não dispomos de dados excedentes, necessários ao *training-sample*. O trunfo dessa metodologia reside na simplificação e encolhimento do modelo, a ponto do pesquisador preocupar-se apenas na escolha, à sua arbitragem, de poucos parâmetros.

O segundo tópico adicional é uma extensão dos modelos BVAR recentemente proposta em Koop e Korobilis (2013). Apesar de todas as características adequadas à análise temporal linear proposta nesta pesquisa, os BVAR sofrem pelo alto custo computacional. Para montar VAR de baixa dimensão, basta utilizarmos os métodos convencionais Bayesianos até então expostos. Estender esses métodos em direção à alta dimensionalidade torna o problema de estimação infactível devido ao alto número de parâmetros. Novamente, este segundo tópico então trata de encolher o problema via simplificação, cabendo ao caso presente quando queremos aplicar o método BVAR para montar uma carteira amplamente diversificada, o que muito ocorre na prática.

### 7.1 *Priori* de Minnesota

A *priori* de Minnesota foi concebida nos primórdios da abordagem dos modelos VAR sob a ótica Bayesiana com a finalidade de reduzir o problema da dimensionalidade.

Ela foi elaborada por pesquisadores da *University of Minnesota* e do FED de Minneapolis e pode ser, sobretudo, atribuída aos trabalhos de Litterman (1979), Litterman (1985) e Litterman (1986), por vezes também conhecida como *priori* de Litterman.

A ideia inicial desta *priori* era ser utilizada em modelagem de dados macroeconômicos, feita através de aproximações, visando simplificar a estimação do VAR computacionalmente. Considere novamente o seguinte modelo VAR:

$$\mathbf{y}_t = \mathbf{a} + \sum_{m=1}^p \mathbf{B}_m \mathbf{y}_{t-m} + \sum_{o=1}^r \mathbf{H}_o \mathbf{Z}_{t-o} + \mathbf{e}_t \quad (7.1)$$

Onde  $\mathbf{y}$  é um vetor de variáveis endógenas  $n$ -dimensional variando no tempo  $t$ ,  $\mathbf{Z}$  é um vetor de variáveis exógenas  $s$ -dimensional, igualmente variante em  $t$  e  $\mathbf{e}$  é um vetor  $n$ -dimensional de choques aleatórios não correlacionados com o vetor  $\mathbf{Z}$  e  $\mathbf{e}_t \sim i.i.d.(\mathbf{0}, \Sigma)$ . Ainda  $\mathbf{a}$  é um vetor  $n$ -dimensional de médias incondicionais. O  $\mathbf{B}_m$  é um conjunto de matrizes  $n \times n$  que associam de forma linear o vetor de variáveis endógenas  $\mathbf{Y}$  com seus respectivos valores defasados até a ordem  $p$  e  $\mathbf{H}_o$  é um conjunto de matrizes  $n \times s$  que associam de forma linear o vetor de variáveis exógenas  $\mathbf{Z}$  defasados até a ordem  $r$  com o vetor de variáveis endógenas  $\mathbf{Y}$ .

A aproximação envolve substituir a matriz  $\Sigma$  por uma estimação  $\hat{\Sigma}$ . Na proposta inicial, restringia-se  $\Sigma$  a ser uma matriz diagonal, neste caso, cada equação do VAR poderia ser estimada separadamente e  $\text{Var}(e_i) = \sigma_{ii}$  e a estimação da variância poderia ser feita via M.Q.O.  $\hat{\sigma}_{ii} = s_i^2$ . Para o caso quando  $\Sigma$  não é diagonal, a estimação por M.Q.O. se mantém e  $\hat{\Sigma} = T^{-1} \mathbf{S}$ , onde  $\mathbf{S} = \sum_{t=1}^T \mathbf{e}_t \mathbf{e}_t'$ .

Para o caso quando  $\Sigma$  é substituída pela estimação, a *priori* de Minnesota volta-se para os demais parâmetros. Chamemos de  $\mathbf{b} = \text{vec}(\mathbf{B})$ , onde  $\mathbf{B}$  é o empilhamento padrão dos coeficientes do modelo, o qual já foi exposto. Então a *priori* assume que:

$$\mathbf{b} \sim N(\underline{\mathbf{b}}_{Min}, \underline{\mathbf{P}}_{Min})$$

Koop e Korobilis (2010) descreve que essa *priori* deve ser pensada como um meio de escolher automaticamente  $\underline{\mathbf{b}}_{Min}$  e  $\underline{\mathbf{P}}_{Min}$  de modo que seja sensível em diversos contextos empíricos. Os coeficientes de um VAR genérico pode ser dividido em três grupos de acordo com a natureza, o primeiro são os coeficientes referentes às variáveis endógenas quanto à elas mesma defasadas. O segundo grupo corresponde às variáveis defasadas de outras variáveis endógenas e o terceiro grupo reúne todos os coeficientes referentes à variáveis exógenas ou determinísticas, incluindo o termo de intercepto.

A *priori* de Minnesota atribui diferentes valores para o vetor  $\mathbf{b}$  de acordo com a natureza dos dados. Para o caso quando o VAR é não estacionário, geralmente quando rodado com dados macroeconômicos em nível, é aceitável que permita que estes

sigam um passeio aleatório. Então para as variáveis endógenas defasadas nelas mesmas,  $\mathbf{b}_{Min} = 1$ , já para os coeficientes das demais variáveis endógenas defasadas, é comum atribuir o valor  $\mathbf{b}_{Min} = 0$ . O outro caso ocorre quando o VAR é estacionário, ou seja, todas as variáveis endógenas já foram estacionarizadas ou são integradas de ordem 0 em nível, que é o caso dos retornos de ativos. Quando tal, a *priori* assume que todo o vetor  $\mathbf{b}_{Min} = 0$ .

Ademais, a *priori* atribui valores para  $\underline{\mathbf{P}}_{Min}$  de modo que esta seja uma matriz diagonal, ou seja, não há covariâncias entre os diferentes parâmetros  $\mathbf{b}_j$ , para  $j = 1, \dots, kn$ , onde  $k = np$  e  $p$  é o número de defasagens. Se pegarmos cada equação do VAR  $i$ , individualmente, chamamos a matriz de covariâncias dos coeficientes relativos à  $i$  como  $\underline{\mathbf{P}}_i$ . Ainda, os elementos da diagonal de  $\underline{\mathbf{P}}_i$  são chamados de  $\underline{\mathbf{P}}_{i,jj}$ , então a *priori* de Minnesota atribui os seguintes valores à esses elementos:

$$\underline{\mathbf{P}}_{i,jj} = \begin{cases} \frac{a_1}{p^2} & \text{para os coeficientes das variáveis defasadas nelas mesmas} \\ \frac{a_2}{p^2} \frac{\sigma_{ii}}{\sigma_{jj}} & \text{para os coeficientes das demais variáveis defasadas} \\ \underline{\mathbf{a}}_3 \sigma_{ii} & \text{para os coeficientes das variáveis exógenas} \end{cases}$$

A grande vantagem dessa *priori* de Minnesota é evitar complicações na escolha correta da especificação da *priori* no caso Bayesiano, reduzindo o problema à escolha dos três escalares  $\underline{a}_1, \underline{a}_2$  e  $\underline{a}_3$ . Sob esse formato, intuitivamente ao dividir  $\underline{a}_i$  por  $p^2$ , os coeficientes tendem a 0 conforme a defasagem aumenta. Intuitivamente a *priori* diz então que para defasagens distantes, podemos inferir com maior precisão que o coeficiente referente é 0.

Outro aspecto importante desta *priori* ocorre da especificação dos valores  $\underline{a}_i$ , comumente as próprias variáveis endógenas defasadas tem um poder explicativo para elas mesmas em  $t$  maior do que as demais variáveis endógenas defasadas. Isso pode ser expresso então colocando  $\underline{a}_1 > \underline{a}_2$ . Em outras palavras, o pesquisador tem uma crença mais forte de que os parâmetros  $\mathbf{b}$  são 0 referentes às defasagens das outras variáveis endógenas do que os parâmetros  $\mathbf{b}$  referentes à elas mesmas.

Uma das vantagens do uso desse método tange à derivação da distribuição *posteriori*. Sob as condições usuais de normalidade, observando-se a amostra  $\tilde{\mathbf{y}}_T$ , então temos a seguinte distribuição:

$$\overline{\mathbf{P}}_{Min}^{-1} = (\mathbf{X}'\mathbf{X} + \underline{\mathbf{P}}_{Min}^{-1}) \quad (7.2)$$

$$\overline{\mathbf{B}} = (\mathbf{X}'\mathbf{X} + \underline{\mathbf{P}}_{Min}^{-1})^{-1} (\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} + \underline{\mathbf{P}}_{Min}^{-1}\mathbf{B}) \quad (7.3)$$

$$= \overline{\mathbf{P}}_{Min} (\mathbf{X}'\hat{\mathbf{Y}} + \underline{\mathbf{P}}_{Min}^{-1}\mathbf{B}) \quad (7.4)$$

Ou de forma compacta, escrevemos que  $\mathbf{B}|\tilde{\mathbf{y}}_T \sim MN_{kn}(\overline{\mathbf{B}}_{Min}, \overline{\mathbf{P}}_{Min})$ .

Deve-se levar em consideração que a *priori* de Minnesota não é uma *priori* completa quanto a todos os parâmetros necessários para a estimação e inferência Bayesiana. Essa *priori* não diz nada sobre o comportamento da matriz de covariâncias dos erros do modelo  $\Sigma$ , sendo considerada dada. Uma vez que não trabalha com essa matriz, é arbitrário tomá-la por outros processos, tal como um processo determinístico, e.g. no casos quando é estimada via mínimos quadrados, ou GARCH, para o caso variante no tempo, ou então tomá-la como parâmetro desconhecido, a ser estimado via métodos MCMC e volatilidade estocástica.

## 7.2 Fatores de desconto

O problema da dimensionalidade faz com que sejam impostas maiores restrições no modelo VAR quanto maior o número de variáveis dependentes. Essa seção segue a metodologia proposta por Koop e Korobilis (2013). Esses autores, então, elaboram um método de calcular VAR de grande dimensão sob a ótica Bayesiana, usando fatores de desconto. Não havendo a necessidade de impor restrições como homoscedasticidade e constância dos parâmetros no tempo, o que, frente ao observado empiricamente, pode ser considerado demasiado, esse modelo se mostra uma alternativa eficaz para a estimação do VAR em duas vertentes principais.

A primeira delas diz respeito à necessidade de que os parâmetros variem no tempo, então é comum a necessidade de trabalhar com modelos de volatilidade estocástica multivariados. A segunda delas refere-se à necessidade de que as restrições impostas ao modelo devem variar no tempo, com o intuito de sanar o problema da sobre-parametrização típica de modelos multivariados com muitas variáveis dependentes.

Os autores utilizam a justificativa de que modelos de baixa dimensão conseguem ser mais flexíveis, podendo-se trabalhar com a heteroscedasticidade dos erros aleatórios e variação temporal dos parâmetros. Estender os tradicionais métodos utilizados para esses modelos para aqueles com grande dimensão não é direto, por questões computacionais. Os modelos TVP-VARs demandam o uso de métodos MCMC para que a inferência bayesiana seja feita, o que em modelos de alta dimensão pode se tornar uma tarefa árdua.

O modelo proposto permite então que os parâmetros variem no tempo de acordo com um passeio aleatório. Algebricamente pode ser descrito pelas expressões:

$$\mathbf{y}'_t = \mathbf{x}'_t + \mathbf{B}_t + \mathbf{e}_t \quad (7.5)$$

$$\mathbf{B}_{t+1} = \mathbf{B}_t + \mathbf{u}_t \quad (7.6)$$

Onde  $\mathbf{e}_t \sim i.i.d.N(\mathbf{0}, \Sigma_t)$  e  $\mathbf{u}_t \sim i.i.d.N(\mathbf{0}, \mathbf{Q}_t)$  são termos de erro,  $\mathbf{y}'_t$  é um vetor em  $\mathbb{R}^n$  de variáveis endógenas e  $\mathbf{x}'_t = [\mathbf{y}'_{t-1} \quad \dots \quad \mathbf{y}'_{t-p}]$  é um vetor  $k$ -dimensional, onde  $k = np$ .

Esse modelo, na sua forma tradicional, necessita de que o pesquisador selecione *prioris* para  $\mathbf{B}_0$ ,  $\mathbf{\Sigma}_0$  e  $\mathbf{Q}_0$  e especifique o processo estocástico que gera os dados desses parâmetros. A inferência então pode ser feita através dos métodos MCMC através da forma como o modelo estado-espaço está estruturado.

Como porém já argumentado, esses métodos sofrem pelo seu alto custo computacional quando que, para uma boa inferência estatística, geralmente são necessitados milhares de simulações. No contexto quando necessita-se calcular a distribuição de probabilidade preditiva de forma recursiva, como no caso de finanças em que os parâmetros precisam ser atualizados diariamente, essa tarefa torna-se ainda mais custosa.

Seja  $\tilde{\mathbf{y}}_s = (\mathbf{y}_1, \dots, \mathbf{y}_s)$  o conjunto de toda amostra obtida até  $t = s$ , o filtro de Kalman diz que:

$$\mathbf{B}_{t-1}|\tilde{\mathbf{y}}_{t-1} \sim N(\mathbf{B}_{t-1|t-1}, \mathbf{V}_{t-1|t-1})$$

Então o filtro promove a predição de acordo com:

$$\mathbf{B}_t|\tilde{\mathbf{y}}_{t-1} \sim N(\mathbf{B}_{t|t-1}, \mathbf{V}_{t|t-1})$$

Onde

$$\mathbf{V}_{t|t-1} = \mathbf{V}_{t-1|t-1} + \mathbf{Q}_t$$

Sendo este o único ponto onde o filtro utiliza-se a informação disponível em  $\mathbf{Q}_t$ , os autores propõem fazer a seguinte substituição:

$$\mathbf{V}_{t|t-1} = \frac{1}{\lambda} \mathbf{V}_{t-1|t-1}$$

O uso do  $\lambda$  substitui a necessidade de estimação da matriz  $\mathbf{Q}_t$ . Esse valor é chamado de fator de desconto, restrito no intervalo  $0 < \lambda < 1$ . Uma interpretação que sai diretamente do uso deste fator é que observações  $j$  períodos passados tem um peso  $\lambda^j$  na estimação de  $\mathbf{B}_t$ .

Substituindo o lado direito de 7.2 em 7.2, obtemos a relação:

$$\lambda^{-1} \mathbf{V}_{t-1|t-1} = \mathbf{V}_{t-1|t-1} + \mathbf{Q}_t \quad (7.7)$$

$$\mathbf{Q}_t = (\lambda^{-1} - 1) \mathbf{V}_{t-1|t-1} \quad (7.8)$$

Botando o fator de desconto no extremo  $\lambda = 1$ , verificamos que não há variação da matriz  $\mathbf{V}_{t|t-1}$ , equivalente ao caso quando os coeficientes são constantes. O uso do fator

de desconto pode ser encontrado em outros trabalhos na literatura, porém a contribuição desses autores encontra-se na busca da correta especificação do valor assumido por  $\lambda$ , ao contrário dos demais autores que se utilizam de um valor escolhido exogenamente perto de 1, no intuito de que os parâmetros  $\mathbf{B}_t$  variem suavemente ao longo do tempo.

Indo além, da necessidade de se estimar a matriz de covariâncias  $\Sigma_t$ , o uso de modelos de volatilidade estocástica pode ser driblado igualmente com o uso de fatores de desconto. Para tal abordagem, é escolhido o uso do método *Exponentially Weighted Moving Average* (EWMA) de modelagem da variância:

$$\hat{\Sigma}_t = \kappa \hat{\Sigma}_{t-1} + (1 - \kappa) \hat{\mathbf{e}}_t \hat{\mathbf{e}}_t'$$

Onde  $\hat{\mathbf{e}}_t = \mathbf{y}_t - \mathbf{x}_t \mathbf{B}_{t|t}$  é o termo de erro aleatório produzido pelo filtro de Kalman,  $\kappa$  é um fator de decaimento, também a ser estimado no intervalo  $0 < \kappa < 1$ . O intervalo de valores mais amplamente aceito segue a metodologia RiskMetrics (1996), a qual sugere que  $0,94 < \kappa < 0,98$ . Para a condição inicial  $\hat{\Sigma}_0$  os autores utilizam uma amostra inicial de  $\tilde{\mathbf{y}}_T$ , também conhecida como *training sample*, geralmente utilizada para calibrar modelos.

### 7.2.0.1 Seleção dinâmica de modelos

Comumente chamada na literatura de *Dinamic Model Selection* (DMS), a seleção dinâmica de modelos permite que haja mudança de regime entre diferentes especificações, o que na prática pode incorporar mudanças mais bruscas no processo gerados dos retornos, para o caso em finanças, ao passo que os modelos tradicionais TVP-VAR são bem ajustados quanto à mudanças suaves nos parâmetros. Os autores utilizam-se dessa metodologia com duas finalidades, a primeira tange o problema da dimensionalidade, permitir que o VAR mude o número de regressores ao longo do tempo, conforme a necessidade, torna a estimação mais parcimoniosa, desprezando aquelas variáveis que não tem alto poder explicativo quando necessário. A segunda finalidade refere-se à escolha dos parâmetros  $\lambda$  e  $\kappa$  de modo que possam variar ao longo de  $t$ .

Dentro das possibilidades que o DMS traz, Koop e Korobilis (2013) estuda o uso do *Dinamic Model Average* (DMA) a fim de calcular qual a probabilidade do modelo  $j$  em prever o vetor  $\mathbf{y}_t$ , dada a informação disponível em  $\tilde{\mathbf{y}}_{t-1}$ , probabilidade esta chamada  $\pi_{t|t-1,j}$ , para  $j = 1, \dots, J$ . Tendo em mãos as probabilidades para os  $j$  modelos, é facultativo o uso da seleção de modelos ou da média ponderada destes. Outro ponto a ser destacado é que, uma vez permitido que  $\pi_{t|t-1,j}$  varie em  $t$ , os modelos de previsão podem variar no tempo. Raftery, Kárný e Ettlér (2010) desenvolve um algoritmo de fácil execução para obter esse conjunto de probabilidades.

A natureza do DMS envolve selecionar, para este caso, modelos que incluem TVP-

VAR de diferentes dimensões, *prioris*, tal como diferentes fatores de desconto e taxas de decaimento. (Raftery et al, 2010) deriva a probabilidade de que o modelo  $j$  deve ser usado para prever o  $\mathbf{y}_t$  de forma análoga a como o filtro de Kalman utiliza seu método de predição e atualização:

$$\pi_{t|t-1,j} = \frac{\pi_{t-1|t-1}^\alpha}{\sum_{l=1}^J \pi_{t-1|t-1,l}^\alpha}$$

Onde  $\pi_{t|t-1,j}$  é a probabilidade do uso do modelo  $j$  com a informação disponível até  $\tilde{\mathbf{y}}_{t-1}$  e  $\alpha$  é um fator de desconto. Com a chegada da nova informação  $\mathbf{y}_t$ , a equação 7.2.0.1 é atualizada como:

$$\pi_{t|t,j} = \frac{\pi_{t|t-1} p_j(\mathbf{y}_t | \tilde{\mathbf{y}}_{t-1})}{\sum_{l=1}^J \pi_{t|t-1,l} p_l(\mathbf{y}_t | \hat{\mathbf{y}}_{t-1})}$$

Onde  $p_j(\mathbf{y}_t | \tilde{\mathbf{y}}_{t-1})$  é a f.d.p. da verossimilhança para a predição, em outras palavras, a função de densidade para o modelo  $j$  avaliado em  $\mathbf{y}_t$ , a qual pode ser considerada uma medida do desempenho da previsão.

De modo compacto, esquiva-se do uso de métodos MCMC, sendo necessário apenas o cômputo de  $\pi_{t|t-1,j}$  e  $\pi_{t|t,j}$ . O uso do fator de desconto para a probabilidade tem as mesmas implicações matemáticas dos demais fatores de desconto, alternativamente,  $\pi_{t|t-1,j}$  é entendido como:

$$\begin{aligned} \pi_{t|t-1,j} &\propto [p_j(\mathbf{y}_{t-1} | \tilde{\mathbf{y}}_2)]^\alpha \times [p_j(\mathbf{y}_{t-2} | \tilde{\mathbf{y}}_3)]^{\alpha^2} \times \dots \times [p_j(\mathbf{y}_1 | \tilde{\mathbf{y}}_0)]^{\alpha^{t-1}} \\ &= \prod_{i=1}^{t-1} [p_j(\mathbf{y}_{t-i} | \tilde{\mathbf{y}}_{(t-i)-1})]^\alpha \end{aligned}$$

Em palavras, a relação acima diz que o quão bom for a probabilidade do modelo  $j$  em prever os resultados recentes, maior será o peso de seu uso para prever em  $t$ . A habilidade em prever eventos recentes decai pelo fator  $\alpha$  que, quanto mais próximo de 1, maior será a persistência. Os autores então sugerem o uso do intervalo  $0,95 \leq \alpha \leq 1$  para o fator de desconto. Um ponto adicional a ser destacado ocorre quando  $\alpha$  tende ao valor extremo 0, neste caso, o método atribui pesos iguais para todos os modelos independentemente de  $t$ .

A metodologia DSM ainda permite que sejam escolhidas diferentes *prioris* para o modelo. Em seu trabalho, Koop e Korobilis (2013) utilizam-se de uma combinação da *priori* de Minnesota, onde  $\mathbf{B}_0$  tem uma distribuição normal multivariada. A escolha da *priori* baseia-se então em diferentes valores para  $\gamma$ , onde:

$$\underline{V}_i = \begin{cases} \frac{\gamma}{p^2} & \text{para } p = 1, \dots, P \\ \underline{a} & \text{para os interceptos} \end{cases}$$

Em que  $\text{var}(\mathbf{B}_0) = \mathbf{V}$  e  $V_i$  são os elementos da diagonal principal e  $p = 1, \dots, P$  é o número de defasagens do modelo VAR. Na prática geralmente é escolhido um valor alto para  $a$  com fim de tornar a *priori* não informativa.

Outro modo ainda como DMS trabalha tange à escolha da dimensão do VAR. A escolha das variáveis que entram no modelo é arbitrária e depende fortemente da teoria que embasa o modelo. Como exemplo, em seu trabalho, os autores aplicam o VAR para dados macroeconômicos com base na escolha de três modelos diferentes, um pequeno, um médio e outro de alta dimensão, com 3, 7 e 25 variáveis respectivamente.

O DMS funciona então como uma combinação de todos os modelos possíveis, definidos pelo pesquisador. Por exemplo, no trabalho de Koop e Korobilis (2013) os autores estipulam 4 valores para  $\lambda$  e 3 para  $\kappa$ , o que dá 12 combinações diferentes. Para  $\gamma$  são atribuídos 6 valores diferentes e, sobre a possibilidade de 3 dimensões distintas do VAR, são obtidos  $j = 12 \times 6 \times 3 = 216$  modelos diferentes. Para todos esses modelos são calculadas as probabilidades definidas por  $\pi_{t|t-1,j}$ , o DMS selecionará, então, a maior probabilidade e a predição será feita com base nos parâmetros relativos a tal, ao passo que o DMA fará uma média ponderada de todos, cuja ponderação é feita pela própria probabilidade.

Por fim, para dar robustez aos resultados, são calculados também diferentes valores para  $\alpha$ , a saber  $[0,001 \quad 0,95 \quad 0,99 \quad 1]$ , os quais indicam uma maior persistência temporal, do menor ao maior valor, dos modelos em prever os resultados em  $t$ .

Por sim, para completar essa abordagem, é necessário ainda que haja a escolha da *priori* para  $\pi_{0|0}$ . Os autores então sugerem que o modelo parta de uma distribuição equilibrada, onde intuitivamente o pesquisador não tenha palpite algum sobre qual parâmetro se ajusta melhor aos dados, o que equivalentemente é dado um peso igual para todos os modelos, no caso  $\pi_{0|0,j} = 1/216$  ou, de forma genérica,  $\pi_{0|0,j} = 1/j$ .

## 8 CONSTRUÇÃO DOS MODELOS E INFERÊNCIA

O objetivo deste capítulo é verificar empiricamente se há dependência temporal na média dos retornos. Grande parte da literatura ignora essa relação e se empenha em modelar o segundo momento, a variância, devido ao fenômeno do agrupamento da volatilidade, já amplamente estudado pelos modelos econométricos. Se há algum tipo de dependência temporal nos retornos, os investidores podem explorar esse tipo de informação afim de obter um melhor desempenho de suas carteiras.

O modelo proposto para esse tipo de análise é o VAR, o qual é capaz de captar quaisquer relações lineares entre os diferentes retornos. O tipo de abordagem utilizada assemelha-se à proposta em DeMiguel, Nogales e Uppal (2014), com ressalva de que o método utilizado para a estimação dos parâmetros é Bayesiana.

Os modelos VAR sob a abordagem Bayesiana vem ganhando força especialmente na literatura em Macroeconomia. Essa forma de estimação contrasta com a abordagem clássica no que tange à restrição dos parâmetros. Ao passo que a disponibilidade de dados macroeconômicos não é abundante, os modelos VAR sofrem quanto à inferência devido ao problema da dimensionalidade. Para contornar tal dificuldade, não é comum restringir alguns parâmetros do modelo a fim de que se possa fazer inferência sobre outros, o que de forma geral, empobrece o método, sobretudo se sua finalidade for fazer previsões.

Para motivar mais a atenção quanto ao problema da dimensionalidade, leve em consideração que para que a inferência possa ser feita, é necessário que a matriz de erros dos parâmetros seja homoscedástica, o que na prática é uma suposição irreal na maioria das vezes. Atentando-se a esse fato, é necessário, além da estimação dos parâmetros do VAR, estimar os parâmetros da dependência temporal da variância, o que torna a estimação infactível.

O BVAR, assim chamado o VAR Bayesiano, trabalha com reamostragens e quaisquer restrições nos parâmetros podem ser colocadas na distribuição *priori*. Uma vez que os dados são rodados, então a distribuição se ajusta ao que foi observado.

Voltando-se para o contexto em finanças, o proposto é a modelagem da média e da variância conjuntamente. Apesar de que essa área não sofre com a escassez de dados tal como a macroeconomia, modelos em finanças geralmente tem alta dimensionalidade, o que torna atrativo o uso da abordagem Bayesiana.

A metodologia para a estimação dos parâmetros segue de acordo com a proposta em Primiceri (2005). Nesse trabalho, o BVAR volta-se a previsão de dados macroeconômicos. Sob pequenas modificações, a metodologia pode ser aplicada ao que se pretende verificar para o presente caso em finanças.

Os *Softwares* utilizados para todos os cálculos foram o Excel e o Matlab e a coleta

de dados das séries históricas dos preços das ações foi feita através do Economatica, já devidamente ajustadas por proventos, e para os índices, foram coletados os preços no endereço eletrônico da BM&FBovespa.

### 8.0.1 Metodologia

A estrutura do VAR segue a sugerida em DeMiguel, Nogales e Uppal (2014). Os autores entendem que a dependência temporal na média dos retornos é fraca, logo não há necessidade de defasagens maiores do que  $p = 1$ . Desta forma a estrutura seguinte é geral para todas as especificações feitas nesse capítulo:

$$\mathbf{y}_t = \mathbf{c}_t + \mathbf{B}_t \mathbf{y}_{t-1} + \mathbf{e}_t, \quad t = 1, \dots, T$$

Onde  $\mathbf{y}_t$  é um vetor  $n$ -dimensional de variáveis endógenas variando em  $t$  e  $\mathbf{B}_t$  e  $\mathbf{c}_t$  são os parâmetros do modelo variantes no tempo. O termo de erro  $\mathbf{e}_t$  é um vetor  $n$ -dimensional normalmente distribuído com média  $E(\mathbf{e}_t) = \mathbf{0}$  para todo  $t$ , porém heteroscedástico, cuja matriz de covariâncias é dada por  $\Sigma_t$  e pode ser decomposta da seguinte forma:

$$\Sigma_t = \mathbf{A}_t \mathbf{D}_t \mathbf{A}_t$$

Onde  $\mathbf{A}_t$  é uma matriz triangular inferior de correlações:

$$\mathbf{A}_t = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \alpha_{21} & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \dots & 1 \end{bmatrix}_t$$

E a matriz  $\mathbf{D}_t$  é uma matriz diagonal de variâncias:

$$\mathbf{D}_t = \begin{bmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{nn} \end{bmatrix}_t$$

Então, os erros do modelo podem ser padronizados como:

$$\mathbf{y}_t = \mathbf{c}_t + \mathbf{B}_t \mathbf{y}_{t-1} + \mathbf{A}_t^{-1} \mathbf{D}_t^{\frac{1}{2}} \mathbf{e}_t, \quad t = 1, \dots, T$$

Por fim, o modelo é reescrito da seguinte forma:

$$\mathbf{y}_t = \mathbf{I}_n \otimes [1 \ \mathbf{y}'_{t-1}] \text{vec}(\mathbf{B}') + \mathbf{A}_t^{-1} \mathbf{D}_t^{\frac{1}{2}} \mathbf{e}_t \quad (8.1)$$

$$= \mathbf{X}'\boldsymbol{\beta} + \boldsymbol{\epsilon}_t \quad (8.2)$$

A dinâmica do modelo, então, pode ser descrita pelo seguinte sistema de equações na forma estado-espço:

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \mathbf{v}_t \quad (8.3)$$

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} + \boldsymbol{\zeta}_t \quad (8.4)$$

$$\ln \boldsymbol{\sigma}_t = \ln \boldsymbol{\sigma}_{t-1} + \boldsymbol{\eta}_t \quad (8.5)$$

Os elementos de  $\boldsymbol{\beta}_t$  postos na forma acima seguem um passeio aleatório, uma suposição importante para que os parâmetros do modelo tenham a liberdade de assumirem quaisquer valores sem que haja a necessidade de impor estacionariedade sobre eles. Os parâmetros do vetor  $\boldsymbol{\alpha}_t$  igualmente seguem um passeio aleatório, sob a restrição de que os  $\alpha_{i,t}$  são independentes entre as linhas  $i$  da matriz de correlações  $\mathbf{A}_t$ . Por último, as variâncias  $\boldsymbol{\sigma}_t$  seguem um passeio aleatório geométrico, típico da classe de modelos de volatilidade estocástica.

Quanto à variância dos termos de erro  $\boldsymbol{\epsilon}_t$ ,  $\mathbf{v}_t$ ,  $\boldsymbol{\zeta}_t$  e  $\boldsymbol{\eta}_t$ , é assumido que não há correlação entre os diferentes vetores, então de forma matricial, a seguinte relação é válida:

$$\text{Var} \left( \begin{pmatrix} \boldsymbol{\epsilon} \\ \mathbf{v}_t \\ \boldsymbol{\zeta}_t \\ \boldsymbol{\eta}_t \end{pmatrix} \right) = \begin{bmatrix} \mathbf{I}_n & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{W} \end{bmatrix} \quad (8.6)$$

A escolha no método Bayesiano de estimação dos parâmetros é defendida por Primiceri (2005) sob diferentes pontos de vista. Para o autor, quando os parâmetros são variantes, a distinção entre os estados não observáveis e os erros aleatórios não é tão clara, a abordagem Bayesiana trata deste tópico com mais naturalidade. Outro ponto positivo no uso decorre da não linearidade do modelo e da alta dimensionalidade, quando tal, é normal que a verossimilhança apresente um comportamento não desejado como múltiplos picos, o que da inferência clássica, pode levar a estimações não desejadas. A estimação Bayesiana novamente trata do problema com naturalidade quando que, da especificação das *prioris*, podemos restringir o espaço paramétrico ao desejado. Outro ponto de ganho da inferência Bayesiana para o caso dos parâmetros variantes encontra-se no processo recursivo de estimação, o que fornece estimativas suavizadas para o parâmetro, ao passo que da clássica, apenas o filtro é utilizado.

Tal como Primiceri (2005), é assumido que os estados iniciais dos coeficientes, covariâncias, log-volatilidades e hiperparâmetros são independentes uns dos outros. *A priori* para os parâmetros  $\beta_0$ ,  $\alpha_0$  e  $\ln \sigma_0$  segue uma distribuição normal. Então *a priori* para todos os parâmetros  $\tilde{\beta}_T$ ,  $\tilde{\alpha}_T$  e  $\ln \tilde{\sigma}_T$  também são normais condicionadas às distribuições das matrizes de covariâncias, as quais são distribuídas pela Wishart inversa:

$$\begin{aligned}\beta_0 &\sim N(\beta_{Min} = \mathbf{0}, \text{Var}(\beta_{Min}) = \mathbf{I}) \\ \mathbf{A}_0 &\sim N(\mathbf{A}_{Min} = \mathbf{0}, \text{Var}(\mathbf{A}_{Min}) = \mathbf{I}) \\ \ln \sigma_0 &\sim N(\mathbf{0}, \mathbf{I}) \\ \mathbf{Q} &\sim iW(k_Q^2 \text{Var}(\beta_{Min}), 1) \\ \mathbf{W} &\sim iW(k_W^2 \mathbf{I}_n, (1+n)) \\ \mathbf{S}_i &\sim iW(k_S^2, (1+i)\text{Var}(\mathbf{A}_{Min}), (1+i)), \quad i = 2, \dots, n\end{aligned}$$

Onde:

- a)  $n$  - tamanho da amostra;
- b) Min - *Priori* de Minnesota, com parâmetro escolhido igual a 1 para todos os casos;
- c)  $k_Q = 0,01$ ,  $k_S = 0,1$  e  $k_W = 0,01$  de acordo com Primiceri (2005) para deixar as *prioris* difusas;
- d)  $i$  refere-se a  $i$  -ésima linha em  $\mathbf{A}_t$ .

A aplicação do modelo descrito acima é feita então sobre dois diferentes tipos de retornos. O primeiro visa a análise dos índices setoriais, afim de verificar a relação temporal existente entre eles, tal como a relação temporal auto defasada, em outras palavras, dada pelo coeficiente do próprio retorno defasado. O segundo conjunto de retornos refere-se à montagem das carteiras necessárias para os fatores propostos por Fama-French como extensão do modelo CAPM. Essas carteiras foram utilizadas para a análise, como encontrado em DeMiguel, Nogales e Uppal (2014), e possuem importância relativa ao tamanho da capitalização de mercado das empresas, e do seu grau de crescimento no mercado, fornecida pelo índice *book-to-market*.

## 8.1 Índices setoriais

A primeira parte da análise consiste em aplicar o BVAR aos índices setoriais negociados na BM&FBovespa, seis foram considerados:

- a) IFNC - Índice Financeiro;
- b) INDX - Índice Industrial;

- c) IEE - Índice Energia Elétrica;
- d) ICON - Índice Consumo;
- e) IMAT - Índice Materiais Básicos;
- f) UTIL - Índice Utilidade Pública.

O período da análise foi decidido de acordo com a disponibilidade dos dados, a saber, de 2 de janeiro de 2007 até 23 de fevereiro de 2015, totalizando 1994 dias de cotação. Sendo o preço do índice em  $t$  dado por  $P_t$ , os retornos dos índices, e dos demais em toda extensão deste trabalho, foram feitos da seguinte maneira:

$$\text{Retorno}_t = y_t = \ln\left(\frac{P_{t+1}}{P_t}\right)$$

Então, deste modo foi rodado o BVAR com parâmetros variantes no tempo, com o objetivo de verificar que tipo de dependência temporal esses retornos tem apresentado. Na forma matricial, todas as relações lineares entre os ativos em  $t$  e  $t - 1$  são dadas pelo modelo:

$$\mathbf{y}_{t+1} = \mathbf{c}_t + \mathbf{B}_t \mathbf{y}_t$$

Onde  $\mathbf{y}_t$  é um vetor composto pelos retornos dos índices, cuja dimensão é  $6 \times 1$ ,  $\mathbf{c}_t$  é um vetor de constantes e a matriz de interesse  $\mathbf{B}_t$  tem dimensão  $6 \times 6$  cujos elementos  $B_{ij}$  são indexados de acordo com os índices descritos acima, para todo  $i, j = 1, \dots, 6$ .

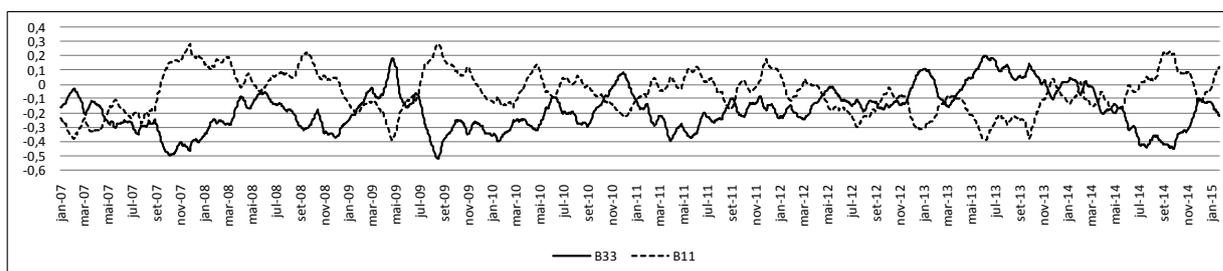
### 8.1.1 Análise dos componentes autorregressivos

Nesta seção são apresentados os principais resultados encontrados para os componentes autorregressivos, ou seja, os parâmetros referentes à relação de cada índice com seu próprio valor defasado.

No primeiro gráfico, 1, são confrontados os valores referentes aos índices IFNC e IEE. A relação encontrada entre estes dois parâmetros pode ser destacada quanto ao seu comportamento temporal, ao passo que um cai, o outro sobe, e vice-versa. A interpretação destes coeficientes, tal como os próximos que seguirão, é clara, quanto maior o valor deste, a dependência temporal aumenta e retornos positivos tenderão a ser acompanhados por retornos positivos no dia seguinte, e negativos por retornos negativos. Verifica-se então que o índice setorial tende a permanecer acima de zero na maior parte do tempo.

O índice IEE apresenta um comportamento inverso do apresentado para o setor financeiro. Este mantém-se negativo na maior parte do tempo, um indicativo de que os retornos dos ativos de empresas do setor de energia oscilam diariamente, se em  $t$  o retorno for positivo, há grandes chances deste ser negativo no dia seguinte, em  $t + 1$ .

Figura 1 – Coeficientes IFNC e IEE

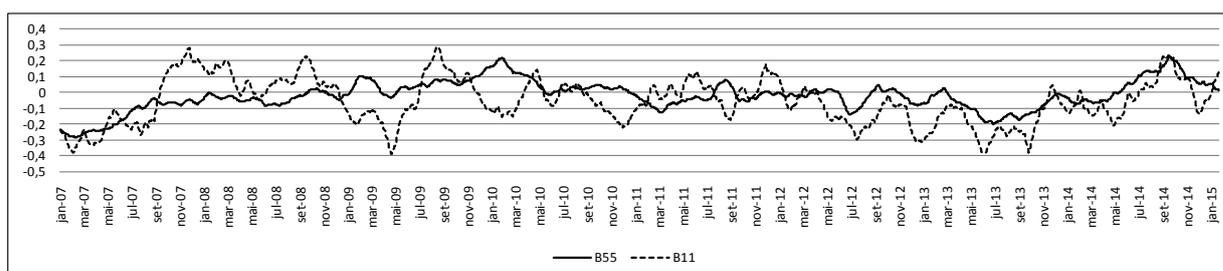


Fonte: elaborada pelo autor (2015)

Então de modo geral pode-se dizer que quando a persistência temporal de retornos de ativos financeiros aumenta, é um sinal de que os retornos do setor de energia terão um padrão de oscilação entre positivo e negativo maior.

O próximo gráfico, 2 apresenta, tal como o anterior, resultados para o índice do setor financeiro, com a diferença que está confrontado com a variação do parâmetro do componente autorregressivo do setor de materiais básicos. Esse gráfico foi exposto devido ao co-movimento de ambos os betas. Apesar de divergirem em alguns intervalos do período da análise, no geral pode-se dizer que caminham juntos, com maior estabilidade para o setor de materiais básicos, o qual apresenta o parâmetro mais próximo de zero na maior parte do tempo, significando que é difícil prever o retorno para esse tipo de mercado sempre um passo à frente.

Figura 2 – Coeficientes IFNC e IMAT

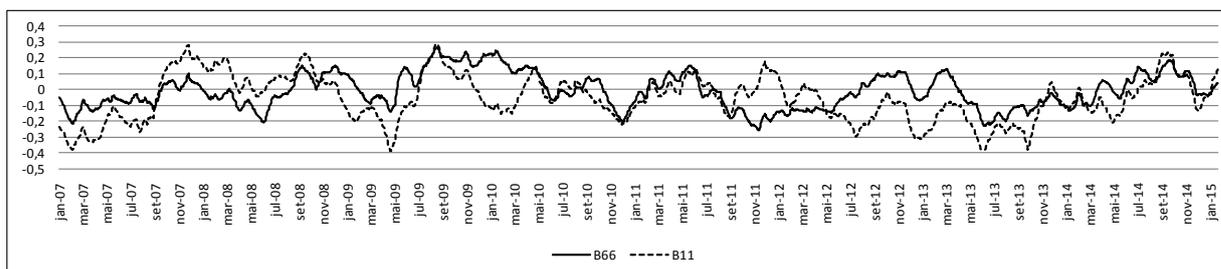


Fonte: elaborada pelo autor (2015)

O terceiro gráfico desta seção, igualmente aborda o setor financeiro, porém agora com o setor de utilidade pública, de acordo com o gráfico exposto em 3. Tal como o anterior, a trajetória dos parâmetros tendem a ser similares, discordando de si em curtos períodos de tempo. Um padrão extra para esses dois setores ocorre a partir da metade de 2010, quando o co-movimento aumenta. No geral o setor de materiais básicos tem apresentado uma trajetória temporal negativa, apesar de próxima de zero, significando que os retornos desse setor costumam oscilar diariamente quanto ao preço de fechamento.

Seguindo a análise, o próximo gráfico apresenta os resultados para o setor industrial

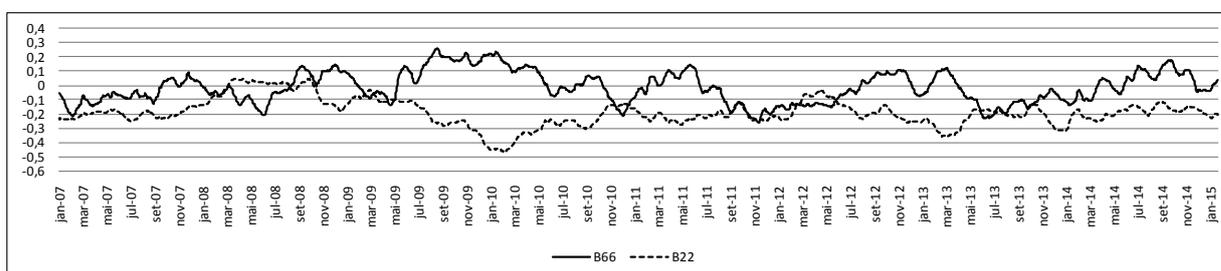
Figura 3 – Coeficientes IFNC e UTIL



Fonte: elaborada pelo autor (2015)

e de energia elétrica, novamente, representado em 4. O resultado da sobreposição destes gráficos indica um comportamento inverso, tal como o apresentado em 1, porém não tão forte. Esse resultado é mais consistente no que tange aos valores assumidos pelos coeficientes, para o caso do setor financeiro, este oscila bem em torno de zero, porém na maior parte do tempo podemos dizer que há algum nível de persistência temporal nos retornos deste. Todavia, vale ressaltar novamente que o setor de energia elétrica apresenta coeficientes negativos na maior parte do tempo.

Figura 4 – Coeficientes INDX e IEE

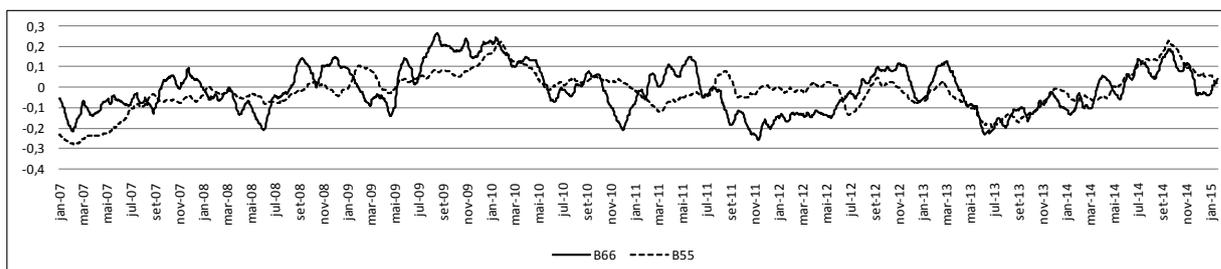


Fonte: elaborada pelo autor (2015)

A motivação da seleção do gráfico 5 ocorre do comportamento muito similar apresentado pelos coeficientes do componente autorregressivo destes dois setores a partir de maio de 2012. É um indicativo de que a persistência temporal destes nos últimos anos tem apresentado o mesmo padrão, o que na prática, olhando-se para apenas um destes dois setores, pode-se ter uma noção boa do comportamento dos retornos do outro.

Para encerrar esta seção, todos os coeficientes relacionados ao componente autorregressivo foram plotados no mesmo gráfico. Percebe-se que, levando-se em consideração os setores individualmente, a informação contida em  $t$  para fazer previsão do retorno em  $t + 1$  apenas quanto à própria defasagem não tem um padrão igual para todos os setores, percebemos que estes oscilam bastante em torno de zero, como esperado, porém há momentos em que estes assumem valores positivos e negativos relativamente grandes em módulo. Dois resultados podem, então, ser retirados desta análise, o primeiro diz respeito à oscilação dos coeficientes, são mercados muito instáveis se levarmos em con-

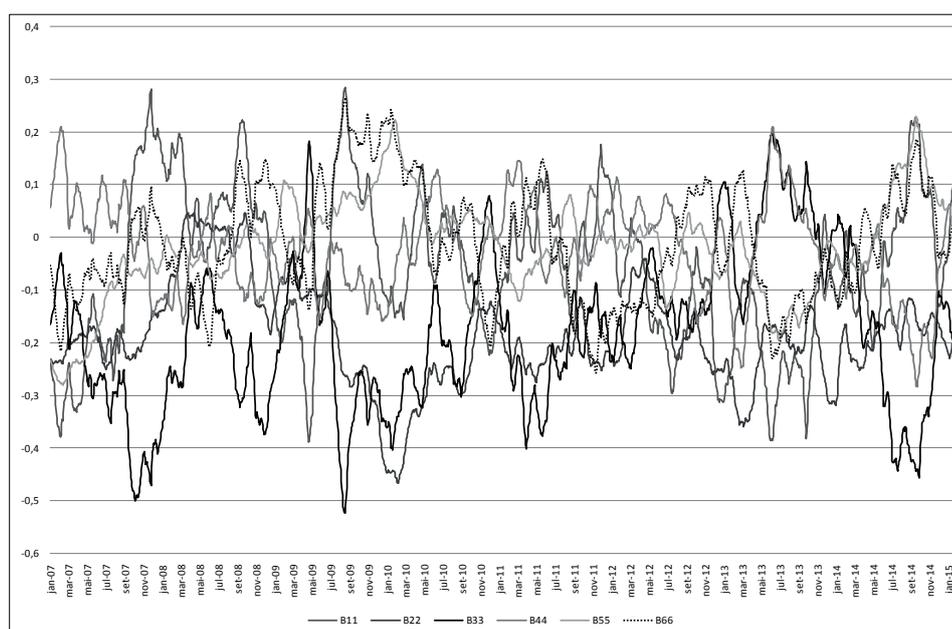
Figura 5 – Coeficientes IMAT e UTIL



Fonte: elaborada pelo autor (2015)

sideração o processo gerador dos dados, o segundo ponto de vista refere-se justamente aos períodos em que os parâmetros diferem-se do valor nulo. Com essa informação, é possível fazer previsões um passo a frente melhor quando tal acontece, com ressalvas de que nenhum permanece muito tempo grande em módulo, um sinal de que o mercado se ajusta rapidamente às informações contidas nessas séries.

Figura 6 – Coeficientes dos componentes autorregressivos

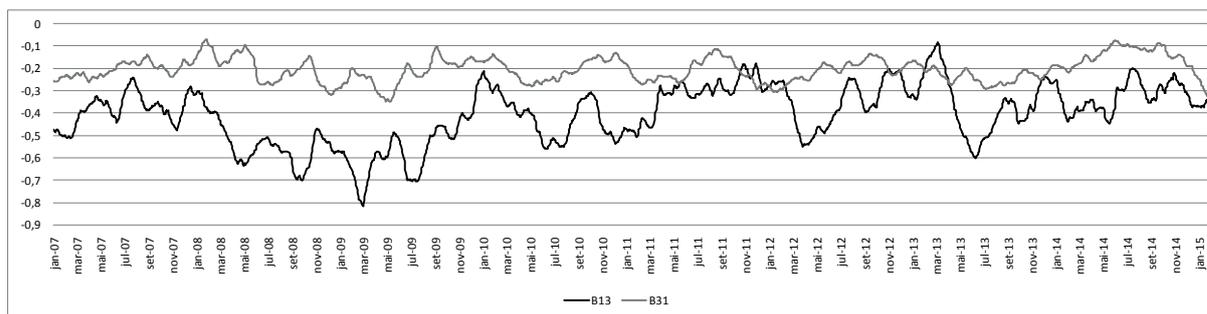


Fonte: elaborada pelo autor (2015)

### 8.1.2 Análise dos demais componentes

Alguns resultados podem ser destacados quando a análise é feita sobre os parâmetro de fora da diagonal principal. De modo geral, esses coeficientes representam o quão bem uma série pode explicar a outra. Se alto em módulo, os resultados são mais satisfatórios para fazer previsões um passo à frente.

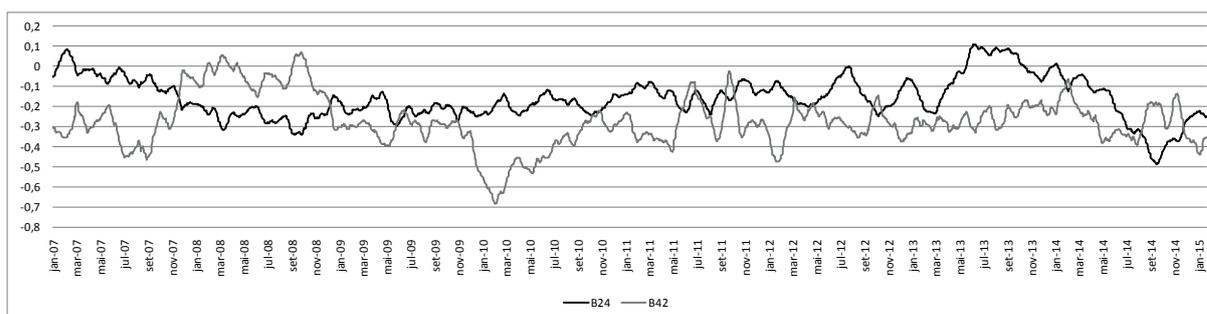
Figura 7 – Coeficientes dos índices IFNC-IEE



Fonte: elaborada pelo autor (2015)

O primeiro resultado escolhido apresenta-se no gráfico 7, cujas trajetórias temporais dos parâmetros referem-se à relação entre os índices IFNC e IEE. Pela análise gráfica é possível observar que o  $\beta_{13}$  mantém-se na maior parte do tempo mais negativo que  $\beta_{31}$ . Em termos práticos os retornos do setor de energia podem explicar melhor o comportamento dos ativos do setor financeiro, porém a relação inversa é mais fraca, devido a proximidade do parâmetro  $\beta_{31}$  de zero. Atente que, se os ativos do setor de energia são bem negativos, o que na prática há uma relação inversa entre esses dois setores muito clara, quando o setor de energia vai bem, é provável que no período seguinte os retornos do setor financeiro não mostrem um bom desempenho, ao passo que essa relação não é tão forte do setor financeiro para o setor de energia.

Figura 8 – Coeficientes dos índices INDX-ICON

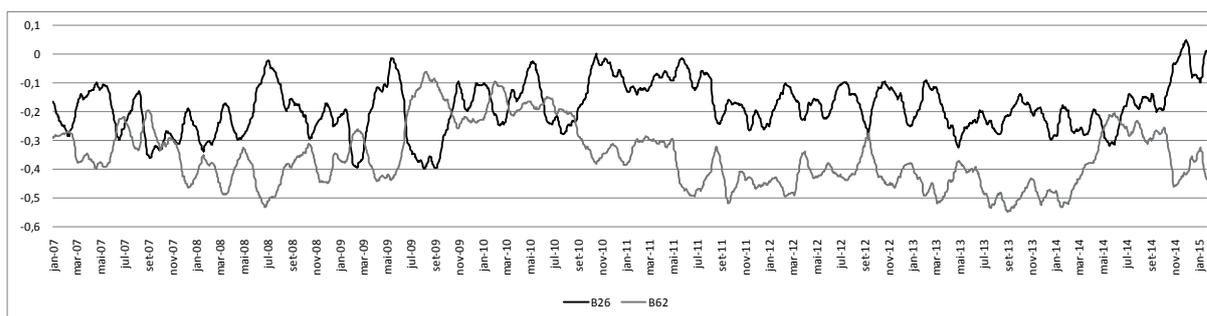


Fonte: elaborada pelo autor (2015)

Uma avaliação semelhante pode ser feita para a relação entre os setores industrial e de consumo, percebe-se que na maior parte do tempo o parâmetro  $\beta_{42}$  permanece mais negativo que  $\beta_{24}$ , indicando que os retornos do setor industrial afetam de modo mais incisivo aqueles do setor de consumo. A relação inversa permanece para este caso, com ressalvas para o ano de 2008, quando houve a Crise do *Subprime*. Os dados mais recentes da amostra tem mostrado essa inversão no padrão das séries, porém de modo não claro, o que não nos permite dizer ao certo como esses setores podem se comportar adiante.

Diferentemente do apresentado nos casos anteriores, a relação entre os setor in-

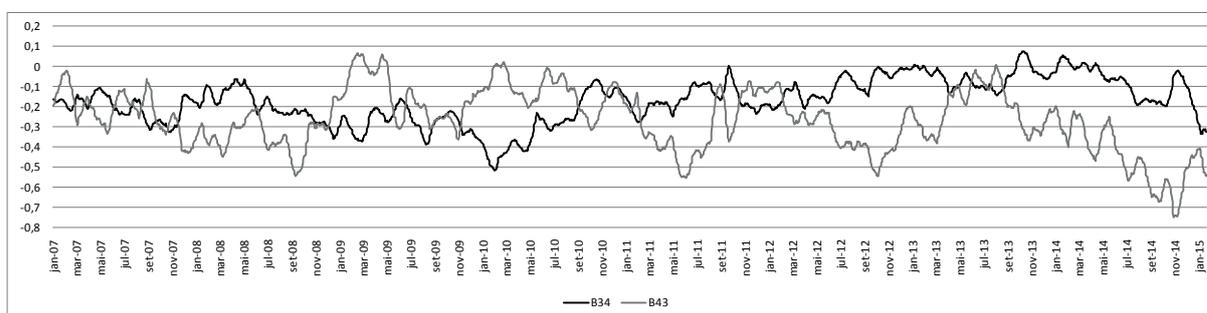
Figura 9 – Coeficientes dos índices INDX-UTIL



Fonte: elaborada pelo autor (2015)

dustrial e de utilidade pública tem mostrado uma alternância significativa ao longo do tempo. Esses dois setores foram escolhidos devido ao padrão inverso apresentado pelos  $\beta$ , quando um diminui, o outro aumenta. Apesar de não haver uma relação de liderança, na qual uma série explica melhor os resultados da outra, ambas são informativas, no que se pode observar que se o retorno do setor industrial deixa de explicar os do setor de utilidade pública, os ativos deste último passam a explicar melhor os do industrial.

Figura 10 – Coeficientes dos índices IEE-ICON



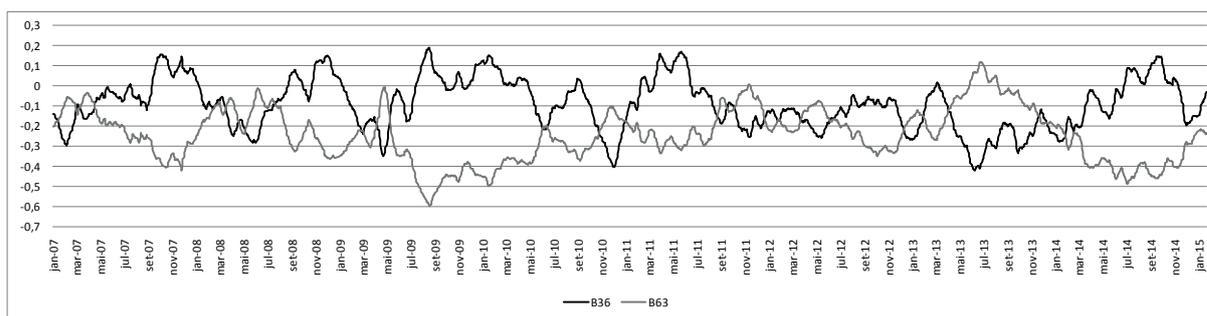
Fonte: elaborada pelo autor (2015)

O gráfico seguinte, 10, foi escolhido pela mesma causa, porém esta relação já não é tão estável para os setores de energia elétrica e de consumo. Desta análise gráfica vale ressaltar que desde o final de 2010, o parâmetro  $\beta_{43}$  tende a permanecer abaixo de  $\beta_{34}$ , de forma que o comportamento do setor de energia elétrica tem um poder explicativo maior para os papéis do setor de consumo.

O último gráfico desse tipo de análise gráfica foca nos setores de energia elétrica e de utilidade pública, dados pelo gráfico 11. A relação novamente é de inversão das trajetórias, alternando-se quanto ao poder explicativo. Ao passo que o  $\beta$  de um setor cai, o outro sobe, e vice-versa.

Para completar essa seção, foram plotados todos os coeficientes de fora da diagonal principal num único gráfico. Apesar de não ser informativo quanto ao perfil individual das séries, ele nos fornece uma visão geral da estrutura de mercado dos setores da bolsa.

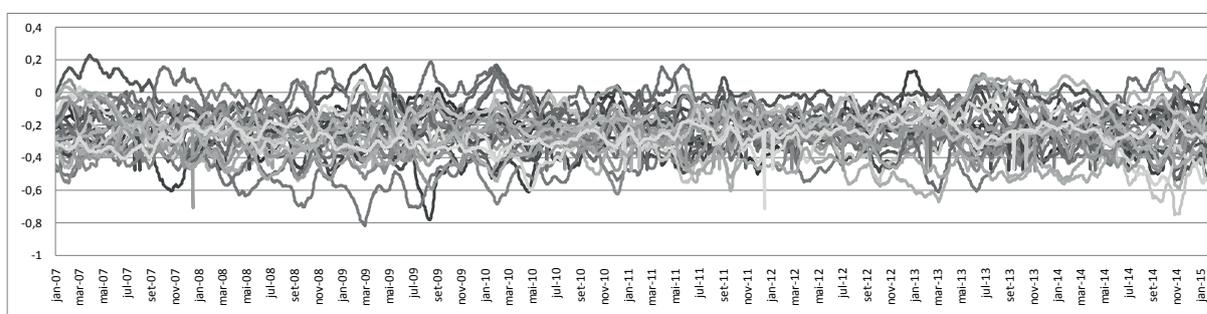
Figura 11 – Coeficientes dos índices IEE-UTIL



Fonte: elaborada pelo autor (2015)

Percebe-se que os parâmetros variam grande parte do tempo dentro do intervalo  $[0, -0, 4]$ , com raras vezes ultrapassando para a área positiva do gráfico. É um indicador que de forma geral o retorno dos setores não apresentam co-movimentos nas negociações. Ou seja, se um setor tem apresentado resultados positivos, é esperado que os demais não tenham uma boa performance. O resultado não é de se admirar, uma vez que os papéis em negociação disputam a atenção dos investidores. Se um setor vai bem, é esperado então que o capital migre para este, de forma que essa hipótese pode ser corroborada pelos betas relativamente mais positivos apresentados na seção anterior.

Figura 12 – Todos os coeficientes fora da diagonal principal



Fonte: elaborada pelo autor (2015)

## 8.2 Carteiras formadas com base nos fatores de Fama-French

A segunda parte da análise foi feita sobre as seis carteiras necessárias para a montagem dos fatores de Fama-French, com base no valor de mercado e no índice *Book-to-market*. O critério de montagem e a aplicação para o mercado brasileiro segue os registrados em Caldeira, Moura e Santos (2013), com ressalva de que foram coletados todos os ativos que entraram no índice IBrX100 no período de julho de 2001 até Fevereiro de 2015, totalizando 3375 dias de cotação. O índice IBrX100 foi escolhido de acordo com a sua representatividade ampla quanto as negociações no mercado acionário brasileiro e pelos seus critérios de liquidez. As carteiras foram montadas no final de junho de cada ano

dentro do período, sob a justificativa de que nesse período todas as informações divulgadas no começo de cada ano já foram assimiladas pelo mercado. Foram excluídas as seguintes ações:

- aquelas que não apresentaram cotações diárias para o período de um ano após a formação das carteiras;
- aquelas que não apresentaram valor de mercado ou patrimônio líquido na montagem das carteiras de acordo com a disponibilidade no Economática, banco de dados utilizado para a análise;
- empresas que não apresentaram patrimônio líquido positivo no período;
- empresas financeiras, devido a estrutura de capital diferenciada das demais, neste caso, o patrimônio líquido não tem o mesmo significado que nas demais.

### 8.2.1 Carteiras formadas pelo tamanho de mercado

Das carteiras selecionadas de acordo com os critérios expostos, a cada final de junho, foram divididas de acordo com o seu valor de mercado, um grupo contendo as maiores empresas e no outro, as menores. Então, aplicado o BVAR aos retornos, similar a DeMiguel(2014):

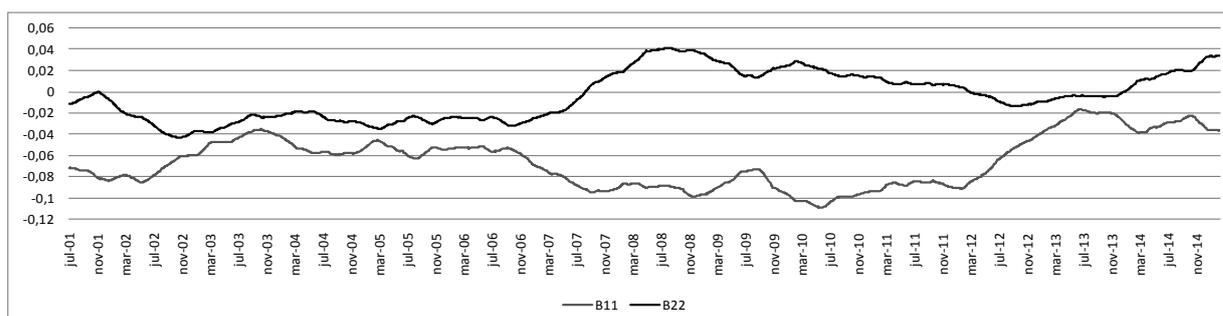
$$y_{t+1}^{grande} = c_{1,t} + \beta_{11,t}y_t^{grande} + \beta_{12,t}y_t^{pequeno}$$

$$y_{t+1}^{pequena} = c_{2,t} + \beta_{21,t}y_t^{grande} + \beta_{22,t}y_t^{pequeno}$$

### 8.2.2 Análise dos componentes autorregressivos

A interpretação dos coeficientes da diagonal principal são semelhantes do feito para os índices setoriais, porém de forma mais ampla, agora foram montadas apenas duas carteiras, o  $\beta_{11}$  referente à persistência temporal de empresas grandes, com grande volume de negociação da BM&FBovespa, e  $\beta_{22}$  para as pequenas.

Figura 13 – Coeficientes dos componentes autorregressivos



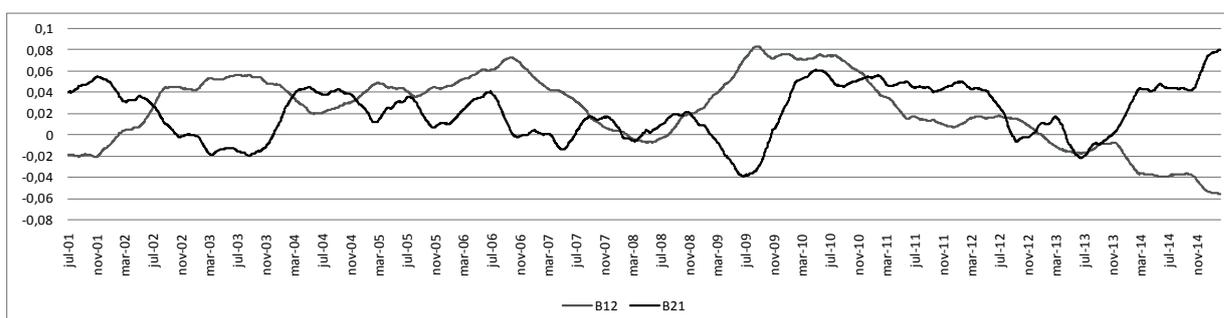
Fonte: elaborada pelo autor (2015)

O gráfico 13 apresenta a trajetória de ambas. Desta análise gráfica é possível inferir que compor carteiras com base no tamanho de mercado é menos informativa do que para os índices setoriais, vide que os parâmetros permanecem muito mais próximos de zero do que para o outro caso. A vantagem destas carteiras reside na sua relativa estabilidade, para o caso das empresas com grande capitalização, o coeficiente permanece em todo o período abaixo de zero, significando que, se um retorno desta carteira apresenta um resultado positivo hoje, há uma chance, mesmo que pequena, deste apresentar um resultado negativo amanhã. Devido a relativa estabilidade da trajetória, essa informação pode ser relevante na hora da inclusão de ativos de empresas grandes na carteira. O coeficiente relativo à empresas com baixa capitalização geralmente oscila em torno de zero, dizendo que esse conjunto de retornos não são tão informativos para obter-se previsões um passo à frente. Apesar de que a trajetória de ambos divergem em alguns intervalos da análise, não é possível descrever um padrão conjunto para ambos.

### 8.2.3 Análise dos demais componentes

Os componentes restantes são os dois fora da diagonal principal. A interpretação destes assemelha-se ao caso dos índices setoriais. Como visto no gráfico 14, tal como no caso dos componentes autorregressivos, os  $\beta_t$  não são tão informativos. Nos últimos meses há uma clara tendência de queda para  $\beta_{12}$ , representando que resultados positivos para ativos de empresas pequenas no geral tem a capacidade de gerar retornos negativos para os ativos de empresas grandes, talvez por conta de que o maior risco assumido nestas empresas proporcionam maiores retornos, atraindo o capital aplicado em empresas mais consolidadas para esse setor.

Figura 14 – Coeficientes dos componentes fora da diagonal principal



Fonte: elaborada pelo autor (2015)

O outro parâmetro,  $\beta_{21}$  encontra-se na maior parte do tempo acima de zero, representando que, se retornos de ativos de empresas grandes são positivos, de modo geral os de empresas pequenas também serão, e vice-versa. É um cenário esperado, uma vez que se empresas com maior capitalização apresentam bons resultados, o mercado de forma geral se sente bem. Quando do caso negativo, como de grandes escândalos corporativos ou má

performance de grandes empresas, vide que o mercado acionário brasileiro é muito concentrado nestes papéis, o mercado como um todo vai mal, um comportamento observado na prática.

#### 8.2.4 Carteiras formadas pelo índice book-to-market

O próximo passo da análise então foi feita com base na razão do patrimônio líquido pelo valor de mercado de todas as empresas que compuseram a análise, de acordo com os critérios expostos. Foram então divididas em três grupos aproximadamente iguais:

- a) o primeiro grupo, contendo aproximadamente 3/10 dos valores mobiliários, com alta razão *book-to-market*, chamado aqui de alto;
- b) o segundo grupo, igualmente contendo aproximadamente 3/10 dos valores mobiliários, contendo aqueles ativos que apresentaram baixa razão *book-to-market*;
- c) o terceiro grupo, contendo aproximadamente 4/10 dos valores mobiliários, composto pelos ativos que restaram, chamado aqui de neutro.

Foi então rodado o VAR de forma semelhante à exposta na seção anterior, porém com o grupo que apresentou alta razão *book-to-market* e o grupo que apresentou baixa razão:

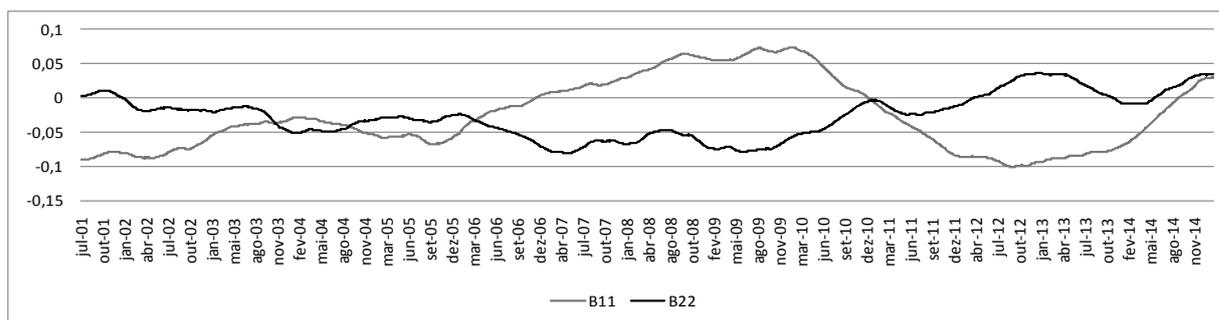
$$\begin{aligned} y_{t+1}^{alto} &= c_{1,t} + \beta_{11,t}y_t^{alto} + \beta_{12,t}y_t^{baixo} \\ y_{t+1}^{baixo} &= c_{2,t} + \beta_{21,t}y_t^{alto} + \beta_{22,t}y_t^{baixo} \end{aligned}$$

#### 8.2.5 Análise dos componentes autorregressivos

O índice *book-to-market* representa a proporção do patrimônio líquido de uma determinada empresa em termos do seu valor de mercado, dado pelo volume de negociação de seus papéis na bolsa. Se a razão é alta, então o valor contábil da empresa sobrepõe-se ao seu valor de mercado, um indício de que o preço dessa empresa está subavaliado pelos investidores, um indício de crescimento no mercado acionário para tal. De forma análoga, temos a mesma relação quando o *book-to-market*, um sinal de que o mercado sobrevalorizou o preço da empresa. Desta forma, a carteira alto representa uma média das empresas com baixa expectativa de crescimento no preço das ações, ao passo que a carteira baixo representa uma boa expectativa de crescimento.

O coeficiente referente ao componente auto defasado das empresas com baixa expectativa de crescimento é dado pelo  $\beta_{11}$ , e sua trajetória é apresentada no gráfico 15. Em boa parte do tempo o retorno dessa carteira mantém-se abaixo de zero. A análise gráfica ainda permite-nos dizer que esse componente oscila muito mais que o referente à carteira com alta expectativa de crescimento, a qual no período da análise gira aproximadamente em torno de zero. Em termos práticos, o mercado de ações com alta expectativa

Figura 15 – Coeficientes dos componentes autorregressivos



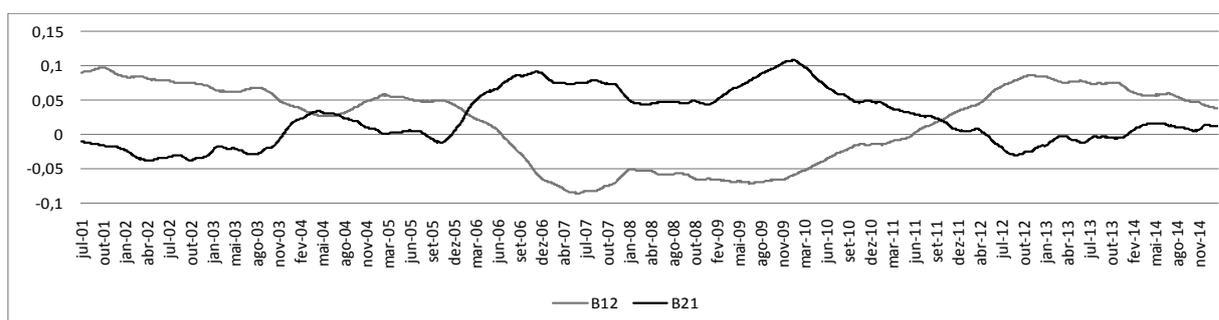
Fonte: elaborada pelo autor (2015)

de crescimento possui baixa dependência temporal, logo seu retorno previsto em  $t + 1$  dificilmente será explicado pelo desempenho mostrado em  $t$ , ao passo que não ocorre de forma tão incisiva para as empresas com baixa expectativa de crescimento.

### 8.2.6 Análise dos demais componentes

Os componentes de fora da diagonal, de acordo com o gráfico 16, em contraste com os coeficientes da auto defasagem, tem uma variabilidade temporal maior. É importante notar-se que há movimentos explícitos de queda e subida entre eles, alternando-se. Até meados de maio de 2006, a carteira composta pelos ativos com alta expectativa de crescimento, representada pelo  $\beta_{12}$ , apresentou valores mais distantes de zero, de forma que os resultados dessas empresas podem explicar melhor o retorno das empresas com baixa expectativa de crescimento, ao passo que o contrário não é tão evidente.

Figura 16 – Coeficientes dos componentes fora da diagonal principal



Fonte: elaborada pelo autor (2015)

Claramente há uma mudança de perfil após esta data e ambos divergem de zero até o final de 2011. Neste período pode-se dizer que resultados positivos em  $t$  para empresas com alta expectativa de crescimento promovem bons resultados nas que tem baixa expectativa, ao passo que os bons resultados das empresas que tem baixo crescimento esperado promovem um resultado negativo para o outro grupo. É plausível do ponto de

vista mais geral se pensarmos que, se as empresas com baixo *book-to-market* estão cumprindo suas expectativas de crescimento, o mercado como um todo se sentirá confortável. Sob o outro ponto de vista, empresas já valorizadas pelo mercado com baixo crescimento, se apresentam retornos positivos, promovem um baixo desempenho para o outro grupo.

Este cenário permanece até o final de 2011, quando a amostra indica que a primeira relação volta a vigorar. Pode-se pensar ainda que esse comportamento de inversão dos coeficientes ocorreu de forma similar para os componentes da auto defasagem. O período compreendido foi um período de grande instabilidade no mercado, compreendendo a euforia pré-crise, o movimento de baixa e a recuperação dos ativos na bolsa.

### 8.2.7 Todas as carteiras formadas pelo valor de mercado e índice book-to-market

Esta seção, ao contrário das anteriores, busca resultados mais gerais. Para tal, foram consideradas todas as combinações de carteiras possíveis com base no valor de mercado e no índice *book-to-market*, a saber:

- a) grande-alto;
- b) grande-neutro;
- c) grande-baixo;
- d) pequeno-alto;
- e) pequeno-neutro;
- f) pequeno-baixo;

O período da análise feito encontra-se no intervalo de 2 de junho de 2001, da montagem da primeira carteira, até 23 de fevereiro de 2015, totalizando 3375 dias de negociação na bolsa.

Tal como feito nas demais análises, então, deste modo foi rodado o BVAR com parâmetros variantes no tempo, com o objetivo de verificar que tipo de dependência temporal esses retornos tem apresentado. Na forma matricial, todas as relações lineares entre os ativos em  $t$  e  $t - 1$  são dadas pelo modelo:

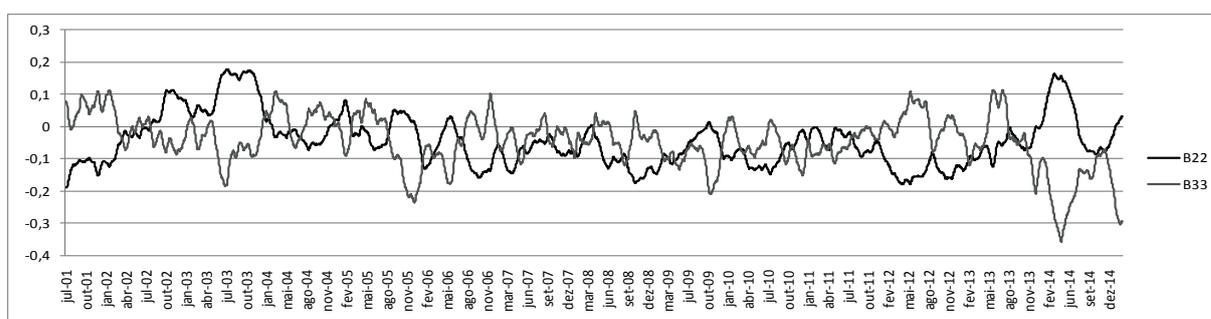
$$\mathbf{y}_{t+1} = \mathbf{c}_t + \mathbf{B}_t \mathbf{y}_t$$

Onde, similarmente ao caso quando dos índices setoriais,  $\mathbf{y}_t$  é um vetor composto pelos retornos dos índices, cuja dimensão é  $6 \times 1$ ,  $\mathbf{c}_t$  é um vetor de constantes e a matriz de interesse  $\mathbf{B}_t$  tem dimensão  $6 \times 6$  cujos elementos  $B_{ij}$  são indexados de acordo com os índices descritos acima, para todo  $i, j = 1, \dots, 6$ .

## 8.2.8 Análise dos componentes autorregressivos

A primeira parte da análise, tal como feito anteriormente, objetiva conferir padrões existentes para as séries dos termos auto defasados. A interpretação que segue é a mesma feita para os casos anteriores.

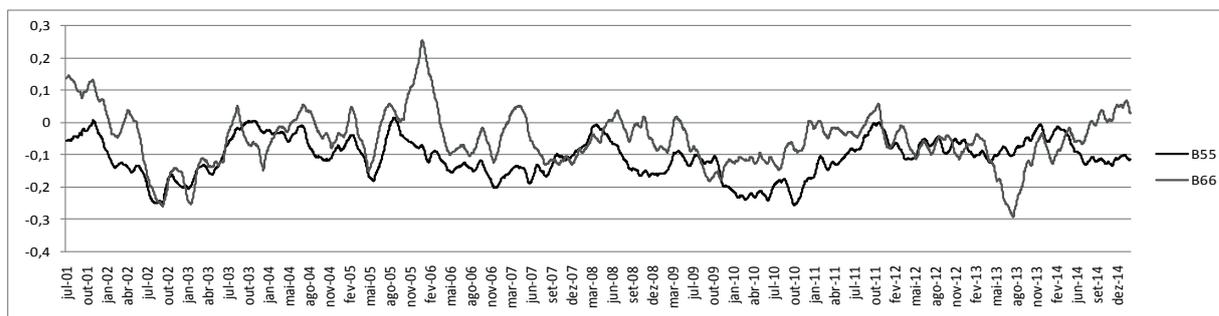
Figura 17 – Carteiras Grande-neutro e Grande-baixo



Fonte: elaborada pelo autor (2015)

O primeiro gráfico selecionado da pesquisa, 17, sobrepõe as trajetórias das carteiras com grande valor de mercado. O  $\beta_{22}$  e o  $\beta_{33}$  correspondem aos índices *book-to-market* neutro e baixo respectivamente. A trajetória desses dois grupos de ativos indicam uma relação inversa, ao passo que um sobe, outro cai. Percebe-se ainda que ambas estão suavemente abaixo de zero, um sinal de que o retorno de ambas oscila entre positivo e negativo, porém o padrão não é tão forte, sobretudo se contarmos que nos últimos meses da análise esse movimento tem aumentado consideravelmente, colocando os parâmetros na área positiva do gráfico em intervalos de tempo de forma distinta.

Figura 18 – Carteiras Pequeno-neutro e Pequeno-baixo

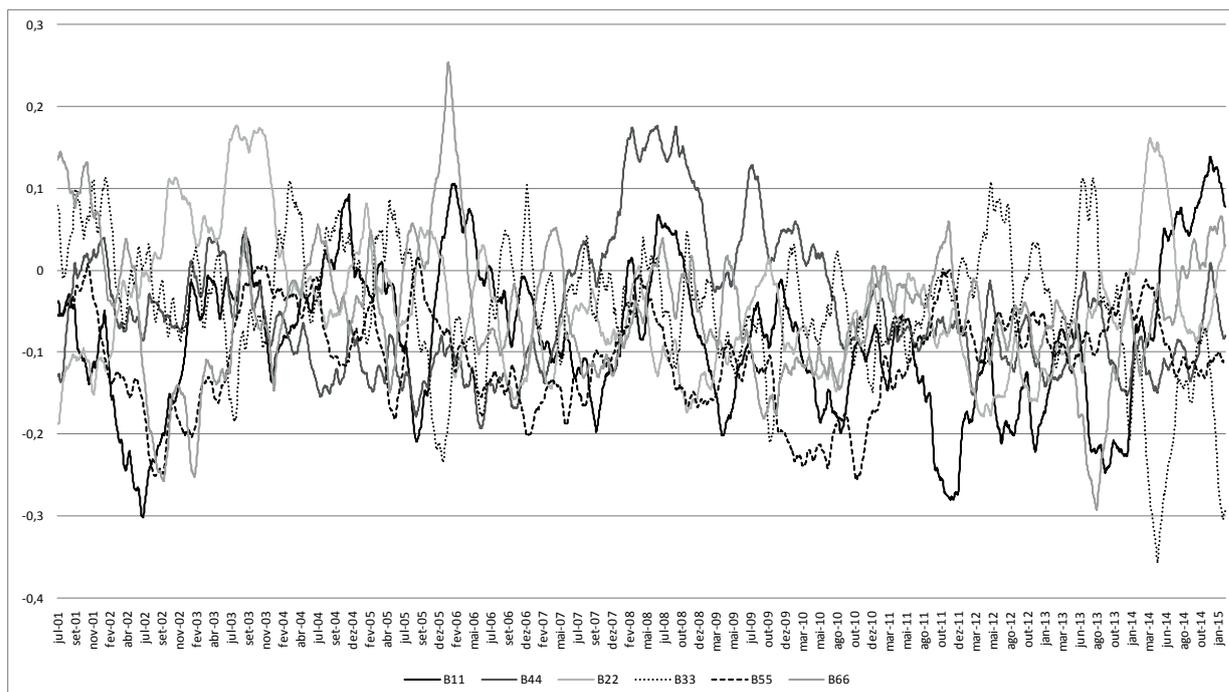


Fonte: elaborada pelo autor (2015)

O segundo resultado escolhido refere-se às carteiras com baixo valor de mercado, igualmente ao caso anterior, os resultados encontrados foram achados para empresas com razão *book-to-market* neutra e baixa, representadas pelos parâmetros  $\beta_{55}$  e  $\beta_{66}$  respectivamente. Diferentemente do caso anterior, verifica-se que esses parâmetros tendem a caminhar juntos na maior parte do tempo, divergindo de meados de maio de 2013 para

dados mais recentes da amostra. Esse resultado é importante quando contrastado com aquele encontrado quando do VAR sobre as empresas com alta e baixa razão *book-to-market*. Uma vez que as empresas agora foram divididas pelo valor de mercado, percebe-se que o perfil temporal da dependência muda, logo tem-se um indício que é importante levar em consideração informações adicionais, tal como o tamanho da empresa, quando montar uma carteira pensando no seu crescimento, no que tange ao índice *book-to-market*.

Figura 19 – Todos os parâmetros dos componentes autorregressivos das carteiras



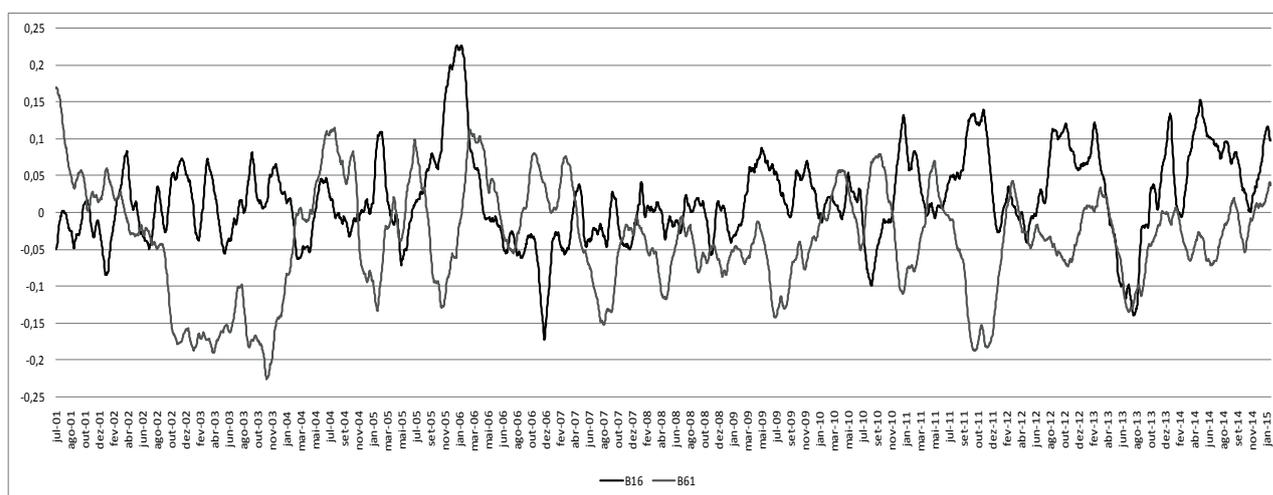
Fonte: elaborada pelo autor (2015)

O último gráfico desta seção sobrepõe todos os  $\beta$  responsáveis por captar a persistência temporal destas carteiras. É um gráfico informativo da estrutura geral do mercado quanto a esse tipo de segmentação, valor de mercado e *book-to-market*. Percebe-se que o comportamento destas carteiras oscilam em torno de zero, com uma leve tendência para a área negativa do gráfico, indicando que os retornos tendem a oscilar diariamente, com baixa persistência temporal. Um fato interessante diz a respeito das carteiras montadas puramente pelo valor de mercado e *book-to-market*, que no caso, apresentam parâmetros mais estáveis. Essa observação pode ser enxergada sob duas óticas, a primeira ocorre que as carteiras tem menos ativos neste caso que para o caso anterior, logo a outra possui maior estabilidade devido a sua diversificação, ao passo que nesta seção temos uma visão mais segmentada do mercado, a segunda ótica decorre da amplitude dos valores assumidos pelos parâmetros, para este caso, em curtos períodos de tempo, os  $\beta$  assumem valores relativamente extremos quanto ao outro caso. Sob certos cuidados um investidor atento pode obter oportunamente retornos maiores se souber utilizar corretamente esta informação, conjuntamente com estratégias adequadas de investimento.

### 8.2.9 Análise dos demais componentes

A última parte deste conjunto de análises gráficas compreende a relação temporal intra carteiras, ou seja, como o comportamento de ativos quando ao seu perfil de tamanho de mercado e crescimento (*book-to-market*) influenciam uns aos outros.

Figura 20 – Carteiras Grande-alto/Pequeno-baixo

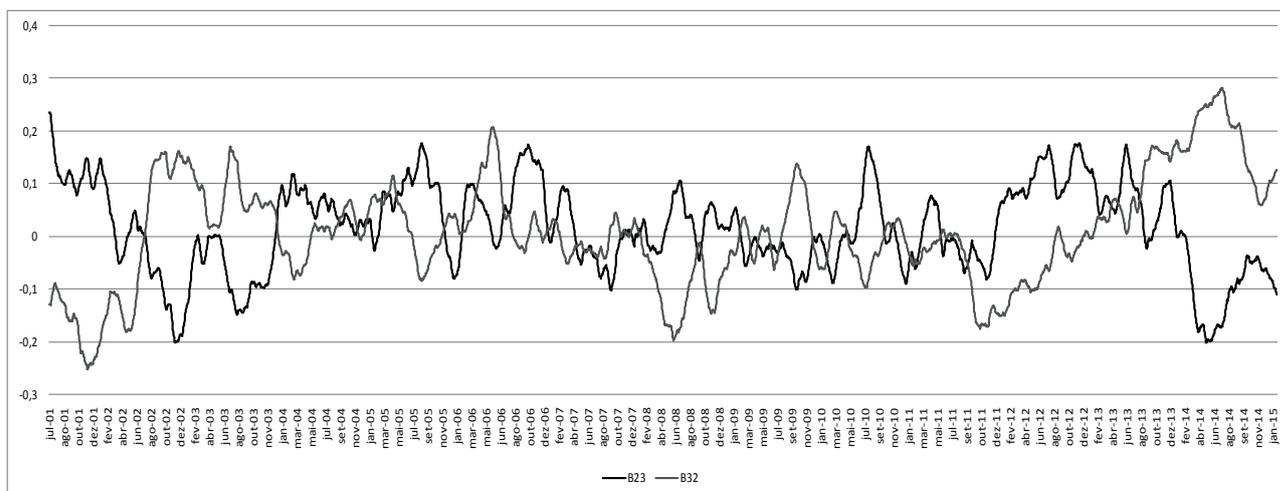


Fonte: elaborada pelo autor (2015)

O primeiro gráfico selecionado, 20, diz exatamente essa relação para carteiras extremas, onde o parâmetro  $\beta_{16}$  indica a influência de empresas pequenas com baixo *book-to-market* sobre empresas grandes com alto *book-to-market*. Ao longo dos anos elas tem oscilado consideravelmente em torno de zero, porém o poder preditivo de ambas é recíproco, uma vez que de modo geral, quando uma fica negativa, a outra fica positiva. A quebra desse padrão apresentou-se apenas no final da amostra. Apesar do fenômeno observado, é difícil para um investidor aplicar seus recursos levando-se em consideração a relação entre essas duas carteiras, uma vez que, quando uma tem retornos positivos com o  $\beta$  na região positiva do gráfico, leva-nos a crer que a outra carteira também apresentará um retorno positivo em  $t + 1$ , porém o  $\beta$  da outra carteira encontra-se na área negativa do gráfico, significando que há uma boa chance da primeira carteira de ativos apresentar um retorno negativo no período seguinte, o que acarreta num retorno negativo para a segunda. De forma geral, há bastante instabilidade quando tal acontece.

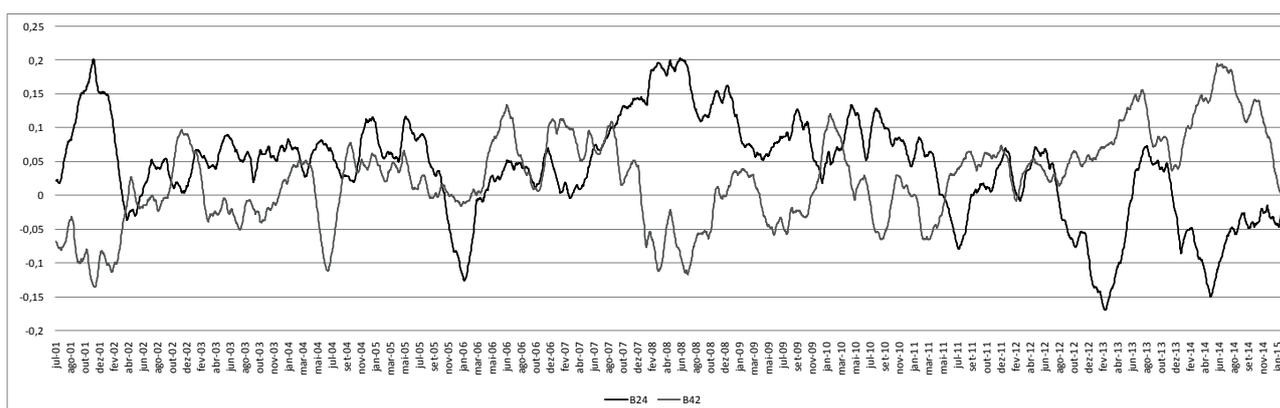
A mesma análise feita anteriormente pode ser feita para os próximos gráficos selecionados, com a ressalva de que o padrão não é tão forte como para a análise anterior. Em 21, são plotados os  $\beta$  de empresas com grande valor de mercado, porém *book-to-market* neutro e baixo. Em 22, para as carteiras uma com grande valor de mercado e *book-to-market* neutro, e a outra pequeno e alto respectivamente. A diferença sutil deve ser notada quanto à escala dos parâmetros, dado que a primeira apresenta uma oscilação maior que a segunda.

Figura 21 – Carteiras Grande-neutro/Grande-baixo



Fonte: elaborada pelo autor (2015)

Figura 22 – Carteiras Grande-neutro/Pequeno-alto

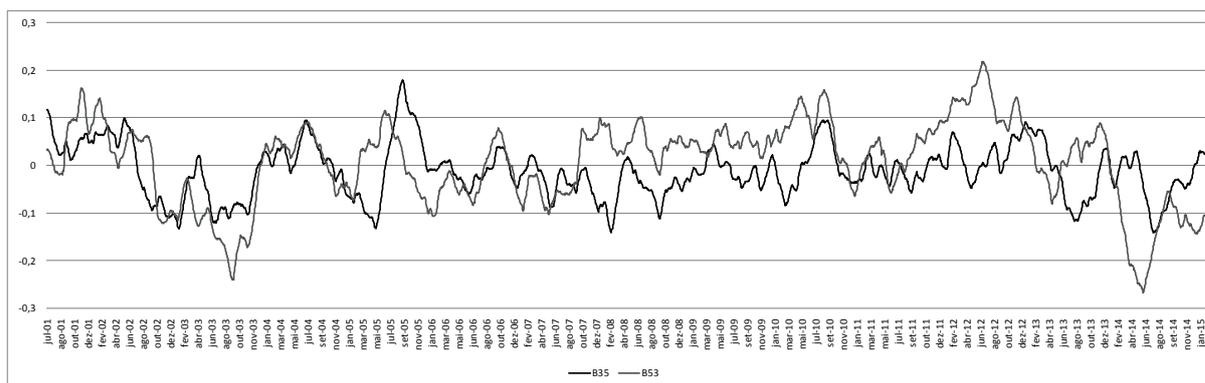


Fonte: elaborada pelo autor (2015)

O penúltimo gráfico, 23, é um dos únicos que apresenta esse tipo de formato, uma série maior do que a outra na maior parte do tempo. Percebe-se que a trajetória de ambas é muito similar, porém de outubro de 2007 até meado de fevereiro de 2013,  $\beta_{35}$ , responsável pela relação da carteira de empresas pequenas com *book-to-market* neutro para com a carteira de empresas grandes e baixo índice, permanece ligeiramente inferior, quando tal, os retornos desta pode explicar melhor, de modo inverso, o retorno da segunda carteira. Esse padrão não se manteve ao final da amostra.

De modo geral, a relação entre essas carteiras é bem instável quando analisadas essas relações, o que não as torna muito informativas. Para verificar isto, foram selecionadas as carteiras que na média do tempo apresentaram os parâmetros mais alto e baixo, curiosamente dados pelos coeficientes  $\beta_{35}$  e  $\beta_{53}$  em 24, relativos à empresas com alto valor de mercado e baixo *book-to-market* com baixo valor de mercado e *book-to-market* neutro. Vide que ao passo que um parâmetro permanece boa parte do tempo na área positiva,

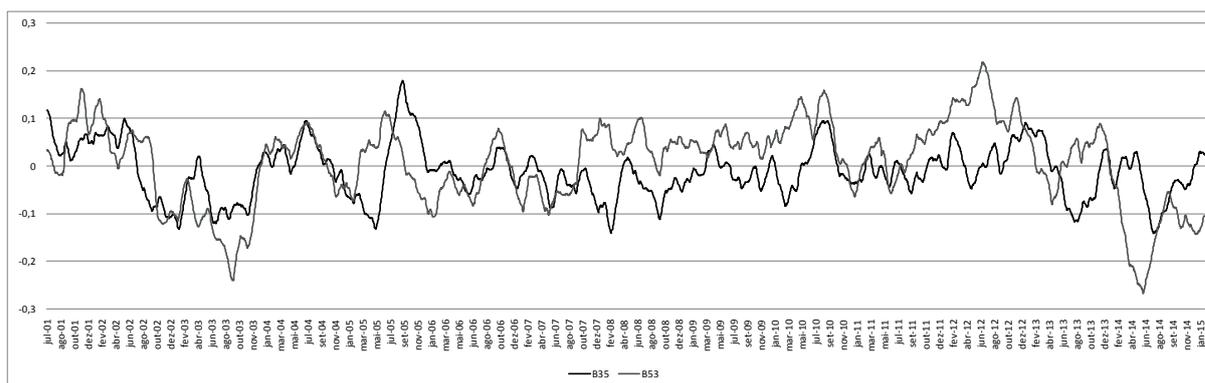
Figura 23 – Carteiras Grande-pequeno/Pequeno-neutro



Fonte: elaborada pelo autor (2015)

outro permanece na negativa. Essas carteiras ainda apresentam uns dos valores mais extremos para todos os parâmetros avaliados fora da diagonal principal.

Figura 24 – Coeficientes extremos



Fonte: elaborada pelo autor (2015)

## 9 CONSTRUÇÃO DAS CARTEIRAS E MEDIDAS DE DESEMPENHO

Em DeMiguel, Nogales e Uppal (2014), encontramos motivação suficiente para considerar que há persistência temporal nas séries de retornos dos ativos, não só no segundo momento, a variância, como no primeiro, a média. Essa dependência, mesmo que pequena, pode ser incorporada na construção de diferentes carteiras, o que pode gerar resultados mais desejáveis para o investidor.

Esse capítulo presta-se a verificar aplicabilidade dessa dependência seguindo uma metodologia de construção para as carteiras similar à proposta encontrada em DeMiguel, Nogales e Uppal (2014). A abordagem Bayesiana novamente aparece, agora sobre outra forma, seguindo a abordagem apresentada em Koop e Korobilis (2013), cujos autores empenham-se em elaborar métodos cujos custos computacionais reduzidos permitem calcular sistemas tipo TVP-VAR heteroscedástico para grandes dimensões.

Para a construção de uma carteira hipotética a fim de que sejam testados os modelos e estratégias de investimento, o critério escolhido de seleção para os ativos baseou-se na busca de todas as ações que entraram pelo menos uma vez no índice IBrX100 negociado na BM&FBovespa. Esse índice foi escolhido pela sua amplitude de mercado, pois os ativos que o compõem advêm de empresas de diversos setores e perfis, de modo que se tenha um investimento bem diversificado, o que é buscado na prática. Foram excluídos todos os ativos que não apresentaram cotação para todo o período entre 2 de janeiro de 2007 a 18 de fevereiro de 2015, tal como foram excluídos aqueles que não apresentaram liquidez necessária para a estimação dos parâmetros necessários, o critério de seleção então foi baseado na contagem dos dias sem cotação para cada ativo, de modo que foram excluídos aqueles que apresentaram mais de 100 dias sem cotação neste período, cerca de 1/5 do total da amostra. Então, restaram  $n = 44$  ativos em  $t = 2010$  dias de negociações.

Os *softwares* utilizados para todos os cálculos foram o Excel e o Matlab e a coleta de dados das séries históricas dos preços das ações foi feita através do Economica, já devidamente ajustadas por proventos.

### 9.1 Metodologia

Para rodar modelos VAR de grande dimensão, sobretudo sob a ótica Bayesiana, é necessário um método de calcular os parâmetros de forma mais parcimoniosa, porém mantendo as hipóteses de variabilidade temporal dos parâmetros e heteroscedasticidade, necessária à montagem das carteiras, sobretudo as do tipo média-variância.

No trabalho de Koop e Korobilis (2013), os autores justificam que estender os tradicionais métodos utilizados para esses modelos no sentido para aqueles com grande dimensão não é direto, devido a questões computacionais. Então eles propõem uma

simplificação do método Bayesiano de TVP-VAR heteroscedástico com a finalidade de abrir caminho para modelos com maior número de variáveis. Esse método, inicialmente proposto para modelagem de dados macroeconômicos, insere-se no contexto de finanças deste trabalho de forma ótima, contendo os pressupostos necessários para a estimação de carteiras com um grande número de ativos.

A primeira suposição segue que os parâmetros variam no tempo de acordo com um passeio aleatório, propício ao caso quando queremos que os coeficientes tenham a liberdade de assumir quaisquer valores no espaço paramétrico:

$$\mathbf{y}_t = \mathbf{B}_t \mathbf{y}_{t-1} + \mathbf{e}_t \quad (9.1)$$

$$\text{vec}(\mathbf{B})_{t+1} = \text{vec}(\mathbf{B})_t + \mathbf{u}_t \quad (9.2)$$

Onde  $\mathbf{e}_t \sim i.i.d.N(\mathbf{0}, \boldsymbol{\Sigma}_t)$  e  $\mathbf{u}_t \sim i.i.d.N(\mathbf{0}, \mathbf{Q}_t)$  são termos de erro,  $\mathbf{y}'_t$  é um vetor em  $\mathbb{R}^n$ .

A distribuição *priori* para  $\mathbf{B}_0$  escolhida foi do tipo Minnesota, a ser abordada mais detalhadamente a frente. Para permitir que os parâmetros variem no tempo de acordo com o processo dado, é usado o filtro de Kalman, o qual assume que:

$$\mathbf{B}_{t-1} | \tilde{\mathbf{y}}_{t-1} \sim N(\mathbf{B}_{t-1|t-1}, \mathbf{V}_{t-1|t-1})$$

Então o filtro promove a predição de acordo com:

$$\mathbf{B}_t | \tilde{\mathbf{y}}_{t-1} \sim N(\mathbf{B}_{t|t-1}, \mathbf{V}_{t|t-1})$$

Onde

$$\mathbf{V}_{t|t-1} = \mathbf{V}_{t-1|t-1} + \mathbf{Q}_t$$

Os autores, então, promovem a substituição deste processo visando sua simplificação, a qual ocorre quanto ao uso de um fator de desconto  $\lambda$ :

$$\mathbf{V}_{t|t-1} = \frac{1}{\lambda} \mathbf{V}_{t-1|t-1}$$

O qual substitui a necessidade de estimação da matriz  $\mathbf{Q}_t$ . Esse valor é restrito no intervalo  $0 < \lambda < 1$ . Os valores para  $\lambda$  pesquisados nesta dissertação variam entre  $[0,97 \ 0,98 \ 0,99]$ . A metodologia ainda permite que o fator de desconto varie temporalmente de acordo com:

$$\lambda_t = \lambda + (1 - \lambda)(1, 1)^{e_t' e_t}$$

Onde  $e_t$  é o erro de predição do modelo.

Para o caso quando  $\lambda = 0,99$ , a variação dos parâmetros é bem pequena, o que permite estimações dos coeficientes em  $\mathbf{B}_t$  mais suavizadas, interessante para carteiras com menor variação nos rebalanceamentos, uma vez que diminui os custos de transação.

Para o conjunto de valores testados para o modelo, neste trabalho serão apresentados apenas os resultados referentes à  $\lambda = 0,97$ , uma vez que as medidas finais mostraram uma robustez moderada quanto à escolha deste parâmetro neste intervalo.

O uso de modelos de volatilidade estocástica, típico de modelos Bayesianos, cujos custos computacionais são elevados, é evitado com o uso de fatores de desconto. Para tal abordagem, é escolhido o uso do método *Exponentially Weighted Moving Average* (EWMA) de modelagem da variância:

$$\hat{\Sigma}_t = \kappa \hat{\Sigma}_{t-1} + (1 - \kappa) \hat{\mathbf{e}}_t \hat{\mathbf{e}}_t'$$

Onde  $\hat{\mathbf{e}}_t = \mathbf{y}_t - \mathbf{y}_{t-1} \mathbf{B}_{t|t}$  é o termo de erro aleatório produzido pelo filtro de Kalman,  $\kappa$  é um fator de decaimento. O intervalo de valores mais amplamente aceito segue a metodologia RiskMetrics em Morgan, a qual sugere que  $0,94 < \kappa < 0,98$ . Neste trabalho assumimos  $\kappa = 0,94$ , próprio para dados de frequência diária. A matriz inicial do modelo,  $\Sigma_0$ , foi calculada através de *training-sample* de 123 dias anteriores à vigência da carteira.

O trabalho de Koop e Korobilis (2013) incorpora ainda o uso do *Dinamic Model Average* (DMA) a fim de calcular qual a probabilidade do modelo  $j$  em prever o vetor  $\mathbf{y}_t$ , dada a informação disponível em  $\tilde{\mathbf{y}}_{t-1}$ , probabilidade esta, chamada  $\pi_{t|t-1,j}$  para todo  $j = 1, \dots, J$ . Raftery, Kárný e Ettlér (2010) deriva a probabilidade de que o modelo  $j$  deve ser usado para prever o  $\mathbf{y}_t$  de forma análoga a como o filtro de Kalman utiliza seu método de predição e atualização:

$$\pi_{t|t-1,j} = \frac{\pi_{t-1|t-1}^\alpha}{\sum_{l=1}^J \pi_{t-1|t-1,l}^\alpha}$$

Onde  $\pi_{t|t-1,j}$  é a probabilidade do uso do modelo  $j$  com a informação disponível até  $\tilde{\mathbf{y}}_{t-1}$  e  $\alpha$  é um fator de desconto. Com a chegada da nova informação  $\mathbf{y}_t$ , a equação é atualizada como:

$$\pi_{t|t,j} = \frac{\pi_{t|t-1} p_j(\mathbf{y}_t | \tilde{\mathbf{y}}_{t-1})}{\sum_{l=1}^J \pi_{t|t-1,l} p_l(\mathbf{y}_t | \tilde{\mathbf{y}}_{t-1})}$$

Onde  $p_j(\mathbf{y}_t|\tilde{\mathbf{y}}_{t-1})$  é a f.d.p. da verossimilhança para a predição, em outras palavras, a função de densidade para o modelo  $j$  avaliado em  $\mathbf{y}_t$ , a qual pode ser considerada uma medida do desempenho da previsão.

Deste modo ainda é evitado o uso de métodos MCMC, sendo necessário apenas o cômputo de  $\pi_{t|t-1,j}$  e  $\pi_{t|t,j}$ . Essa metodologia permite que sejam escolhidas diferentes *prioris* para o modelo. Neste trabalho, tal como Koop e Korobilis (2013), é utilizada uma combinação da *priori* de Minnesota para VAR estacionário, onde  $\mathbf{B}_0$  tem uma distribuição normal multivariada com média zero. A variância dos termos de intercepto é posta como 100, no intuito de que essa distribuição seja não informativa. Ademais, os parâmetros da variância das variáveis de estados de  $\mathbf{y}_{t-1}$  são calculadas nos valores  $\gamma = [1^{-10} \quad 1^{-5} \quad 0.001 \quad 0.005 \quad 0.01 \quad 0.05 \quad 0.1]$ :

$$V_i = \begin{cases} \frac{\gamma}{p^2} & \text{para } p = 1 \\ \underline{a} = 100 & \text{para os interceptos} \end{cases}$$

Por fim, para dar robustez aos resultados, são calculados também diferentes valores para  $\alpha$ , a saber  $[0,001 \quad 0,95 \quad 0,99 \quad 1]$ , os quais indicam uma maior persistência temporal, do menor ao maior valor, no que tange aos modelos quanto à previsão dos resultados em  $t$ .

Por fim, é necessário que haja a escolha da *priori* para  $\pi_{0|0}$ . Seguindo os autores, o modelo parte de uma distribuição equilibrada, onde intuitivamente o pesquisador não tenha palpite algum sobre qual parâmetro se ajusta melhor aos dados, o que equivalentemente é dado um peso igual para todos os modelos dos diferentes parâmetros para a *priori*.

### 9.1.1 Medidas de desempenho

Para todas as estratégias apresentadas neste trabalho, foram calculadas as seguintes medidas de desempenho para as carteiras:

$$\hat{\mu} = (T - 1)^{-1} \sum_{t=1}^{T-1} (\mathbf{w}'_t \mathbf{y}_{t+1}) \quad (9.3)$$

$$\hat{\sigma}^2 = (T - 1)^{-1} \sum_{t=1}^{T-1} (\mathbf{w}'_t \mathbf{y}_{t+1} - \hat{\mu})^2 \quad (9.4)$$

$$IS = \frac{\hat{\mu}}{\hat{\sigma}} \quad (9.5)$$

$$turnover = (T - 1)^{-1} \sum_{t=1}^{T-1} \sum_{i=1}^n |w_{i,t+1} - w_{i,t}| \quad (9.6)$$

Onde  $T$  é o período amostral completo e  $n$  é o número de ativos da carteira. A medida  $\hat{\mu}$  representa a média amostral da carteira, sendo dada pela média aritmética dos retornos diários. A segunda medida de desempenho é a variância da carteira, dado por  $\hat{\sigma}^2$ , é igualmente calculada levando-se em consideração todo o período amostral. Considerando-se a variância como uma medida de risco, valores menores assumidos por esse estimador são mais desejáveis. A terceira medida de desempenho leva em consideração ambas as primeiras medidas, o  $IS$ , conhecido como índice de Sharpe, pondera o retorno da carteira pelo risco assumido, logo valores mais altos são mais desejáveis. Por último, temos o *turnover*, o qual está intimamente ligado aos custos de transação, os quais são proporcionalmente maiores de acordo com essa medida de desempenho. Ele é interpretado como a porcentagem média diária da carteira que é reponderada, em outros termos, se rebalanceada diariamente, qual será a porcentagem média do total da carteira que a ser negociada, assim então, carteiras mais estáveis apresentam menor *turnover* em relação às mais voláteis.

## 9.2 Carteiras de arbitragem

Esta seção do trabalho segue a proposta em DeMiguel, Nogales e Uppal (2014), a qual busca avaliar empiricamente as carteiras do tipo arbitragem, ou de zero custos, para o caso brasileiro. Da montagem das carteiras, usou-se da abordagem Bayesiana para a estimação dos parâmetros do VAR, tal como apresentado na metodologia. Ao passo que temos ganhos em torno de parâmetros suavizados no tempo, o que pode reduzir o *turnover*, a análise da origem da predição dos retornos, como apresentada no primeiro capítulo quando da decomposição das carteiras, já não pode ser feita diretamente para o caso Bayesiano, uma vez que elas baseiam-se na inferência clássica, todavia o total dos retornos obtidos fora da amostra são comparáveis do mesmo modo.

Os custos de transação não devem ser desconsiderados, porém diferentemente da proposta original dos autores, este trabalho empenha-se em mostrar os resultados para o *turnover*, visando não agrupar os resultados em distintos cenários de custos, mas deixar a critério do leitor, aumentando seu arbítrio.

A primeira carteira de arbitragem considerada segue a proposta em Lo e MacKinlay (1990), também chamada de contrária, cujos pesos são calculados da seguinte forma:

$$\mathbf{w}_{t+1}^c = -\frac{1}{n}(\mathbf{y}_t - \bar{y}_t \mathbf{1}) \quad (9.7)$$

$$\bar{y}_t = \frac{\mathbf{1}'\mathbf{y}_t}{n} \quad (9.8)$$

Onde  $\mathbf{1}$  é um vetor  $n$ -dimensional de 1 e  $\bar{y}_t$  é a média de todos os retornos individuais em  $t$ . Os retornos deste portfólio e todos do tipo arbitragem considerados somam zero,

ou em termos práticos, o volume de ativos que se entra na forma vendida, também o é na forma comprada. A interpretação dos pesos dessa estratégia é clara, se um determinado ativo apresenta retorno positivo em  $t$ , então a carteira de arbitragem contrária garantirá uma posição vendida, ou negativa para este em  $t + 1$ .

A segunda estratégia considerada segue a proposta de Pan, Liano e Huang (2004), também chamada de momento, obtida similarmente à carteira de arbitragem contrária, porém com o sinal oposto:

$$\mathbf{w}_{t+1}^m = \frac{1}{n}(\mathbf{y}_t - \bar{y}_t \mathbf{1}) \quad (9.9)$$

$$\bar{y}_t = \frac{\mathbf{1}' \mathbf{y}_t}{n} \quad (9.10)$$

A diferença crucial dessas estratégias pode ser explicada pela relação entre os retornos dos ativos. Para a carteira contrária, o retorno esperado será positivo se a auto correlação for negativa, uma vez que, se um determinado ativo  $i$  apresenta bom desempenho em  $t$ , é esperado uma reversão nesse desempenho em  $t + 1$ . Por outro lado, se um outro determinado ativo, chamamos de  $j$ , apresentar retorno positivo em  $t$ , e ele for positivamente correlacionado com  $i$ , então essa estratégia também proporcionará retornos positivos.

Uma vez que há a possibilidade de não haver o cenário descrito, a carteira momento trabalha de forma contrária, esta considera que há alguma persistência temporal nos retornos de  $t$  para  $t + 1$  quando assume um sinal contrário à carteira de arbitragem contrária.

A terceira carteira, então, leva em consideração todas as informações disponíveis da dependência temporal entre os retornos de  $t$  para  $t+1$ , chamada por DeMiguel, Nogales e Uppal (2014) de carteira de arbitragem com VAR, cujos pesos são dados pela seguinte expressão:

$$\mathbf{w}_{t+1}^{VAR} = \frac{1}{n}(\hat{\mathbf{y}}_{t+1} - \bar{y}_{t+1}^{VAR} \mathbf{1}) \quad (9.11)$$

$$\hat{\mathbf{y}}_{t+1} = \mathbf{a} + \mathbf{y}_t \mathbf{B}_t \quad (9.12)$$

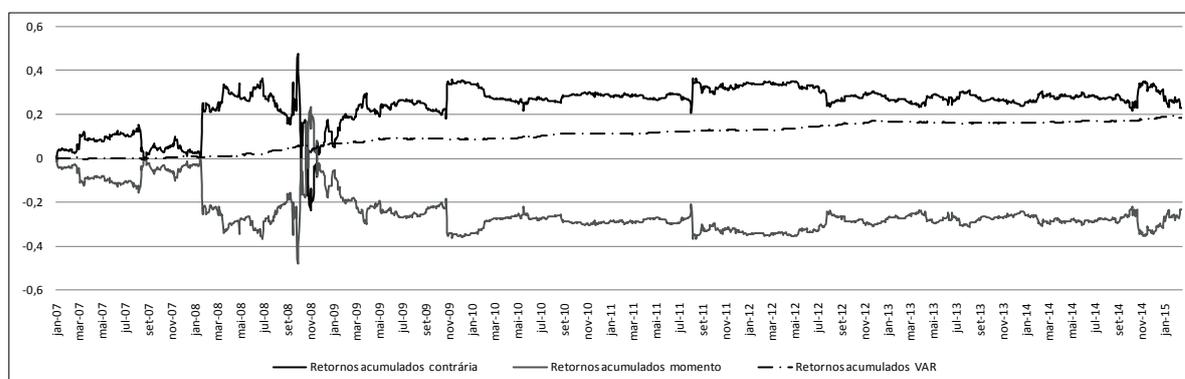
$$\bar{y}_{t+1}^{VAR} = \frac{(\mathbf{a} + \mathbf{y}_t \mathbf{B}_t)' \mathbf{1}}{n} \quad (9.13)$$

### 9.2.1 Resultados

Esta seção aborda os resultados encontrados para as estratégias de arbitragem no mercado acionário no período vigente da carteira. Para confrontar melhor as medidas de desempenho, foi plotado o gráfico 25 no que tange ao acumulado dos retornos diários de

cada carteira. Esse método é preferível, uma vez que é mais fácil de visualizar o que ocorre temporalmente se permanecermos rebalanceando as carteiras diariamente de acordo com as estratégias descritas acima.

Figura 25 – Retornos acumulados - Carteiras de arbitragem



Fonte: elaborada pelo autor (2015)

O primeiro resultado visível dessa análise encontra-se na variância dos retornos da carteira, extremamente mais voláteis para os casos das carteiras contrária e momento. Ambos portfólios levam em consideração os retornos apresentados em  $t$  para otimizar a carteira em  $t$ , essas estratégias então, incorporam toda a volatilidade do mercado, esse aspecto fica mais visível quando olhamos para o período conturbado da crise do *subprime* no ano de 2008.

O segundo resultado deste gráfico ainda refere-se a essas duas carteiras. Percebe-se então que a estratégia contrária é preferível para o mercado brasileiro quanto à seleção dos ativos mais líquidos feita para o período proposto. Como esperado, os retornos de uma são opostos ao de outra. O fato de termos a carteira contrária com o melhor desempenho é um indício forte de que no geral esses ativos apresentam uma autocorrelação negativa. Esse perfil é fortemente afetado no ano de 2008, provavelmente por uma quebra no regime estocástico dos dados no período, em outras palavras, o mercado passa a ter um efeito momento de  $t$  para  $t + 1$ . Esse efeito pode ser atribuído sobretudo à um comportamento de "manada" dos investidores em períodos de crise, levando os papéis à desvalorizações recursivas.

Por último, esse gráfico ainda apresenta de forma visível as vantagens de uma aplicação VAR para a definição dos pesos dos ativos na carteira. Trata-se então de uma carteira muito estável, principalmente se comparada às demais. Essa estratégia ainda apresenta retornos acumulados sempre crescentes, conforme garantido em DeMiguel, Nogales e Uppal (2014). Um ponto não menos notável ainda refere-se à época de crise. Um comportamento mais padrão, por assim dizer, dos investidores nessa época, provavelmente foi captada pelos parâmetros variantes do modelo, os quais podem apresentar

uma reversão no sinal conforme se faça necessário. Desse ponto de vista, o período em que essa estratégia mais acumulou retornos foi justamente o período de crise.

Tabela 2 – Medidas de desempenho para as carteiras de arbitragem

	$\mu_{\times 10^{-4}}$	$\sigma^2_{\times 10^{-4}}$	Sharp $_{\times 10^{-1}}$
Carteira contrária	1,14922689	2,73207755	6,95279250
Carteira momento	-1,14922689	2,73207755	-6,95279250
Carteira VAR	0,91395458	0,07685449	32,96780152

Fonte: elaborada pelo autor (2015)

Para finalizar a análise dessas estratégias, foram compiladas em 2 as medidas de desempenho médias diárias para todo o período completo da análise. Esses dados estão de acordo com o apresentado em 25, a média diária dos retornos é maior para a carteira contrária, e suavemente menor para a estratégia VAR. Por outro lado, a carteira VAR tem ganhos extremamente altos quanto à sua estabilidade, dada pela variância muito mais baixa que as outras estratégias. Decorrente deste fato, o Índice de Sharpe ficou bem mais alto que as demais, logo essa carteira apresenta-se como uma boa alternativa para aqueles investidores mais avessos ao risco.

### 9.3 Carteiras baseadas na média-variância

Desse grupo de carteiras, são estudadas no trabalho algumas que desconsideram a dependência temporal da média nas séries de retornos, utilizadas aqui como *benchmark* a nível de comparação.

A primeira delas trata-se de uma carteira ingênua, onde todos os ativos tem peso igual, dado por  $w_i = (1/n)$ . Essa estratégia é levada em consideração pela sua atratividade, pois não há *turnover*, logo o investidor não tem a necessidade de rebalanceá-la a cada  $t$ .

A segunda estratégia, muito comum na literatura, leva em consideração a estrutura de covariâncias entre os ativos. Assumindo que o investidor seja totalmente avesso ao risco, este estará interessado apenas em reduzir o risco do portfólio de forma que os pesos são dados pelo processo de minimização:

$$\min_w \mathbf{w}'_t \Sigma_t \mathbf{w}_t \quad (9.14)$$

$$\text{s.a } \mathbf{w}' \mathbf{1} = 1 \quad (9.15)$$

A restrição sobre os pesos desta carteira são necessários, uma vez que se  $\Sigma_t$  é positiva definida, o processo de otimização atribuiria peso zero à todos os ativos. O valor

unitário então foi escolhido de forma que 100% do investimento seja alocado otimamente entre os ativos. Perceba que, sob essa estratégia, é permitido que  $w_i < 0$  quando necessário, o que na prática aumenta o *turnover* da carteira. Então dentro da mesma estratégia, ainda foi considerado a restrição no espaço dos pesos de forma que  $w_i > 0$  para todo  $i = 1, \dots, n$ , muito comum na prática.

Demais considerações sobre esta estratégia, é necessária a estimação da matriz de covariâncias. Restringimo-nos então a duas possibilidades, a primeira delas é uma estimação simples do tipo EWMA, cujo parâmetro foi escolhido de acordo com a recomendação RiskMetrics, similar ao caso do BVAR do Koop e Korobilis (2013) para a estimação da matriz de covariâncias:

$$\Sigma_{t+1} = (1 - B)(\mathbf{y}_{t-1}\mathbf{y}'_{t-1}) + B\Sigma_{t-1} \quad (9.16)$$

$$B = 0,94 \quad (9.17)$$

A segunda estimação foi escolhida pelo seu perfil mais informativo, o modelo BEKK-GARCH(1,1), com a restrição de que as matrizes de parâmetros sejam diagonais, uma vez que a versão escalar se assemelha muito ao EWMA e o BEKK-GARCH completo sofre pela sua dimensionalidade e instabilidade nos parâmetros:

$$\Sigma_t = \text{diag}(\mathbf{A})\mathbf{y}_{t-1}\mathbf{y}'_{t-1}\text{diag}(\mathbf{A}) + \text{diag}(\mathbf{B})\mathbf{H}_{t-1}\text{diag}(\mathbf{B})$$

A próxima estratégia assemelha-se à mínima variância, com a diferença que agora levamos em consideração os retornos na média:

$$\min_w \mathbf{w}'_t \Sigma_t \mathbf{w}_t - \frac{1}{\gamma} \mathbf{w}'_t \boldsymbol{\mu}_t \quad (9.18)$$

$$\text{s.a } \mathbf{w}' \mathbf{1} = 1 \quad (9.19)$$

Onde  $\boldsymbol{\mu}_t$  é o vetor de médias dos retornos e  $\gamma$  é o grau de aversão ao risco do investidor, cujos valores testados nesse trabalho seguem DeMiguel, Nogales e Uppal (2014), para  $\gamma = [1 \quad 2 \quad 10]$ . Atente que se  $\gamma \rightarrow \infty$ , então a carteira converge para o caso da variância mínima.

Sob esta estratégia então, dois tipos de abordagem para a estimação foram escolhidos, a primeira leva em consideração que os retornos seguem um processo tipo VAR(1), cuja estimação segue Koop e Korobilis (2013) tal como as estratégias de arbitragem com base no VAR, estimação dos parâmetro a qual fora dada na metodologia deste capítulo, própria para modelos de grandes dimensões. Desta forma:

$$\boldsymbol{\mu}_{t+1}^{VAR} = \hat{\mathbf{y}}_{t+1} = \mathbf{a}_t + \mathbf{B}_t \mathbf{y}_t$$

E a matriz de covariâncias  $\boldsymbol{\Sigma}_t$  é dada pelo processo EWMA intrínseca ao modelo.

A segunda estratégia segue a proposta não paramétrica em DeMiguel, Nogales e Uppal (2014), cujas estimações dos parâmetros da carteira são condicionados à informação disponível em  $t$  através da média amostral e matriz de covariâncias amostral, dadas colocando-se pesos iguais aos  $K$  dias que antecedem à estimação, fazendo-se desta forma uma janela móvel:

$$\boldsymbol{\mu}_{t+1}^{NAR} = \frac{1}{k} \sum_{k=1}^K \mathbf{y}_{t-k} \quad \boldsymbol{\Sigma}_{t+1}^{NAR} = \frac{1}{K-1} \sum_{k=1}^K (\mathbf{y}_{t-k} - \boldsymbol{\mu}_{t+1})(\mathbf{y}_{t-k} - \boldsymbol{\mu}_{t+1})' \quad (9.20)$$

Onde a fim de referência para a análise, foi tomado  $k = 5$ , supondo que os retornos da semana que passou são suficientes para predizer o retorno do dia seguinte.

Para as carteiras apresentadas de média-variância até então, foram considerados ambos os casos quando há venda a descoberto e quando não há, para este último caso,  $w_i > 0$  para todo  $i = 1, \dots, n$ .

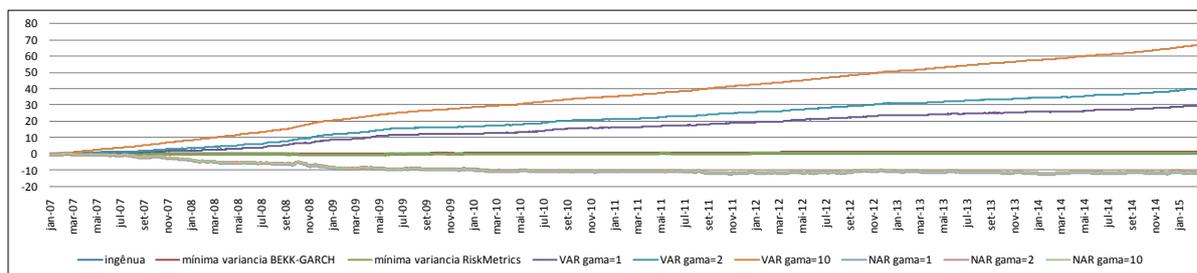
Por fim, há de ser levado em consideração que carteiras de média-variância são muito instáveis, sobretudo para os casos quando o parâmetro  $\gamma$  é baixo. Para contornar esse problema, foi restringido que, para quaisquer estratégias, seja com venda a descoberto ou sem, foi restringido então que nenhum ativo possa ocupar mais de 50% na alocação da carteira.

### 9.3.1 Resultados

Esta seção apresenta os resultados referentes às estratégias de alocação de portfólio baseadas na média-variância. O primeiro conjunto avaliado foram as carteiras menos restritivas, onde permitiu-se que estas alocassem os ativos com pesos negativos conforme fosse necessário.

Como pode-se observar em 26, as carteiras que apresentaram o melhor desempenho foram aquelas baseadas no VAR, provavelmente porque este foi responsável por captar bem os retornos um passo à frente. Percebe-se ainda que dentre estas carteiras, a que apresentou o melhor resultado foi aquela cujo investidor teria maior aversão ao risco, contrariando a suposição inicial de que assumir um risco maior implica em maiores retornos. Vale ressaltar aqui que essas carteiras foram restritas de forma que nenhum ativo possa ter um peso maior que 50% na alocação da carteira. Essa restrição mostra-se aplicável na prática, uma vez que essas carteiras são extremamente instáveis. Todavia, apesar de apresentadas aqui, são carteiras teóricas, a restrição imposta aos pesos não foi capaz de conter o alto *turnover* destas estratégias, como apresentado em 3.

Figura 26 – Retornos acumulados - Retorno acumulado - Carteiras de média-variância com venda a descoberto



Fonte: elaborada pelo autor (2015)

O segundo grupo tem seus retornos acumulados muito próximo de zero, e referem-se às carteiras com base somente na média. A junção dessa informação com a anterior induz a pensar que deve haver um grau de aversão ao risco ótimo, uma vez que se  $\gamma \rightarrow \infty$ , então esta deveria convergir à curva referente à estratégia RiskMetrics (perto de zero), dado que a metodologia para o cálculo da matriz de covariâncias feita através de Koop e Korobilis (2013) foi a mesma, com a mesma escolha do parâmetro.

Por fim, do gráfico e da tabela referentes à estratégia com venda a descoberto, temos o último grupo de estratégias, baseadas então no NAR com 5 dias de cotação anteriores para o cálculo dos pesos ótimos. Essas carteiras não mostraram um bom desempenho, além da alta variância, o retorno médio diário foi negativo. Um indício que, se os ativos apresentaram um bom desempenho em média na semana que passou, não pode ser considerado um sinal que esses resultados persistirão. Outro ponto a ser notado ocorre quanto à robustez dessa estratégia quanto ao grau de aversão ao risco, todas as curvas e retornos foram muito similares, mesmo para o caso quando  $\gamma = 10$ .

Tabela 3 – Medidas de desempenho para as carteiras média-variância com venda a descoberto

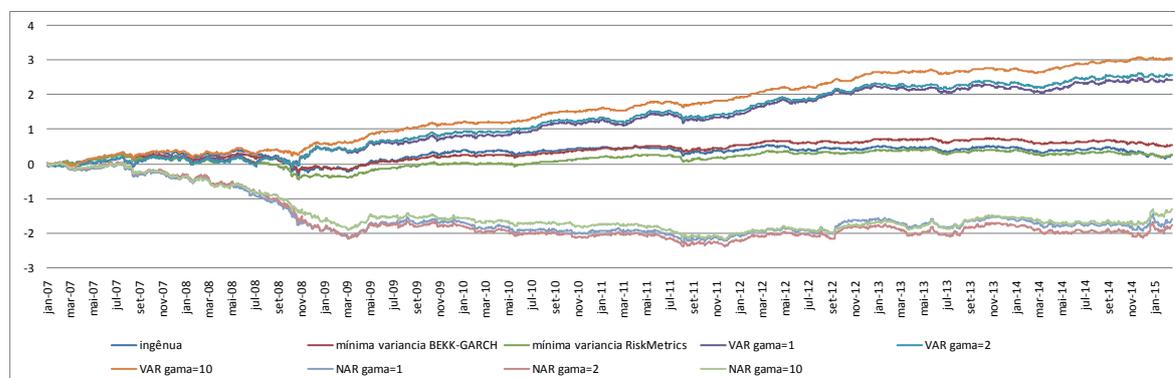
	$\mu \times 10^{-2}$	$\sigma^2 \times 10^{-3}$	Sharpe	turnover
Carteira ingênua	0,012799553	0,241970348	0,026020386	-
<b>Carteira mínima variância</b>				
RiskMetrics	0,031651424	0,146774873	0,082616614	0,928852922
BEKK-GARCH(1,1)	0,049156143	0,082268198	0,171380658	0,151875536
<b>Carteira média-variância</b>				
Carteira VAR( $\gamma = 1$ )	1,467732206	2,574751135	0,914701087	17,14164939
Carteira VAR( $\gamma = 2$ )	1,970546658	1,783715139	1,475448222	16,56527144
Carteira VAR( $\gamma = 10$ )	3,328179503	0,560741534	4,444525262	15,20363572
<b>Carteira NAR</b>				
Carteira NAR( $\gamma = 1$ )	-0,585277653	8,812280462	-0,197159536	9,372708238
Carteira NAR( $\gamma = 2$ )	-0,560399222	7,721016493	-0,201678816	9,221424911
Carteira NAR( $\gamma = 10$ )	-0,585277653	8,812280462	-0,197159536	8,698552607

Fonte: elaborada pelo autor (2015)

A segunda parte da análise foi feita em torno das carteiras mais restritas, cujos pesos puderam variar apenas com valores positivos, caso quando o investidor assume

apenas posições compradas na bolsa.

Figura 27 – Retornos acumulados - Retorno acumulado - Carteiras de média-variância sem venda a descoberto



Fonte: elaborada pelo autor (2015)

O gráfico 27 apresenta os resultados para as estratégias baseadas em tal restrição. Em conformidade com o esperado, o desenho do gráfico assemelha-se ao caso com vendas a descoberto, porém com retornos menos (atente à escala). Novamente, os gráficos baseados no VAR apresentaram melhores resultados porém, por outro lado, o *turnover* foi maior, porém essa estratégia parece ser mais factível de ser aplicada se os custos de operações não forem altos. Outro ponto similar tange ao grupo sob a estimação via NAR. Novamente, não mostrou-se como uma boa estratégia, mesmo quando o  $\gamma$  varia, há certa resistência dessa estratégia mesmo quando o grau de aversão ao risco aumenta. Deste gráfico, percebemos que o período quando mais caiu foi justamente durante a instabilidade da crise de 2008 (atente que as estratégias com VAR, principalmente para  $\gamma = 10$ , mostrou certa resistência na queda). Os dados para esse grupo está sumarizado na tabela a seguir, (4).

Tabela 4 – Medidas de desempenho para as carteiras média-variância sem venda a descoberto

	$\mu \times 10^{-3}$	$\sigma^2 \times 10^{-4}$	Sharpe	<i>turnover</i>
Carteira ingênua	0,127995532	2,419703476	0,822836856	-
<b>Carteira mínima variância</b>				
RiskMetrics	0,127430339	1,251339911	1,139161219	0,152702821
BEKK-GARCH(1,1)	0,275259689	1,243721035	2,468204414	0,041224172
<b>Carteira média-variância</b>				
Carteira VAR( $\gamma = 1$ )	1,218235662	3,289029542	6,717341403	1,697408249
Carteira VAR( $\gamma = 2$ )	1,284318982	2,909814864	7,529050712	1,682937708
Carteira VAR( $\gamma = 10$ )	1,518381745	1,605740998	11,98238379	1,422930891
Carteira NAR( $\gamma = 1$ )	-0,792083666	5,817357123	-3,284038159	1,046675044
Carteira NAR( $\gamma = 2$ )	-0,878956328	5,581305054	-3,720483795	1,056357654
Carteira NAR( $\gamma = 10$ )	-0,64852091	4,183245439	-3,170788764	0,978320972

Fonte: elaborada pelo autor (2015)

## 10 CONSIDERAÇÕES FINAIS

Os modelos baseados na variância dos retornos dos ativos tem sido amplamente abordados na literatura. Essa área da pesquisa, então, encontrou avanços significativos, ao passo que a modelagem na média, não.

Esse trabalho cumpriu o papel de inserir, além das usuais estratégias de alocação de ativos baseadas na volatilidade, a previsão dos retornos na média, feita então a partir do já popular modelo linear de vetor autorregressivo VAR.

Para tal abordagem, foi escolhida a recente estratégia proposta por DeMiguel, Nogales e Uppal (2014), o qual remonta duas classes de carteiras diferentes sob a predição dos retornos com base no VAR, a saber, carteiras de arbitragem, ou custo zero, e carteiras de média-variância. Em ambas classes, as novas carteiras foram comparadas com estratégias já consagradas na literatura.

Deste modo, essa dissertação visa dar resposta duas questões relevantes à literatura na área: os retornos no mercado acionário podem ser previstos? Caso positivo, é possível montar estratégias que leve essa informação à mercados mais voláteis, e.g. o caso de países emergentes como o Brasil?

A resposta encontrada foi sim para as duas perguntas, o que nos leva a crer que, por menor que seja a informação contida em  $t$  para prever retornos em  $t + 1$ , esta deve ser utilizada, ou em outras palavras, a previsão dos retornos diários pôde ser feita com base nas informações contidas nesses mercados um dia anterior à previsão. O método VAR então foi capaz de dar pesos aos ativos de acordo com o processos estocásticos em cada série de retornos, sobretudo no que tange ao sinal da autocorrelação apresentada por elas.

Visando dar flexibilidade ao modelo, permitiu-se então que os parâmetros variassem no tempo de acordo com um passeio aleatório. Essa estratégia mostrou-se eficaz quanto à captação da dinâmica volátil dos mercados. Diferentemente de DeMiguel, Nogales e Uppal (2014), esse trabalho utilizou-se de métodos mais sofisticados de estimação dos parâmetros variantes, os quais foram feitos através da inferência Bayesiana, a qual se utiliza de métodos de reamostragem do tipo MCMC. Apesar dos altos custos computacionais, esses métodos vêm ganhando força na literatura pela sua flexibilidade, e por driblar de forma mais natural o já conhecido problema da sobre parametrização de modelos multivariados.

Para à análise da dinâmica de mercado, foi utilizada a metodologia proposta em Primiceri (2005), a qual fora inicialmente utilizada para dados macroeconômicos. Essa metodologia adaptou-se de forma natural ao problema proposto aqui e foi capaz de indicar padrões apresentados nos mercados de ações. Dos melhores resultados, pode-se destacar que o valor dos parâmetros divergiu de 0 em grande escala em alguns períodos de tempo,

indicando que os retornos das carteiras não apresentam as propriedades de um ruído branco. Outro resultado notável tange à variância alta dos parâmetros, ao passo que estes divergem de 0, não mantém-se por muito tempo numa posição fixa, um indicativo que os mercados se ajustam rapidamente às informações contidas nas séries. Um terceiro resultado importante dessa análise pode ainda ser atribuído à variância dos parâmetros, a qual apresentou-se menor para o caso de carteiras mais diversificadas, o que é importante quando se espera estabilidade no investimento.

Uma vez feita a análise da dinâmica dos parâmetros, seguiu-se para a construção de carteiras que possam incorporar a informação contida neles. Para as estratégias de arbitragem, o VAR não garantiu maiores retornos que aqueles obtidos para a carteira contrária, por outro lado, o VAR apresentou uma estabilidade dos retornos muito maior, e constantemente acima de 0. Como consequência, o Índice de Sharpe foi bem mais alto que das demais estratégias.

Para as carteiras de média-variância, as estratégias baseadas no VAR garantiram maiores retornos que as demais, mesmo com a variância mais alta, o que pôde ser expresso pelo índice de Sharp mais alto. Em contraste com o alto desempenho do VAR, o NAR apresentou retornos médios negativos, nesse trabalho foi testado apenas uma janela para os últimos 5 dias de cotação, ou de outra forma, é possível dizer que se um ativo apresentou um bom retorno na semana que passou, essa informação não pode ser levada em consideração na expectativa que o desempenho permaneça constante para o próximo dia. Outro ponto relevante da análise foi o alto *turnover* para o caso VAR. ao montar essa carteira, é importante levar em consideração os custos de transação, restringindo a variação dos pesos dos ativos.

Por fim, este trabalho espera ter aberto caminho para a incorporação da modelagem na média para os retornos de ativos. Por menor que seja a informação obtida na previsão um passo à frente, os resultados foram satisfatórios e incorporou ganhos às estratégias, motivando, deste modo, o uso da metodologia em outras estratégias aqui não abordadas.

## REFERÊNCIAS

- AGUILAR, O. et al. Bayesian inference on latent structure in time series. *Bayesian Statistics*, v. 6, n. 1, p. 1–16, 1998.
- AGUILAR, O.; WEST, M. Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics*, Washington, v. 18, n. 3, p. 338–357, 2000. Disponível em: <<http://ftp.stat.duke.edu/WorkingPapers/98-03.pdf>>.
- ALEXANDER, C. Orthogonal garch. *Mastering risk*, v. 2, p. 21–38, 2001.
- ANDERSON, T. W. et al. *An introduction to multivariate statistical analysis*. New York: Wiley, 1958.
- BARNDORFF-NIELSEN, O. E. Normal inverse gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of statistics*, Stockholm, v. 24, n. 1, p. 1–13, 1997.
- BARNDORFF-NIELSEN, O. E.; SHEPHARD, N. Non-gaussian ornstein–uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Oxford, v. 63, n. 2, p. 167–241, 2001.
- BAUWENS, L.; PREMINGER, A.; ROMBOUTS, J. V. Regime switching garch models. 2006. Disponível em SSRN: <<http://ssrn.com/abstract=914144>> ou <<http://dx.doi.org/10.2139/ssrn.914144>>.
- BILLIO, M.; CAPORIN, M.; GOBBO, M. *Block dynamic conditional correlation multivariate GARCH models*. [S.l.], 2003.
- BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, Amsterdam, v. 31, n. 3, p. 307–327, 1986. Disponível em: <[http://public.econ.duke.edu/~boller/Published\\_Papers/joe\\_86.pdf](http://public.econ.duke.edu/~boller/Published_Papers/joe_86.pdf)>.
- BOLLERSLEV, T. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *The Review of Economics and Statistics*, New York, p. 498–505, 1990. Disponível em: <[http://public.econ.duke.edu/~boller/Published\\_Papers/restat\\_90.pdf](http://public.econ.duke.edu/~boller/Published_Papers/restat_90.pdf)>.
- BOLLERSLEV, T.; ENGLE, R. F.; WOOLDRIDGE, J. M. A capital asset pricing model with time-varying covariances. *The Journal of Political Economy*, New York, p. 116–131, 1988. Disponível em: <[http://public.econ.duke.edu/~boller/Published\\_Papers/jpe\\_88.pdf](http://public.econ.duke.edu/~boller/Published_Papers/jpe_88.pdf)>.
- CALDEIRA, J. F.; MOURA, G. V.; SANTOS, A. A. Seleção de carteiras utilizando o modelo fama-french-carhart. *Revista Brasileira de Economia*, Rio de Janeiro, v. 67, n. 1, p. 45–65, 2013.
- CAMPBELL, J. Y. et al. *The econometrics of financial markets*. Princeton: University press Princeton, 1997. Disponível em: <[http://rum.prf.jcu.cz/public/mecirova/eng\\_ekonomka/Campbell\\_Ho\\_Mackinley\\_Ebooks\\_Econometrics\\_The.pdf](http://rum.prf.jcu.cz/public/mecirova/eng_ekonomka/Campbell_Ho_Mackinley_Ebooks_Econometrics_The.pdf)>.

- CAPOBIANCO, E. State-space stochastic volatility models: A review of estimation algorithms. *Applied Stochastic Models and Data Analysis*, v. 12, n. 4, p. 265–279, 1996.
- CAPPIELLO, L.; ENGLE, R. F.; SHEPPARD, K. Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial econometrics*, Cary, v. 4, n. 4, p. 537–572, 2006.
- CARHART, M. M. On persistence in mutual fund performance. *Journal of finance*, New York, v. 52, n. 1, p. 57–82, 1997. Disponível em: <[http://faculty.chicagobooth.edu/john.cochrane/teaching/35150\\_advanced\\_investments/Carhart\\_funds\\_jf.pdf](http://faculty.chicagobooth.edu/john.cochrane/teaching/35150_advanced_investments/Carhart_funds_jf.pdf)>.
- CARTER, C. K.; KOHN, R. On gibbs sampling for state space models. *Biometrika*, London, v. 81, n. 3, p. 541–553, 1994.
- CHIB, S.; NARDARI, F.; SHEPHARD, N. Markov chain monte carlo methods for stochastic volatility models. *Journal of Econometrics*, Amsterdam, v. 108, n. 2, p. 281–316, 2002.
- CHIB, S.; NARDARI, F.; SHEPHARD, N. Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, Amsterdam, v. 134, n. 2, p. 341–371, 2006. Disponível em: <<http://scholar.harvard.edu/files/multichib.pdf>>.
- COGLEY, T.; SARGENT, T. J. Evolving post-world war ii us inflation dynamics. In: *NBER Macroeconomics Annual*. [S.l.: s.n.]. v. 16.
- COGLEY, T.; SARGENT, T. J. Drifts and volatilities: monetary policies and outcomes in the post wwii us. *Review of Economic dynamics*, Orlando, v. 8, n. 2, p. 262–302, 2005.
- COORNISH, E. The multivariate t-distribution associated with a set of normal sample deviates. *Australian Journal of Physics*, Melbourne, v. 7, n. 4, p. 532–542, 1954.
- CUTHBERTSON, K.; HALL, S. G.; TAYLOR, M. P. *Applied econometric techniques*. [S.l.]: Harvester Wheatsheaf Hemel Hempstead, 1992.
- DANIELSSON, J. Stochastic volatility in asset prices estimation with simulated maximum likelihood. *Journal of Econometrics*, Amsterdam, v. 64, n. 1, p. 375–400, 1994.
- DEMIGUEL, V.; NOGALES, F. J.; UPPAL, R. Stock return serial dependence and out-of-sample portfolio performance. *Review of Financial Studies*, p. hhu002, 2014. Disponível em: <<http://faculty.london.edu/avmiguel/DNU-RFS.pdf>>.
- ENGLE, R. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, Washington, v. 20, n. 3, p. 339–350, 2002. Disponível em: <<http://pages.stern.nyu.edu/~rengle/dccfinal.pdf>>.
- ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, Chicago, p. 987–1007, 1982. Disponível em: <<http://www.econ.uiuc.edu/~econ508/Papers/engle82.pdf>>.
- ENGLE, R. F.; KRONER, K. F. Multivariate simultaneous generalized arch. *Econometric theory*, Cambridge, v. 11, n. 01, p. 122–150, 1995.

- ENGLE, R. F.; NG, V. K.; ROTHSCILD, M. Asset pricing with a factor-arch covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics*, Amsterdam, v. 45, n. 1, p. 213–237, 1990. Disponível em: <<https://www.stern.nyu.edu/rengle/EngleNgRothschild.pdf>>.
- FAMA, E. F.; FRENCH, K. R. Multifactor explanations of asset pricing anomalies. *The journal of finance*, New York, v. 51, n. 1, p. 55–84, 1996. Disponível em: <[http://faculty.chicagobooth.edu/john.cochrane/teaching/35904\\_asset\\_pricing/Fama\\_French\\_multifactor\\_explanations.pdf](http://faculty.chicagobooth.edu/john.cochrane/teaching/35904_asset_pricing/Fama_French_multifactor_explanations.pdf)>.
- FRANCQ, C.; ZAKOIAN, J.-M. *GARCH models: structure, statistical inference and financial applications*. Chichester: John Wiley, 2011.
- GALLANT, A. R.; HSIEH, D.; TAUCHEN, G. Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics*, Amsterdam, v. 81, n. 1, p. 159–192, 1997. Disponível em: <[http://dukespace.lib.duke.edu/dspace/bitstream/handle/10161/2057/Hsieh\\_estimation\\_of\\_stochastic\\_volatility.pdf%3Fsequenc](http://dukespace.lib.duke.edu/dspace/bitstream/handle/10161/2057/Hsieh_estimation_of_stochastic_volatility.pdf%3Fsequenc)>.
- GHYSELS, E. et al. *Stochastic volatility*. Montreal: CIRANO, 1995.
- GLOSTEN, L. R.; JAGANNATHAN, R.; RUNKLE, D. E. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, New York, v. 48, n. 5, p. 1779–1801, 1993. Disponível em: <<http://faculty.washington.edu/ezivot/econ589/GJRJOF1993.pdf>>.
- GOURIÉROUX, C.; JASIAK, J.; SUFANA, R. The wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, Amsterdam, v. 150, n. 2, p. 167–181, 2009. Disponível em: <[http://www.ims.nus.edu.sg/Programs/econometrics/files/cg\\_paper3.pdf](http://www.ims.nus.edu.sg/Programs/econometrics/files/cg_paper3.pdf)>.
- HAMILTON, J. D. *Time series analysis*. Princeton: University Press Princeton, 1994.
- HARVEY, A.; RUIZ, E.; SHEPHARD, N. Multivariate stochastic variance models. *The Review of Economic Studies*, Oxford, v. 61, n. 2, p. 247–264, 1994. Disponível em: <[http://www.stat.ntnu.edu.tw/~njhsu/Nonlinear%20Time%20Series/Multivariate%20SV\(Harvey-Ruiz-Shephard\).pdf](http://www.stat.ntnu.edu.tw/~njhsu/Nonlinear%20Time%20Series/Multivariate%20SV(Harvey-Ruiz-Shephard).pdf)>.
- HARVEY, A. C.; SHEPHARD, N. Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business & Economic Statistics*, Washington, v. 14, n. 4, p. 429–434, 1996.
- JACQUIER, É.; POLSON, N. G.; ROSSI, P. E. *Models and priors for multivariate stochastic volatility*. Montreal, 1995.
- JACQUIER, E.; POLSON, N. G.; ROSSI, P. E. Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, Washington, v. 20, n. 1, p. 69–87, 2002.
- JACQUIER, E.; POLSON, N. G.; ROSSI, P. E. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, Amsterdam, v. 122, n. 1, p. 185–212, 2004.
- JEANTHEAU, T. Strong consistency of estimators for multivariate arch models. *Econometric theory*, Cambridge, v. 14, n. 01, p. 70–86, 1998.

- JENSEN, M. C. Some anomalous evidence regarding market efficiency. *Journal of financial economics*, Lausanne, v. 6, n. 2, p. 95–101, 1978. Disponível em: <<http://www.e-m-h.org/Jens78.pdf>>.
- JUNGBACKER, B.; KOOPMAN, S. J. Monte carlo likelihood estimation for three multivariate stochastic volatility models. *Econometric Reviews*, New York, v. 25, n. 2-3, p. 385–408, 2006.
- KARLSSON, S. et al. *Conditional posteriors for the reduced rank regression model*. Örebro, 2012. (Working Papers, 2012-11).
- KIM, C.-J.; NELSON, C. R. *State-space models with regime switching*. Cambridge: Massachusetts Institute of Technology, 1999.
- KIM, S.; SHEPHARD, N.; CHIB, S. Stochastic volatility: likelihood inference and comparison with arch models. *The Review of Economic Studies*, Oxford, v. 65, n. 3, p. 361–393, 1998. Disponível em: <<http://finance.martinsewell.com/stylized-facts/volatility/KimShephardChib1998.pdf>>.
- KOOP, G.; KOROBILIS, D. *Bayesian multivariate time series methods for empirical macroeconomics*. Rimini: RCEA, 2010. (Working Paper Series, 47-09).
- KOOP, G.; KOROBILIS, D. Large time-varying parameter vars. *Journal of Econometrics*, Amsterdam, v. 177, n. 2, p. 185–198, 2013. Disponível em: <[https://mpra.ub.uni-muenchen.de/38591/1/MPRA\\_paper\\_38591.pdf](https://mpra.ub.uni-muenchen.de/38591/1/MPRA_paper_38591.pdf)>.
- KOOPMAN, S. J.; USPENSKY, E. H. The stochastic volatility in mean model: empirical evidence from international stock markets. *Journal of applied Econometrics*, Chichester, v. 17, n. 6, p. 667–689, 2002.
- LIESENFELD, R.; RICHARD, J.-F. Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of empirical finance*, Amsterdam, v. 10, n. 4, p. 505–531, 2003.
- LINTNER, J. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The review of economics and statistics*, Cambridge, p. 13–37, 1965. Disponível em: <<http://jpkc.fudan.edu.cn/picture/article/296/e8/42/91ba14764a0a8fa1617e387407af/9b1f9d38-cb13-4248-88f5-49dc3c1f545a.pdf>>.
- LITTERMAN, R. B. *Techniques of forecasting using vector autoregressions*. [S.l.], 1979. Disponível em: <<https://www.minneapolisfed.org/research/wp/wp115.pdf>>.
- LITTERMAN, R. B. *A Bayesian Procedure for Forecasting with Vector Autoregressions and Forecasting with Bayesian Vector Autoregressions—four Years of Experience*. [S.l.]: Federal Reserve Bank of Minneapolis, 1985.
- LITTERMAN, R. B. Forecasting with bayesian vector autoregressions?five years of experience. *Journal of Business & Economic Statistics*, Washington, v. 4, n. 1, p. 25–38, 1986.
- LO, A. W.; MACKINLAY, A. C. When are contrarian profits due to stock market overreaction? *Review of Financial studies*, Cary, v. 3, n. 2, p. 175–205, 1990. Disponível em: <<http://www.finance.martinsewell.com/stylized-facts/dependence/LoMacKinlay1990.pdf>>.

- MAHIEU, R.; SCHOTMAN, P. An empirical application of stochastic volatility models. *Journal of Applied Econometrics*, Chichester, v. 13, n. 4, p. 333–360, 1998. Disponível em: <<http://repub.eur.nl/pub/16768/mahieu.pdf>>.
- MARKOWITZ, H. Portfolio selection\*. *The journal of finance*, New York, v. 7, n. 1, p. 77–91, 1952. Disponível em: <[http://www.efalken.com/LowVolClassics/markowitz\\_JF1952.pdf](http://www.efalken.com/LowVolClassics/markowitz_JF1952.pdf)>.
- MORGAN, J. *Riskmetrics: technical document*. [S.l.]: Morgan Guaranty Trust Company of New York, 1996.
- NELSON, D. B. *The time series behavior of stock market volatility and returns*. Tese (Doutorado) — Massachusetts Institute of Technology, 1988.
- NELSON, D. B. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, Chicago, p. 347–370, 1991. Disponível em: <[http://www.samsi.info/sites/default/files/Nelson\\_1991.pdf](http://www.samsi.info/sites/default/files/Nelson_1991.pdf)>.
- PAN, M.-S.; LIANO, K.; HUANG, G.-C. Industry momentum strategies and autocorrelations in stock returns. *Journal of Empirical Finance*, Amsterdam, v. 11, n. 2, p. 185–202, 2004.
- PELLETIER, D. Regime switching for dynamic correlations. *Journal of econometrics*, Amsterdam, v. 131, n. 1, p. 445–473, 2006. Disponível em: <[http://www4.ncsu.edu/~dpellet/papers/dynamic\\_correlations.pdf](http://www4.ncsu.edu/~dpellet/papers/dynamic_correlations.pdf)>.
- PHILIPOV, A.; GLICKMAN, M. E. Multivariate stochastic volatility via wishart processes. *Journal of Business & Economic Statistics*, Washington, v. 24, n. 3, p. 313–328, 2006. Disponível em: <<http://www.hedgefund-index.com/philipov-multivariate.pdf>>.
- PITT, M.; SHEPHARD, N. Time varying covariances: a factor stochastic volatility approach. *Bayesian statistics*, Oxford, v. 6, p. 547–570, 1999.
- PLATANIOTI, K.; MCCOY, E.; STEPHENS, D. *A review of stochastic volatility: univariate and multivariate models*. [S.l.], 2005. Disponível em: <<http://www.imperial.ac.uk/~das01/Papers/TechReport2.pdf>>.
- PRIMICERI, G. E. Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, Oxford, v. 72, n. 3, p. 821–852, 2005. Disponível em: <[http://faculty.wcas.northwestern.edu/~gep575/tvsvar\\_final\\_july\\_04.pdf](http://faculty.wcas.northwestern.edu/~gep575/tvsvar_final_july_04.pdf)>.
- RABEMANANJARA, R.; ZAKOIAN, J.-M. Threshold arch models and asymmetries in volatility. *Journal of Applied Econometrics*, Chichester, v. 8, n. 1, p. 31–49, 1993.
- RAFTERY, A. E.; KÁRNÝ, M.; ETTLER, P. Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics*, Washington, v. 52, n. 1, p. 52–66, 2010. Disponível em: <<https://www.stat.washington.edu/raftery/Research/PDF/Karny2010.pdf>>.
- RAY, B. K.; TSAY, R. S. Long-range dependence in daily stock volatilities. *Journal of Business & Economic Statistics*, Washington, v. 18, n. 2, p. 254–262, 2000.
- ROSS, S. A. The arbitrage theory of capital asset pricing. *Journal of economic theory*, San Diego, v. 13, n. 3, p. 341–360, 1976. Disponível em: <<http://down.cenet.org.cn/upfile/36/2009323121657103.pdf>>.

- RUIZ, E. Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of econometrics*, Amsterdam, v. 63, n. 1, p. 289–306, 1994.
- SANDMANN, G.; KOOPMAN, S. J. Estimation of stochastic volatility models via monte carlo maximum likelihood. *Journal of Econometrics*, Amsterdam, v. 87, n. 2, p. 271–301, 1998. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.329.2011&rep=rep1&type=pdf>>.
- SÄRKKÄ, S. *Bayesian filtering and smoothing*. Cambridge: Cambridge University Press, 2013.
- SHARPE, W. F. Capital asset prices: A theory of market equilibrium under conditions of risk\*. *The journal of finance*, New York, v. 19, n. 3, p. 425–442, 1964. Disponível em: <[http://psc.ky.gov/psccef/2012-00221/rateintervention@ag.ky.gov/10252012f/sharpe\\_-\\_CAPM.pdf](http://psc.ky.gov/psccef/2012-00221/rateintervention@ag.ky.gov/10252012f/sharpe_-_CAPM.pdf)>.
- SHEPHARD, N. Fitting nonlinear time-series models with applications to stochastic variance models. *Journal of Applied Econometrics*, Chichester, v. 8, n. S1, p. S135–S152, 1993.
- SHEPHARD, N. *Statistical aspects of ARCH and stochastic volatility*. London: Chapman & Hall, 1996. 1–68 p.
- SILVENNOINEN, A.; TERÄSVIRTA, T. Modeling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation garch model. *Journal of Financial Econometrics*, Oxford, v. 7, n. 4, p. 373–411, 2009.
- SILVENNOINEN, A.; TERÄSVIRTA, T. Multivariate garch models. In: *Handbook of financial time series*. Berlin: [s.n.], 2009. p. 201–229.
- SIMS, C. A. Macroeconomics and reality. *Econometrica*, Chicago, p. 1–48, 1980. Disponível em: <[http://www.ekonometria.wne.uw.edu.pl/uploads/Main/macroeconomics\\_and\\_reality.pdf](http://www.ekonometria.wne.uw.edu.pl/uploads/Main/macroeconomics_and_reality.pdf)>.
- SMITH, M.; PITTS, A. Foreign exchange intervention by the bank of japan: Bayesian analysis using a bivariate stochastic volatility model. *Econometric Reviews*, New York, v. 25, n. 2-3, p. 425–451, 2006.
- SO, M. K.; LI, W.; LAM, K. A threshold stochastic volatility model. *Journal of Forecasting*, Chichester, v. 21, n. 7, p. 473–500, 2002.
- TAYLOR, S. Modelling financial time series. *Vlfiley*, New York, 1986.
- TAYLOR, S. J. Modeling stochastic volatility: A review and comparative study. *Mathematical finance*, Oxford, v. 4, n. 2, p. 183–204, 1994. Disponível em: <[http://www.stat.purdue.edu/~mlevins/STAT598K\\_2012/Taylor\\_1994.pdf](http://www.stat.purdue.edu/~mlevins/STAT598K_2012/Taylor_1994.pdf)>.
- TSAY, R. S. *Analysis of financial time series*. Cambridge: [s.n.], 2005. Disponível em: <<http://down.cenet.org.cn/upfile/28/2008111495713151.pdf>>.
- TSE, Y. K.; TSUI, A. K. C. A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics*, Washington, v. 20, n. 3, p. 351–362, 2002. Disponível em: <<http://homes.chass.utoronto.ca/~jmaheu/4050/TT.pdf>>.

UHLIG, H. Bayesian vector autoregressions with stochastic volatility. *Econometrica*, Chicago, p. 59–73, 1997. Disponível em: <<http://home.uchicago.edu/~huhlig/papers/uhligh.econometrica.1997.pdf>>.

WEIDE, R. Van der. Go-garch: a multivariate generalized orthogonal garch model. *Journal of Applied Econometrics*, Chichester, v. 17, n. 5, p. 549–564, 2002. Disponível em: <<http://www.eea-esem.com/papers/eea-esem/esem2002/1034/gogarch.pdf>>.

WISHART, J. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, JSTOR, p. 32–52, 1928.

WOLD, H. A study in the analysis of stationary time series. 1938.

YU, J.; MEYER, R. Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews*, New York, v. 25, n. 2-3, p. 361–384, 2006. Disponível em: <[http://www.mysmu.edu/faculty/yujun/Research/Yu\\_ER\\_2006B.pdf](http://www.mysmu.edu/faculty/yujun/Research/Yu_ER_2006B.pdf)>.

YU, J.; YANG, Z.; ZHANG, X. *A class of nonlinear stochastic volatility models and its implications on pricing currency options*. [S.l.], 2002. Disponível em: <<https://pdfs.semanticscholar.org/6f7f/3c220963578fc8fa0c621f0bc4eb10381b66.pdf>>.